Package ‘fasstr’

March 28, 2024

Title Encode, Summarize, and Visualize Daily Streamflow Data

Version 0.5.2

Description The Flow Analysis Summary Statistics Tool for R, ‘fasstr’, provides various func-
tions to tidy and screen daily stream discharge data, calculate and visualize various sum-
mary statistics and metrics, and compute annual trending and volume frequency analyses.
It features useful function arguments for filtering of and handling dates, customizing data and met-
rics, and the ability to pull daily data directly from the Water Survey of Canada hydromet-

Depends R (>= 3.3.0)

License Apache License 2.0


BugReports https://github.com/bcgov/fasstr/issues

Encoding UTF-8

Imports dplyr (>= 0.8.1), e1071 (>= 1.7.0.1), fitdistrplus (>=
1.0.14), ggplot2 (>= 3.1.0), grDevices, lubridate, openxlsx (>=
4.1.0), PearsonDS (>= 1.1), plyr (>= 1.8.4), purrr (>= 0.3.2),
RcppRoll (>= 0.3.0), scales (>= 1.0.0), tidyhydat (>= 0.4.0),
tidyR (>= 0.8.3), zyp (>= 0.10.1.1)

Suggests knitr, rmarkdown, testthat

RoxygenNote 7.3.1

VignetteBuilder knitr

NeedsCompilation no

Author Jon Goetz [aut, cre] (<https://orcid.org/0000-0002-4993-1119>),
Carl James Schwarz [aut],
Sam Albers [ctb] (<https://orcid.org/0000-0002-9270-7884>),
Robin Pike [ctb],
Province of British Columbia [cph]

Maintainer Jon Goetz <jon.goetz@gov.bc.ca>

Repository CRAN

Date/Publication 2024-03-28 19:20:02 UTC


R topics documented:

- `add_basin_area`.......................... 3
- `add_cumulative_volume`.................. 4
- `add_cumulative_yield`................... 6
- `add_daily_volume`....................... 8
- `add_daily_yield`........................ 9
- `add_date_variables`...................... 10
- `add_rolling_means`...................... 11
- `add_seasons`............................. 13
- `calc_all_annual_stats`................... 14
- `calc_annual_cumulative_stats`........... 18
- `calc_annual_extremes`................... 21
- `calc_annual_flow_timing`................ 24
- `calc_annual_highflows`.................. 26
- `calc_annual_lowflows`................... 29
- `calc_annual_normal_days`................. 32
- `calc_annual_outside_normal`............. 34
- `calc_annual_peaks`...................... 36
- `calc_annual_stats`...................... 39
- `calc_daily_cumulative_stats`............ 42
- `calc_daily_stats`....................... 45
- `calc_flow_percentile`................... 47
- `calc_longterm_daily_stats`.............. 49
- `calc_longterm_mean`..................... 52
- `calc_longterm_monthly_stats`......... 54
- `calc_longterm_percentile`............. 57
- `calc_monthly_cumulative_stats`........ 59
- `calc_monthly_stats`..................... 62
- `compute_annual_frequencies`............ 65
- `compute_annual_trends`.................. 68
- `compute_frequency_analysis`............ 73
- `compute_frequency_quantile`........... 75
- `compute_full_analysis`.................. 77
- `compute_hydat_peak_frequencies`....... 80
- `fill_missing_dates`..................... 83
- `plot_annual_cumulative_stats`......... 84
- `plot_annual_extremes`................... 87
- `plot_annual_extremes_year`............. 89
- `plot_annual_flow_timing`............... 92
- `plot_annual_flow_timing_year`......... 95
- `plot_annual_highflows`.................. 97
- `plot_annual_lowflows`................... 99
- `plot_annual_means`...................... 102
- `plot_annual_normal_days`............... 104
- `plot_annual_normal_days_year`......... 106
- `plot_annual_outside_normal`........... 109
- `plot_annual_stats`...................... 111
**add_basin_area**

*Add a basin area column to daily flows*

**Description**

Add a column of basin areas to a daily streamflow data set, in units of square kilometres.

**Usage**

```r
add_basin_area(data, groups = STATION_NUMBER, station_number, basin_area)
```

**Arguments**

- `data`: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
- `groups`: Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not `STATION_NUMBER`. Function will automatically group by a column named `STATION_NUMBER` if present. Remove the `STATION_NUMBER` column beforehand to remove this grouping. Leave blank if using station_number argument.
add_cumulative_volume

station_number  Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

basin_area  Upstream drainage basin area, in square kilometres, to apply to observations. Three options:
(1) Leave blank if groups is STATION_NUMBER with HYDAT station numbers to extract basin areas from HYDAT.
(2) A single numeric value to apply to all observations.
(3) List each basin area for each group/station in groups (can override HYDAT value if listed) as such c("08NM116" = 795,"08NM242" = 10). If group is not listed the HYDAT area will be applied if it exists, otherwise it will be NA.

Value

A tibble data frame of the original source data with an additional column:

Basin_Area_sqkm

area of upstream drainage basin area, in square kilometres

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {
  # Add the HYDAT basin area to a data frame with station numbers
  flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
  add_basin_area(data = flow_data)
  # Add the HYDAT basin area to data from HYDAT
  add_basin_area(station_number = "08NM116")
  # Set a custom basin area
  add_basin_area(station_number = "08NM116",
                 basin_area = 800)
  # Set multiple custom basin areas for multiple stations
  add_basin_area(station_number = c("08NM116", "08NM242"),
                 basin_area = c("08NM116" = 800, "08NM242" = 10))
}

add_cumulative_volume  Add a daily cumulative volumetric flows column to daily flows
**add_cumulative_volume**

**Description**

Add a column of rolling daily cumulative volumetric flows on an annual basis to a daily streamflow data set. Adds the volumetric discharge from each day with the previous day(s) for each year, in units of cubic metres. The cumulative flows restart every year and are only calculated in years with complete data.

**Usage**

```r
add_cumulative_volume(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  water_year_start = 1,
  months = 1:12
)
```

**Arguments**

- **data**: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
- **dates**: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
- **values**: Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
- **groups**: Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
- **station_number**: Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
- **water_year_start**: Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
- **months**: Numeric vector of months to add cumulative flows (e.g. 6:8 for Jun-Aug). Default accumulates to full years using all months (1:12).

**Value**

A tibble data frame of the source data with an additional column:
add_cumulative_yield

Cumul Volume m³

Cumulative volumetric flows for each day for each year, in units of cubic metres

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Add a column based on water years starting in August
add_cumulative_volume(station_number = "08NM116",
                      water_year_start = 8)
}

Description

Add a column of rolling daily cumulative water yields on an annual basis to a daily streamflow dataset. Adds the water yields from each day with the previous day(s) for each year, in units of millimetres. Converts cumulative discharge to a depth of water based on the upstream drainage basin area from basin_area argument. The cumulative flows restart every year and are only calculated in years with complete data.

Usage

add_cumulative_yield(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  basin_area,
  water_year_start = 1,
  months = 1:12
)

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
add_cumulative_yield

values  Name of column in data that contains numeric flow values, in units of cubic
metres per second. Only required if values column name is not 'Value' (default).
Leave blank if using station_number argument.

groups  Name of column in data that contains unique identifiers for different data sets, if
applicable. Only required if groups column name is not 'STATION_NUMBER'.
Function will automatically group by a column named 'STATION_NUMBER'
if present. Remove the 'STATION_NUMBER' column beforehand to remove
this grouping. Leave blank if using station_number argument.

station_number  Character string vector of seven digit Water Survey of Canada station numbers
(e.g. "08NM116") of which to extract daily streamflow data from a HYDAT
database. Requires tidyhydat package and a HYDAT database. Leave blank if
using data argument.

basin_area  Upstream drainage basin area, in square kilometres, to apply to observations.
Three options:
(1) Leave blank if groups is STATION_NUMBER with HYDAT station num-
bers to extract basin areas from HYDAT.
(2) A single numeric value to apply to all observations.
(3) List each basin area for each group/station in groups (can override HYDAT
value if listed) as such c("08NM116" = 795,"08NM242" = 10). If group is not
listed the HYDAT area will be applied if it exists, otherwise it will be NA.

water_year_start  Numeric value indicating the month (1 through 12) of the start of water year for
analysis. Default 1.

months  Numeric vector of months to add cumulative flows. For example, 3 for March,
6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start
= 10 (Oct). Default summarizes all months (1:12).

Value

A tibble data frame of the source data with an additional column:

Cumul_Yield_mm  cumulative yield flows for each day for each year, in units of millimetres

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

  # Add a column based on water years starting in August
  add_cumulative_yield(station_number = "08NM116",
                       water_year_start = 8)

  # Add a column based on water years starting in August with a custom basin area to calculate yield
  add_cumulative_yield(station_number = "08NM116",
                       water_year_start = 8,
                       basin_area = 800)

}
add_daily_volume

Add a daily volumetric flows column to daily flows

Description

Add a column of daily volumetric flows to a daily streamflow data set, in units of cubic metres. Converts the discharge to a volume.

Usage

add_daily_volume(data, values = Value, station_number)

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.

values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.

station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

Value

A tibble data frame of the source data with an additional column:

Volume_m3 daily total volumetric flow, in units of cubic metres

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Add a column of daily flow volumes
add_daily_volume(station_number = "08NM116")

}
**add_daily_yield**  
**Add a daily volumetric water yield column to daily flows**

**Description**
Add a column of daily water yields to a daily streamflow data set, in units of millimetres. Converts the discharge to a depth of water based on the upstream drainage basin area.

**Usage**

```r
add_daily_yield(
    data,
    values = Value,
    groups = STATION_NUMBER,
    station_number,
    basin_area
)
```

**Arguments**
- `data` Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to `NULL` if using `station_number` argument.
- `values` Name of column in `data` that contains numeric flow values, in units of cubic metres per second. Only required if `values` column name is not 'Value' (default). Leave blank if using `station_number` argument.
- `groups` Name of column in `data` that contains unique identifiers for different data sets, if applicable. Only required if `groups` column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using `station_number` argument.
- `station_number` Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
- `basin_area` Upstream drainage basin area, in square kilometres, to apply to observations. Three options:
  1. Leave blank if `groups` is `STATION_NUMBER` with HYDAT station numbers to extract basin areas from HYDAT.
  2. A single numeric value to apply to all observations.
  3. List each basin area for each group/station in `groups` (can override HYDAT value if listed) as such c("08NM116" = 795, "08NM242" = 10). If group is not listed the HYDAT area will be applied if it exists, otherwise it will be NA.
add_date_variables

Value

A tibble data frame of the source data with an additional column:

Yield_mm  daily water yield, in units of millimetres

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Add a column of yields based on HYDAT basin area
add_daily_yield(station_number = "08NM116")

# Add a column of yields based on a custom basin area
add_daily_yield(station_number = "08NM116",
    basin_area = 800)

}

add_date_variables  Add year, month, and day of year variable columns to daily flows

Description

Add columns of CalendarYear (YYYY), Month (MM), MonthName (e.g. 'Jan'), WaterYear (YYYY), and DayofYear (1-365 or 366; of WaterYear); to a data frame with a column of dates called 'Date'. Water years are designated by the year in which they end. For example, Water Year 1999 (starting Oct) is from 1 Oct 1998 (DayofYear 1) to 30 Sep 1999 (DayofYear 365)).

Usage

add_date_variables(data, dates = Date, station_number, water_year_start = 1)

Arguments

data  Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.

dates  Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.

station_number  Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

water_year_start  Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
add_rolling_means

Value

A tibble data frame of the source data with additional columns:

- **CalendarYear**: calendar year
- **Month**: numeric month (1 to 12)
- **MonthName**: month abbreviation (Jan-Dec)
- **WaterYear**: year starting from the selected month start, water_year_start
- **DayofYear**: day of the year from the selected month start (1-365 or 366)

Examples

```r
# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {
  # Add date variables using calendar years
  add_date_variables(station_number = "08NM116")
  
  # Add date variables using water years starting in August
  add_date_variables(station_number = "08NM116",
                     water_year_start = 8)
}
```

Description

Adds selected n-day rolling means to a daily streamflow data set. Based on selected n-days and alignment, the rolling mean for a given day is obtained by averaging the adjacent dates of daily mean values. For example, rolling days of '7' and 'right' alignment would obtain a mean of the given and previous 6 days of daily mean flow.

Usage

```r
add_rolling_means(
  data,  
  dates = Date,  
  values = Value,  
  groups = STATION_NUMBER,  
  station_number,  
  roll_days = c(3, 7, 30),  
  roll_align = "right"
)
```
add_rolling_means

Arguments

- **data**: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
- **dates**: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
- **values**: Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
- **groups**: Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
- **station_number**: Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
- **roll_days**: Numeric values of the number of days to apply a rolling mean. Default c(3, 7, 30).
- **roll_align**: Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.

Value

A data frame of the source data with an additional column(s):

- **QnDay**: rolling means of the n-day flow values of the designated date and adjacent dates, direction of mean specified by roll_align

Default additional columns:

- **Q3Day**: rolling means of the 3-day flow values of the designated date and previous 2 days (roll_align = "right")
- **Q7Day**: rolling means of the 7-day flow values of the designated date and previous 6 days (roll_align = "right")
- **Q30Day**: rolling means of the 30-day flow values of the designated date and previous 29 days (roll_align = "right")

Examples

```r
# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Add default 3, 7, and 30-day rolling mean columns, with "right" alignment
add_rolling_means(station_number = "08NM116")
```

# Add custom 5 and 10-day rolling mean columns
add_rolling_means(station_number = "08NM116",
    roll_days = c(5,10))

# Add default 3, 7, and 30-day rolling mean columns, with "left" alignment
add_rolling_means(station_number = "08NM116",
    roll_align = "left")

---

**add_seasons**  
Add a column of seasons

---

**Description**  
Adds a column of seasons identifiers to a data frame with a column of dates called 'Date'. The length of seasons, in months, is provided using the `seasons_length` argument. As seasons are grouped by months the length of the seasons must be divisible into 12 with one of the following season lengths: 1, 2, 3, 4, 6, or 12 months. The start of the first season coincides with the start month of each year; 'Jan-Jun' for 6-month seasons starting with calendar years or 'Dec-Feb' for 3-month seasons starting with water year starting in December.

**Usage**

```r
add_seasons(
    data,
    dates = Date,
    station_number,
    water_year_start = 1,
    seasons_length
)
```

**Arguments**

- `data` Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using `station_number` argument.
- `dates` Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using `station_number` argument.
- `station_number` Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
- `water_year_start` Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
seasons_length  Numeric value indicating the desired length of seasons in months, divisible into 12. Required.

Value

A tibble data frame of the source data with additional column:

Season  season identifier labelled by the start and end month of the season

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Add a column with four annual seasons starting in January
add_seasons(station_number = "08NM116",
            seasons_length = 4)

# Add a column with two annual seasons (of 6 months length) starting in October
add_seasons(station_number = "08NM116",
            water_year_start = 10,
            seasons_length = 6)
}

---

\textit{calc_all_annual_stats}  \textit{Calculate all fasstr annual statistics}

Description

Calculates annual statistics from all annual fasstr functions from a daily streamflow data set. Data is ideally long-term and continuous with minimal missing/seasonal data as annual statistics are calculated. Calculates statistics from all values, unless specified. Returns a tibble with statistics. Data calculated using the following functions:

- calc_annual_stats()
- calc_annual_lowflows()
- calc_annual_cumulative_stats()
- calc_annual_flow_timing()
- calc_annual_normal_days()
- calc_monthly_stats()
Usage

calc_all_annual_stats(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  basin_area,
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  annual_percentiles = c(10, 90),
  monthly_percentiles = c(10, 20),
  stats_days = 1,
  stats_align = "right",
  lowflow_days = c(1, 3, 7, 30),
  lowflow_align = "right",
  timing_percent = c(25, 33, 50, 75),
  normal_percentiles = c(25, 75),
  transpose = FALSE,
  complete_years = FALSE,
  ignore_missing = FALSE,
  allowed_missing_annual = ifelse(ignore_missing, 100, 0),
  allowed_missing_monthly = ifelse(ignore_missing, 100, 0)
)

Arguments

data               Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates               Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values              Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups             Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number     Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT
calc_all_annual_stats

database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

**basin_area**

Upstream drainage basin area, in square kilometres, to apply to observations. Three options:
1. Leave blank if groups is STATION_NUMBER with HYDAT station numbers to extract basin areas from HYDAT.
2. A single numeric value to apply to all observations.
3. List each basin area for each group/station in groups (can override HYDAT value if listed) as such c("08NM116" = 795,"08NM242" = 10). If group is not listed the HYDAT area will be applied if it exists, otherwise it will be NA.

**water_year_start**

Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

**start_year**

Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

**end_year**

Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

**exclude_years**

Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

**months**

Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12). If not 1:12, seasonal total yield and volumetric flows will not be included.

**annual_percentiles**

Numeric vector of percentiles to calculate annually. Set to NA if none required. Used for calc_annual_stats() function. Default c(10,90).

**monthly_percentiles**

Numeric vector of percentiles to calculate monthly for each year. Set to NA if none required. Used for calc_monthly_stats() function. Default c(10,20).

**stats_days**

Numeric vector of the number of days to apply a rolling mean on basic stats. Default c(1). Used for calc_annual_stats() and calc_monthly_stats() functions.

**stats_align**

Character string identifying the direction of the rolling mean on basic stats from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'. Used for calc_annual_stats(), calc_monthly_stats(), and calc_annual_normal_days() functions.

**lowflow_days**

Numeric vector of the number of days to apply a rolling mean on low flow stats. Default c(1,3,7,30). Used for calc_lowflow_stats() function.

**lowflow_align**

Character string identifying the direction of the rolling mean on low flow stats from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'. Used for calc_lowflow_stats() function.

**timing_percent**

Numeric vector of percents of annual total flows to determine dates. Used for calc_annual_flow_timing() function. Default c(25,33.3,50,75).
calc_all_annual_stats

## Arguments

- `normal_percentiles`: Numeric vector of two values, lower and upper percentiles, respectively indicating the limits of the normal range. Default `c(25, 75)`.
- `transpose`: Logical value indicating whether to transpose rows and columns of results. Default `FALSE`.
- `complete_years`: Logical values indicating whether to include only years with complete data in analysis. Default `FALSE`.
- `ignore_missing`: Logical value indicating whether dates with missing values should be included in the calculation. If `TRUE` then a statistic will be calculated regardless of missing dates. If `FALSE` then only those statistics from time periods with no missing dates will be returned. Default `FALSE`.
- `allowed_missing_annual`: Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate an annual statistic (0 to 100 percent). If `ignore_missing = FALSE` then it defaults to 0 (zero missing dates allowed), if `ignore_missing = TRUE` then it defaults to 100 (any missing dates allowed); consistent with `ignore_missing` usage. Supersedes `ignore_missing` when used. Only for annual means, percentiles, minimums, and maximums.
- `allowed_missing_monthly`: Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a monthly statistic (0 to 100 percent). If `ignore_missing = FALSE` then it defaults to 0 (zero missing dates allowed), if `ignore_missing = TRUE` then it defaults to 100 (any missing dates allowed); consistent with `ignore_missing` usage. Supersedes `ignore_missing` when used. Only for monthly means, percentiles, minimums, and maximums.

## Value

A tibble data frame with column "Year" and then 107 (default) variables from the fasstr annual functions. See listed functions above for default variables. Transposing data creates a column of "Statistics" and subsequent columns for each year selected.

## See Also

- `calc_annual_stats`
- `calc_annual_lowflows`
- `calc_annual_cumulative_stats`
- `calc_annual_flow_timing`
- `calc_monthly_stats`
- `calc_annual_normal_days`

## Examples

```r
## Not run:

# Working examples:

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())){

# Calculate all annual statistics from this package with default arguments
calc_all_annual_stats(station_number = "08NM116")
```

# Calculate all annual statistics from this package with default arguments
# with some default arguments shown to customize metrics

```r
calc_all_annual_stats(station_number = "08NM116",
  annual_percentiles = c(10,90),
  monthly_percentiles = c(10,20),
  stats_days = 1,
  stats_align = "right",
  lowflow_days = c(1,3,7,30),
  lowflow_align = "right",
  timing_percent = c(25,33,50,75),
  normal_percentiles = c(25,75))

}
```

## End(Not run)

calc_annual_cumulative_stats

*Calculate annual (and seasonal) total cumulative flows*

**Description**

Calculates annual and seasonal total flows, as volumetric discharge or water yields, from a daily streamflow data set. For water year and seasonal data, the year is identified by the year in which the year or season ends. Two-seasons and four-seasons per year are calculated, with each 6 and 3-month seasons starting with the first month of the year (Jan for calendar year, specified for water year). Each season is designated by the calendar or water year in which it occurs. Calculates statistics from all values from complete years, unless specified. Returns a tibble with statistics.

**Usage**

```r
calc_annual_cumulative_stats(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  use_yield = FALSE,
  basin_area,
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  include_seasons = FALSE,
  transpose = FALSE,
  complete_years = FALSE
)
```
Arguments

data | Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates | Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values | Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups | Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number | Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
use_yield | Logical value indicating whether to calculate area-based water yield, in mm, instead of volumetric discharge. Default FALSE.
basin_area | Upstream drainage basin area, in square kilometres, to apply to observations. Three options:
(1) Leave blank if groups is STATION_NUMBER with HYDAT station numbers to extract basin areas from HYDAT.
(2) A single numeric value to apply to all observations.
(3) List each basin area for each group/station in groups (can override HYDAT value if listed) as such c("08NM116" = 795, "08NM242" = 10). If group is not listed the HYDAT area will be applied if it exists, otherwise it will be NA.
water_year_start | Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
start_year | Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.
end_year | Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.
exclude_years | Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.
months | Numeric vector of months to include in analysis (e.g. 6:8 for Jun-Aug). Default summarizes all months (1:12). If not all months, seasonal total yield and volumetric flows will not be included.
include_seasons | Logical value indication whether to include seasonal yields or volumetric discharges. Default TRUE.
transpose Logical value indicating whether to transpose rows and columns of results. Default FALSE.

complete_years Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

Value

A tibble data frame with the following columns, ending with '_Volume_m3' or '_Yield_mm' based on selection:

- **Year** calendar or water year selected
- **Total_*** annual (or selected months) total flow, in m3 or mm

Default seasonal columns:

- MMM-MMM_* first of two season total flows, in m3 or mm
- MMM-MMM_* second of two season total flows, in m3 or mm
- MMM-MMM_* first of four season total flows, in m3 or mm
- MMM-MMM_* second of four season total flows, in m3 or mm
- MMM-MMM_* third of four season total flows, in m3 or mm
- MMM-MMM_* fourth of four season total flows, in m3 or mm

Transposing data creates a column of 'Statistics' and subsequent columns for each year selected.

Examples

```r
# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

  # Calculate annual total volumetric flow statistics
  calc_annual_cumulative_stats(station_number = "08NM116")

  # Calculate annual total yield statistics with default HYDAT basin area
  calc_annual_cumulative_stats(station_number = "08NM116",
                               use_yield = TRUE)

  # Calculate annual total yield statistics with a custom basin area
  calc_annual_cumulative_stats(station_number = "08NM116",
                               use_yield = TRUE,
                               basin_area = 800,
                               start_year = 1980)
}
```
calc_annual_extremes  

Calculate annual high and low flows

Description

Calculates annual n-day minimum and maximum values, and the day of year and date of occurrence of daily flow values from a daily streamflow data set. Calculates statistics from all values, unless specified. Returns a tibble with statistics.

Usage

```r
calc_annual_extremes(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  roll_days = 1,
  roll_days_min = NA,
  roll_days_max = NA,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  months_min = NA,
  months_max = NA,
  transpose = FALSE,
  complete_years = FALSE,
  ignore_missing = FALSE,
  allowed_missing = ifelse(ignore_missing, 100, 0)
)
```

Arguments

- **data**: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
- **dates**: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
- **values**: Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
calc_annual_extremes

groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.

station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

roll_days Numeric value of the number of days to apply a rolling mean. Default 1.

roll_days_min Numeric value of the number of days to apply a rolling mean for low flows. Will override 'roll_days' argument for low flows. Default NA.

roll_days_max Numeric value of the number of days to apply a rolling mean for high flows. Will override 'roll_days' argument for high flows. Default NA.

roll_align Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.

water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

months_min Numeric vector of specified months for window of low flows (3 for March, 6:8 for Jun-Aug). Will override 'months' argument for low flows. Default NA.

months_max Numeric vector of specified months for window of high flows (3 for March, 6:8 for Jun-Aug). Will override 'months' argument for high flows. Default NA.

transpose Logical value indicating whether to transpose rows and columns of results. Default FALSE.

complete_years Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

ignore_missing Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

allowed_missing Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing
calc_annual_extremes

= FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing
= TRUE' then it defaults to 100 (any missing dates allowed); consistent with
ignore_missing usage. Supersedes ignore_missing when used.

Value

A tibble data frame with the following columns:

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>calendar or water year selected</td>
</tr>
<tr>
<td>Min_'n'_Day</td>
<td>annual minimum for selected n-day rolling mean, direction of mean specified by roll_align</td>
</tr>
<tr>
<td>Min_'n'_Day_DoY</td>
<td>day of year for selected annual minimum of n-day rolling mean</td>
</tr>
<tr>
<td>Min_'n'_Day_Date</td>
<td>date (YYYY-MM-DD) for selected annual minimum of n-day rolling mean</td>
</tr>
<tr>
<td>Max_'n'_Day</td>
<td>annual maximum for selected n-day rolling mean, direction of mean specified by roll_align</td>
</tr>
<tr>
<td>Max_'n'_Day_DoY</td>
<td>day of year for selected annual maximum of n-day rolling mean</td>
</tr>
<tr>
<td>Max_'n'_Day_Date</td>
<td>date (YYYY-MM-DD) for selected annual maximum of n-day rolling mean</td>
</tr>
</tbody>
</table>

Default columns:

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min_1_Day</td>
<td>annual 1-day mean minimum (roll_align = right)</td>
</tr>
<tr>
<td>Min_1_Day_DoY</td>
<td>day of year of annual 1-day mean minimum</td>
</tr>
<tr>
<td>Min_1_Day_Date</td>
<td>date (YYYY-MM-DD) of annual 1-day mean minimum</td>
</tr>
<tr>
<td>Max_1_Day</td>
<td>annual 1-day mean maximum (roll_align = right)</td>
</tr>
<tr>
<td>Max_1_Day_DoY</td>
<td>day of year of annual 1-day mean maximum</td>
</tr>
<tr>
<td>Max_1_Day_Date</td>
<td>date (YYYY-MM-DD) of annual 1-day mean maximum</td>
</tr>
</tbody>
</table>

Transposing data creates a column of 'Statistics' and subsequent columns for each year selected. 'Date' statistics not transposed.

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

  # Calculate annual 1-day (default) max/min flow data with
  # default alignment ('right')
  calc_annual_extremes(station_number = "08NM116")

  # Calculate custom 3-day max/min flow data with 'center' alignment
  calc_annual_extremes(station_number = "08NM116",
                       roll_days = 3,
                       roll_align = "center",
                       start_year = 1980)

}
calc_annual_flow_timing

Description

Calculates the timing (day of year and date) of portions of total annual flow of daily flow values from a daily streamflow data set. Calculates statistics from all values from complete years, unless specified. Returns a tibble with statistics.

Usage

calc_annual_flow_timing(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  percent_total = c(25, 33.3, 50, 75),
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  transpose = FALSE
)

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.

dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.

values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.

groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'.

station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT
calc_annual_flow_timing

database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

percent_total Numeric vector of percents of total annual flows to determine dates. Default c(25,33.3,50,75).

water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

transpose Logical value indicating whether to transpose rows and columns of results. Default FALSE.

Value

A tibble data frame with the following columns:

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>calendar or water year selected</td>
</tr>
<tr>
<td>DoY_’n’pct_TotalQ</td>
<td>day of year for each n-percent of total volumetric discharge</td>
</tr>
<tr>
<td>Date_’n’pct_TotalQ</td>
<td>date (YYYY-MM-DD) for each n-percent of total volumetric discharge</td>
</tr>
</tbody>
</table>

Default columns:

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DoY_25pct_TotalQ</td>
<td>day of year of 25-percent of total volumetric discharge</td>
</tr>
<tr>
<td>Date_25pct_TotalQ</td>
<td>date (YYYY-MM-DD) of 25-percent of total volumetric discharge</td>
</tr>
<tr>
<td>DoY_33.3pct_TotalQ</td>
<td>day of year of 33.3-percent of total volumetric discharge</td>
</tr>
<tr>
<td>Date_33.3pct_TotalQ</td>
<td>date (YYYY-MM-DD) of 33.3-percent of total volumetric discharge</td>
</tr>
<tr>
<td>DoY_50pct_TotalQ</td>
<td>day of year of 50-percent of total volumetric discharge</td>
</tr>
<tr>
<td>Date_50pct_TotalQ</td>
<td>date (YYYY-MM-DD) of 50-percent of total volumetric discharge</td>
</tr>
<tr>
<td>DoY_75pct_TotalQ</td>
<td>day of year of 75-percent of total volumetric discharge</td>
</tr>
<tr>
<td>Date_75pct_TotalQ</td>
<td>date (YYYY-MM-DD) of 75-percent of total volumetric discharge</td>
</tr>
</tbody>
</table>
Transposing data creates a column of ‘Statistics’ (just DoY, not Date values) and subsequent columns for each year selected.

References


Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Calculate annual flow timings with default percent of annual totals
calc_annual_flow_timing(station_number = "08NM116")

# Calculate annual flow timings with custom percent of annual totals
calc_annual_flow_timing(station_number = "08NM116",
                        percent_total = 50)
}

---

**calc_annual_highflows**  
*Calculate annual high flows and dates*

**Description**

Calculates annual n-day maximum values, and the day of year and date of occurrence of daily flow values from a daily streamflow data set. Calculates statistics from all values, unless specified. Returns a tibble with statistics.

**Usage**

calc_annual_highflows(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  roll_days = c(1, 3, 7, 30),
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  transpose = FALSE,
calc_annual_highflows

complete_years = FALSE,
ignore_missing = FALSE,
allowed_missing = ifelse(ignore_missing, 100, 0)
)

Arguments

data
Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates
Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values
Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups
Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number
Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
roll_days
Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align
Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
water_year_start
Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
start_year
Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.
end_year
Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.
exclude_years
Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.
months
Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).
transpose
Logical value indicating whether to transpose rows and columns of results. Default FALSE.
complete_years
Logical values indicating whether to include only years with complete data in analysis. Default FALSE.
ignore_missing Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

allowed_missing Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used.

Value

A tibble data frame with the following columns:

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>calendar or water year selected</td>
</tr>
<tr>
<td>Max_1_Day</td>
<td>annual 1-day mean maximum (roll_align = right)</td>
</tr>
<tr>
<td>Max_1_Day_DoY</td>
<td>day of year of annual 1-day mean maximum</td>
</tr>
<tr>
<td>Max_1_Day_Date</td>
<td>date (YYYY-MM-DD) of annual 1-day mean maximum</td>
</tr>
<tr>
<td>Max_3_Day</td>
<td>annual 3-day mean maximum (roll_align = right)</td>
</tr>
<tr>
<td>Max_3_Day_DoY</td>
<td>day of year of annual 3-day mean maximum</td>
</tr>
<tr>
<td>Max_3_Day_Date</td>
<td>date (YYYY-MM-DD) of annual 3-day mean maximum</td>
</tr>
<tr>
<td>Max_7_Day</td>
<td>annual 7-day mean maximum (roll_align = right)</td>
</tr>
<tr>
<td>Max_7_Day_DoY</td>
<td>day of year of annual 7-day mean maximum</td>
</tr>
<tr>
<td>Max_7_Day_Date</td>
<td>date (YYYY-MM-DD) of annual 7-day mean maximum</td>
</tr>
<tr>
<td>Max_30_Day</td>
<td>annual 30-day mean maximum (roll_align = right)</td>
</tr>
<tr>
<td>Max_30_Day_DoY</td>
<td>day of year of annual 30-day mean maximum</td>
</tr>
<tr>
<td>Max_30_Day_Date</td>
<td>date (YYYY-MM-DD) of annual 30-day mean maximum</td>
</tr>
</tbody>
</table>

Transposing data creates a column of 'Statistics' and subsequent columns for each year selected. 'Date' statistics not transposed.
**Examples**

```r
# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

  # Calculate annual 1, 3, 7, and 30-day (default) high flows with
  # default alignment ('right')
  calc_annual_highflows(station_number = "08NM116")

  # Calculate custom 3 and 7-day annual high flows with 'center' alignment
  calc_annual_highflows(station_number = "08NM116",
                        roll_days = c(3,7),
                        roll_align = "center",
                        start_year = 1980)
}
```

---

**calc_annual_lowflows**  
*Calculate annual low flows and dates*

**Description**

Calculates annual n-day minimum values, and the day of year and date of occurrence of daily flow values from a daily streamflow data set. Calculates statistics from all values, unless specified. Returns a tibble with statistics.

**Usage**

```r
calc_annual_lowflows(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  roll_days = c(1, 3, 7, 30),
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  transpose = FALSE,
  complete_years = FALSE,
  ignore_missing = FALSE,
  allowed_missing = ifelse(ignore_missing, 100, 0)
)
```
calc_annual_lowflows

Arguments

data  Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates  Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values  Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups  Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number  Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
roll_days  Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align  Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
water_year_start  Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
start_year  Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.
end_year  Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.
exclude_years  Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.
months  Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).
transpose  Logical value indicating whether to transpose rows and columns of results. Default FALSE.
complete_years  Logical values indicating whether to include only years with complete data in analysis. Default FALSE.
ignore_missing  Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.
allowed_missing

Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used.

Value

A tibble data frame with the following columns:

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>calendar or water year selected</td>
</tr>
<tr>
<td>Min_<code>n</code>_Day</td>
<td>annual minimum for each n-day rolling mean, direction of mean specified by roll_align</td>
</tr>
<tr>
<td>Min_<code>n</code>_Day_DoY</td>
<td>day of year for each annual minimum of n-day rolling mean</td>
</tr>
<tr>
<td>Min_<code>n</code>_Day_Date</td>
<td>date (YYYY-MM-DD) for each annual minimum of n-day rolling mean</td>
</tr>
</tbody>
</table>

Default columns:

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min_1_Day</td>
<td>annual 1-day mean minimum (roll_align = right)</td>
</tr>
<tr>
<td>Min_1_Day_DoY</td>
<td>day of year of annual 1-day mean minimum</td>
</tr>
<tr>
<td>Min_1_Day_Date</td>
<td>date (YYYY-MM-DD) of annual 1-day mean minimum</td>
</tr>
<tr>
<td>Min_3_Day</td>
<td>annual 3-day mean minimum (roll_align = right)</td>
</tr>
<tr>
<td>Min_3_Day_DoY</td>
<td>day of year of annual 3-day mean minimum</td>
</tr>
<tr>
<td>Min_3_Day_Date</td>
<td>date (YYYY-MM-DD) of annual 3-day mean minimum</td>
</tr>
<tr>
<td>Min_7_Day</td>
<td>annual 7-day mean minimum (roll_align = right)</td>
</tr>
<tr>
<td>Min_7_Day_DoY</td>
<td>day of year of annual 7-day mean minimum</td>
</tr>
<tr>
<td>Min_7_Day_Date</td>
<td>date (YYYY-MM-DD) of annual 7-day mean minimum</td>
</tr>
<tr>
<td>Min_30_Day</td>
<td>annual 30-day mean minimum (roll_align = right)</td>
</tr>
<tr>
<td>Min_30_Day_DoY</td>
<td>day of year of annual 30-day mean minimum</td>
</tr>
<tr>
<td>Min_30_Day_Date</td>
<td>date (YYYY-MM-DD) of annual 30-day mean minimum</td>
</tr>
</tbody>
</table>

Transposing data creates a column of 'Statistics' and subsequent columns for each year selected. 'Date' statistics not transposed.

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Calculate annual 1, 3, 7, and 30-day (default) low flows with
# default alignment ('right')
calc_annual_lowflows(station_number = "08NM116")
# Calculate custom 3 and 7-day annual low flows with 'center' alignment

```r
calc_annual_lowflows(station_number = "08NM116",
                     roll_days = c(3,7),
                     roll_align = "center",
                     start_year = 1980)
```

---

calc_annual_normal_days

*Calculate annual days above and below normal*

**Description**

Calculates the number of days per year outside of the 'normal' range (typically between 25 and 75th percentiles) for each day of the year. Upper and lower-range percentiles are calculated for each day of the year of from all years, and then each daily flow value for each year is compared. All days above or below the normal range are included. Analysis methodology is based on Environment and Climate Change Canada’s Water Quantity indicator from the Canadian Environmental Sustainability Indicators. Calculates statistics from all values from complete years, unless specified. Returns a tibble with statistics.

**Usage**

```r
calc_annual_normal_days(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  normal_percentiles = c(25, 75),
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  transpose = FALSE
)
```

**Arguments**

- **data**  
  Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
**dates**  
Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.

**values**  
Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.

**groups**  
Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.

**station_number**  
Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

**normal_percentiles**  
Numeric vector of two values, lower and upper percentiles, respectively indicating the limits of the normal range. Default c(25, 75).

**roll_days**  
Numeric value of the number of days to apply a rolling mean. Default 1.

**roll_align**  
Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.

**water_year_start**  
Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

**start_year**  
Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

**end_year**  
Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

**exclude_years**  
Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

**months**  
Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

**transpose**  
Logical value indicating whether to transpose rows and columns of results. Default FALSE.

**Value**  
A tibble data frame with the following columns:

<table>
<thead>
<tr>
<th>Year</th>
<th>calendar or water year selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below_Normal_Days</td>
<td>number of days per year below the daily normal (default 25th percentile)</td>
</tr>
<tr>
<td>Above_Normal_Days</td>
<td>number of days per year above the daily normal (default 75th percentile)</td>
</tr>
</tbody>
</table>
calc_annual_outside_normal

Days_Outside_Normal

number of days per year below and above the daily normal (default 25/75th percentile)

Transposing data creates a column of "Statistics" and subsequent columns for each year selected.

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Calculate statistics with default limits of normal (25 and 75th percentiles)
calc_annual_normal_days(station_number = "08NM116")

# Calculate statistics with custom limits of normal
calc_annual_normal_days(station_number = "08NM116",
                        normal_percentiles = c(10,90),
                        start_year = 1980)

}

calc_annual_outside_normal

*Calculate annual days above and below normal*

Description

This function has been superseded by the `calc_annual_normal_days()` function.

Calculates the number of days per year outside of the 'normal' range (typically between 25 and 75th percentiles) for each day of the year. Upper and lower-range percentiles are calculated for each day of the year of from all years, and then each daily flow value for each year is compared. All days above or below the normal range are included. Analysis methodology is based on Environment and Climate Change Canada’s Water Quantity indicator from the Canadian Environmental Sustainability Indicators. Calculates statistics from all values from complete years, unless specified. Returns a tibble with statistics.

Usage

```r
calc_annual_outside_normal(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  normal_percentiles = c(25, 75),
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
)```
calc_annual_outside_normal

start_year,
end_year,
exclude_years,
months = 1:12,
transpose = FALSE

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
normal_percentiles Numeric vector of two values, lower and upper percentiles, respectively indicating the limits of the normal range. Default c(25, 75).
roll_days Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
start_year Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.
end_year Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.
exclude_years Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.
months Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).
transpose Logical value indicating whether to transpose rows and columns of results. Default FALSE.

Value

A tibble data frame with the following columns:

- **Year**: calendar or water year selected
- **Days_Below_Normal**: number of days per year below the daily normal (default 25th percentile)
- **Days_Above_Normal**: number of days per year above the daily normal (default 75th percentile)
- **Days_Outside_Normal**: number of days per year below and above the daily normal (default 25/75th percentile)

Transposing data creates a column of "Statistics" and subsequent columns for each year selected.

Examples

```r
# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

  # Calculate statistics with default limits of normal (25 and 75th percentiles)
  calc_annual_outside_normal(station_number = "08NM116")

  # Calculate statistics with custom limits of normal
  calc_annual_outside_normal(station_number = "08NM116",
                            normal_percentiles = c(10,90))
}
```

**calc_annual_peaks**

Calculate annual high and low flows

Description

This function has been superseded by the `calc_annual_extremes()` function.

Calculates annual n-day minimum and maximum values, and the day of year and date of occurrence of daily flow values from a daily streamflow data set. Calculates statistics from all values, unless specified. Returns a tibble with statistics.
Usage

calc_annual_peaks(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  roll_days = 1,
  roll_days_low = NA,
  roll_days_high = NA,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  months_low = NA,
  months_high = NA,
  transpose = FALSE,
  complete_years = FALSE,
  ignore_missing = FALSE,
  allowed_missing = ifelse(ignore_missing, 100, 0)
)

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
roll_days Numeric value of the number of days to apply a rolling mean. Default 1.
roll_days_low Numeric value of the number of days to apply a rolling mean for low flows. Will override 'roll_days' argument for low flows. Default NA.
### `calc_annual_peaks`

- **`roll_days_high`**: Numeric value of the number of days to apply a rolling mean for high flows. Will override 'roll_days' argument for high flows. Default `NA`.
- **`roll_align`**: Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
- **`water_year_start`**: Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
- **`start_year`**: Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.
- **`end_year`**: Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.
- **`exclude_years`**: Numeric vector of years to exclude from analysis. Leave blank or set to `NULL` to include all years.
- **`months`**: Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or `c(10:12,1)` for first four months (Oct-Jan) when `water_year_start = 10` (Oct). Default summarizes all months (1:12).
- **`months_low`**: Numeric vector of specified months for window of low flows (3 for March, 6:8 for Jun-Aug). Will override 'months' argument for low flows. Default `NA`.
- **`months_high`**: Numeric vector of specified months for window of high flows (3 for March, 6:8 for Jun-Aug). Will override 'months' argument for high flows. Default `NA`.
- **`transpose`**: Logical value indicating whether to transpose rows and columns of results. Default `FALSE`.
- **`complete_years`**: Logical values indicating whether to include only years with complete data in analysis. Default `FALSE`.
- **`ignore_missing`**: Logical value indicating whether dates with missing values should be included in the calculation. If `TRUE` then a statistic will be calculated regardless of missing dates. If `FALSE` then only those statistics from time periods with no missing dates will be returned. Default `FALSE`.
- **`allowed_missing`**: Numeric value between 0 and 100 indicating the **percentage** of missing dates allowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with `ignore_missing` usage. Supersedes `ignore_missing` when used.

### Value

A tibble data frame with the following columns:

| Year | calendar or water year selected |
| Min_'n'_Day | annual minimum for selected n-day rolling mean, direction of mean specified by `roll_align` |
| Min_'n'_Day_DoY | day of year for selected annual minimum of n-day rolling mean |
### Description

Calculates means, medians, maximums, minimums, and percentiles for each year from all years of a daily streamflow data set. Calculates statistics from all values, unless specified. Returns a tibble with statistics.

---

**Example**

```r
# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

    # Calculate annual 1-day (default) peak flow data with
    # default alignment ('right')
calc_annual_peaks(station_number = "08NM116")

    # Calculate custom 3-day peak flow data with 'center' alignment
calc_annual_peaks(station_number = "08NM116",
                     roll_days = 3,
                     roll_align = "center")
}
```
Usage

calc_annual_stats(
    data,
    dates = Date,
    values = Value,
    groups = STATION_NUMBER,
    station_number,
    roll_days = 1,
    roll_align = "right",
    percentiles = c(10, 90),
    water_year_start = 1,
    start_year,
    end_year,
    exclude_years,
    months = 1:12,
    transpose = FALSE,
    complete_years = FALSE,
    ignore_missing = FALSE,
    allowed_missing = ifelse(ignore_missing, 100, 0)
)

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
roll_days Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
percentiles Numeric vector of percentiles to calculate. Set to NA if none required. Default c(10,90).
calc_annual_stats

water_year_start  Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year  Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year  Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years  Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months  Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

transpose  Logical value indicating whether to transpose rows and columns of results. Default FALSE.

complete_years  Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

ignore_missing  Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

allowed_missing  Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing usage. Supersedes ignore_missing when used.

Value

A tibble data frame with the following columns:

<table>
<thead>
<tr>
<th>Year</th>
<th>calendar or water year selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>annual mean of all daily flows for a given year</td>
</tr>
<tr>
<td>Median</td>
<td>annual median of all daily flows for a given year</td>
</tr>
<tr>
<td>Maximum</td>
<td>annual maximum of all daily flows for a given year</td>
</tr>
<tr>
<td>Minimum</td>
<td>annual minimum of all daily flows for a given year</td>
</tr>
<tr>
<td>P’n’</td>
<td>each annual n-th percentile selected of all daily flows</td>
</tr>
</tbody>
</table>

Default percentile columns:

| P10 | annual 10th percentile of all daily flows for a given year |
| P90 | annual 90th percentile of all daily flows for a given year |

Transposing data creates a column of "Statistics" and subsequent columns for each year selected.
**Examples**

```r
# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

    # Calculate annual statistics from a data frame using the data argument
    flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
    calc_annual_stats(data = flow_data)

    # Calculate annual statistics using station_number argument
    calc_annual_stats(station_number = "08NM116")

    # Calculate annual statistics regardless if there
    # is missing data for a given year
    calc_annual_stats(station_number = "08NM116",
                      ignore_missing = TRUE)

    # Calculate annual statistics for water years starting in October
    calc_annual_stats(station_number = "08NM116",
                      water_year_start = 10)

    # Calculate annual statistics for 7-day flows for July-September
    # months only, with 25 and 75th percentiles
    calc_annual_stats(station_number = "08NM116",
                      roll_days = 7,
                      months = 7:9,
                      percentiles = c(25,75))
}
```

---

**calc_daily_cumulative_stats**

*Calculate cumulative daily flow statistics*

**Description**

Calculate cumulative daily flow statistics for each day of the year of daily flow values from a daily streamflow data set. Defaults to volumetric cumulative flows, can use use_yield and basin_area to convert to area-based water yield. Calculates statistics from all values from all complete years, unless specified. Returns a tibble with statistics.

**Usage**

```r
calc_daily_cumulative_stats(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  percentiles = c(5, 25, 75, 95),
)```
use_yield = FALSE,
basin_area,
water_year_start = 1,
start_year,
end_year,
exclude_years,
months = 1:12,
transpose = FALSE
)

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.

station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

percentiles Numeric vector of percentiles to calculate. Set to NA if none required. Default c(5, 25, 75, 95).

use_yield Logical value indicating whether to calculate area-based water yield, in mm, instead of volumetric discharge. Default FALSE.

basin_area Upstream drainage basin area, in square kilometres, to apply to observations. Three options:
(1) Leave blank if groups is STATION_NUMBER with HYDAT station numbers to extract basin areas from HYDAT.
(2) A single numeric value to apply to all observations.
(3) List each basin area for each group/station in groups (can override HYDAT value if listed) as such c("08NM116" = 795, "08NM242" = 10). If group is not listed the HYDAT area will be applied if it exists, otherwise it will be NA.

water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
calc_daily_cumulative_stats

start_year Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

day_numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12). Need to be consecutive months for given year/water year to work properly.

transpose Logical value indicating whether to transpose rows and columns of results. Default FALSE.

Value

A data frame with the following columns, default units in cubic metres, millimetres if use_yield and basin_area provided:

- Date date (MMM-DD) of daily cumulative statistics
- DayofYear day of year of daily cumulative statistics
- Mean daily mean of all cumulative flows for a given day of the year
- Median daily mean of all cumulative flows for a given day of the year
- Maximum daily mean of all cumulative flows for a given day of the year
- Minimum daily mean of all cumulative flows for a given day of the year
- P’n’ each daily n-th percentile selected of all cumulative flows for a given day of the year

Default percentile columns:

- P5 daily 5th percentile of all cumulative flows for a given day of the year
- P25 daily 25th percentile of all cumulative flows for a given day of the year
- P75 daily 75th percentile of all cumulative flows for a given day of the year
- P95 daily 95th percentile of all cumulative flows for a given day of the year

Transposing data creates a column of "Statistics" and subsequent columns for each year selected.

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Calculate annual daily cumulative yield statistics
# with default HYDAT basin area
calc_daily_cumulative_stats(station_number = "08NM116",
use_yield = TRUE)

# Calculate annual daily cumulative yield statistics
calc_daily_stats  

Description

Calculates means, medians, maximums, minimums, and percentiles for each day of the year of flow values from a daily streamflow data set. Can determine statistics of rolling mean days (e.g. 7-day flows) using the roll_days argument. Note that statistics are based on the numeric days of year (1-365) and not the date of year (Jan 1 - Dec 31). Calculates statistics from all values, unless specified. Returns a tibble with statistics.

Usage

```
calc_daily_stats(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number, 
  percentiles = c(5, 25, 75, 95),
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year, 
  end_year, 
  exclude_years,
  months = 1:12,
  transpose = FALSE,
  complete_years = FALSE,
  ignore_missing = FALSE
)
```

Arguments

- **data**: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
- **dates**: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
calc_daily_stats

values: Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.

groups: Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.

station_number: Character string vector of seven digit Water Survey of Canada station numbers (e.g. '08NM116') of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

percentiles: Numeric vector of percentiles to calculate. Set to NA if none required. Default c(5,25,75,95).

roll_days: Numeric value of the number of days to apply a rolling mean. Default 1.

roll_align: Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.

water_year_start: Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year: Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year: Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years: Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months: Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

transpose: Logical value indicating whether to transpose rows and columns of results. Default FALSE.

complete_years: Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

ignore_missing: Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

Value

A tibble data frame with the following columns:

Date: date (MMM-DD) of daily statistics
DayofYear: day of year of daily statistics
Mean
daily mean of all flows for a given day of the year
Median
daily mean of all flows for a given day of the year
Maximum
daily mean of all flows for a given day of the year
Minimum
daily mean of all flows for a given day of the year
P’n’
each daily n-th percentile selected of all flows for a given day of the year

Default percentile columns:
P5
daily 5th percentile of all flows for a given day of the year
P25
daily 25th percentile of all flows for a given day of the year
P75
daily 75th percentile of all flows for a given day of the year
P95
daily 95th percentile of all flows for a given day of the year

Transposing data creates a column of "Statistics" and subsequent columns for each year selected.

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Calculate daily statistics using station_number argument with defaults
calc_daily_stats(station_number = "08NM116",
start_year = 1980)

# Calculate daily statistics regardless if there is missing data for a given day of year
calc_daily_stats(station_number = "08NM116",
ignore_missing = TRUE)

# Calculate daily statistics using only years with no missing data
calc_daily_stats(station_number = "08NM116",
complete_years = TRUE)

# Calculate daily statistics for water years starting in October between 1980 and 2010
calc_daily_stats(station_number = "08NM116",
start_year = 1980,
end_year = 2010,
water_year_start = 10)
}

### Examples

**Description**

Calculates the percentile rank of a discharge value compared to all flow values of a streamflow data set. Looks up the value in the distribution (stats::ecdf() function) of all daily discharge values from all years, unless specified. Returns a tibble with statistics.
Usage

calc_flow_percentile(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  roll_days = 1,
  roll_align = "right",
  flow_value,
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  complete_years = FALSE,
  months = 1:12
)

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
roll_days Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
flow_value A numeric flow value of which to determine the percentile rank. Required.
water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
calc_longterm_daily_stats

Description

Calculates the long-term mean, median, maximum, minimum, and percentiles of daily flow values for over all months and all data (Long-term) from a daily streamflow data set. Calculates statistics from all values, unless specified. Returns a tibble with statistics.

Example:

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

  # Calculate the percentile rank of a 10-cms flow value from a full record
  calc_flow_percentile(station_number = "08NM116",
                      flow_value = 10)

  # Calculate the percentile rank of a 10-cms flow value from years with no missing data
  calc_flow_percentile(station_number = "08NM116",
                      complete_years = TRUE,
                      flow_value = 10)

  # Calculate the percentile rank of a 10-cms flow value for June from years with no missing data
  calc_flow_percentile(station_number = "08NM116",
                      complete_years = TRUE,
                      months = 6,
                      flow_value = 10)

}
Usage

calc_longterm_daily_stats(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  percentiles = c(10, 90),
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  complete_years = FALSE,
  include_longterm = TRUE,
  custom_months,
  custom_months_label,
  transpose = FALSE,
  ignore_missing = FALSE
)

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
percentiles Numeric vector of percentiles to calculate. Set to NA if none required. Default c(10, 90).
roll_days Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align  Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.

water_year_start  Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year  Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

don_end  Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years  Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months  Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

custom_months  Numeric vector of months to combine to summarize (ex. 6:8 for Jun-Aug). Adds results to the end of table. If wanting months that overlap calendar years (ex. Oct-Mar), choose water_year_start that begins before the first month listed. Leave blank for no custom month summary.

custom_months_label  Character string to label custom months. For example, if months = 7:9 you may choose "Summer" or "Jul-Sep". Default "Custom-Months".

transpose  Logical value indicating whether to transpose rows and columns of results. Default FALSE.

ignore_missing  Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

Value

A tibble data frame with the following columns:

- **Month**: month of the year, included 'Long-term' for all months, and 'Custom-Months' if selected
- **Mean**: mean of all daily data for a given month and long-term over all years
- **Median**: median of all daily data for a given month and long-term over all years
- **Maximum**: maximum of all daily data for a given month and long-term over all years
- **Minimum**: minimum of all daily data for a given month and long-term over all years
- **P'n'**: each n-th percentile selected for a given month and long-term over all years
Default percentile columns:

- **P10**: annual 10th percentile selected for a given month and long-term over all years
- **P90**: annual 90th percentile selected for a given month and long-term over all years

Transposing data creates a column of "Statistics" and subsequent columns for each year selected.

**Examples**

```r
# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Calculate long-term statistics using data argument with defaults
flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
calc_longterm_daily_stats(data = flow_data,  
                         start_year = 1980)

# Calculate long-term statistics using station_number argument with defaults
calc_longterm_daily_stats(station_number = "08NM116",  
                         start_year = 1980)

# Calculate long-term statistics regardless if there is missing data for a given year
calc_longterm_daily_stats(station_number = "08NM116",  
                         ignore_missing = TRUE)

# Calculate long-term statistics for water years starting in October
calc_longterm_daily_stats(station_number = "08NM116",  
                         start_year = 1980,  
                         water_year_start = 10)

# Calculate long-term statistics with custom years and percentiles
calc_longterm_daily_stats(station_number = "08NM116",  
                         start_year = 1981,  
                         end_year = 2010,  
                         percentiles = c(25, 75))

# Calculate long-term statistics and add custom stats for July-September
calc_longterm_daily_stats(station_number = "08NM116",  
                         start_year = 1980,  
                         custom_months = 7:9,  
                         custom_months_label = "Summer")
}
```

**calc_longterm_mean**

*Calculate the long-term mean annual discharge*

**Description**

Calculates the long-term mean annual discharge (MAD) from a daily streamflow data set. Calculates statistics from all values, unless specified. Returns a tibble with statistics.
Usage

calc_longterm_mean(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  complete_years = FALSE,
  months = 1:12,
  percent_MAD,
  transpose = FALSE
)

Arguments

data
Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates
Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values
Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups
Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number
Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
roll_days
Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align
Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
water_year_start
Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
calc_longterm_monthly_stats

Description

Calculates the long-term mean, median, maximum, minimum, and percentiles of annual monthly mean flow values for all months and all data (Long-term) from a daily streamflow data set. Calculates statistics from all values, unless specified. Returns a tibble with statistics.
calc_longterm_monthly_stats

Usage

calc_longterm_monthly_stats(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  percentiles = c(10, 90),
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  complete_years = FALSE,
  include_annual = TRUE,
  custom_months,
  custom_months_label,
  transpose = FALSE,
  ignore_missing = FALSE
)

Arguments

data          Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates         Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values        Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups        Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
percentiles   Numeric vector of percentiles to calculate. Set to NA if none required. Default c(10, 90).
roll_days     Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.

water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

complete_years Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

include_annual Logical value indicating whether to include annual calculation of all months. Default TRUE.

custom_months Numeric vector of months to combine to summarize (ex. 6:8 for Jun-Aug). Adds results to the end of table. If wanting months that overlap calendar years (ex. Oct-Mar), choose water_year_start that begins before the first month listed. Leave blank for no custom month summary.

custom_months_label Character string to label custom months. For example, if months = 7:9 you may choose "Summer" or "Jul-Sep". Default "Custom-Months".

transpose Logical value indicating whether to transpose rows and columns of results. Default FALSE.

ignore_missing Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

Value

A tibble data frame with the following columns:

<table>
<thead>
<tr>
<th>Month</th>
<th>month of the year, included 'Annual' for all months, and 'Custom-Months' if selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>mean of all annual monthly means for a given month over all years</td>
</tr>
<tr>
<td>Median</td>
<td>median of all annual monthly means for a given month over all years</td>
</tr>
<tr>
<td>Maximum</td>
<td>maximum of all annual monthly means for a given month over all years</td>
</tr>
<tr>
<td>Minimum</td>
<td>minimum of all annual monthly means for a given month over all years</td>
</tr>
<tr>
<td>P'n'</td>
<td>each n-th percentile selected for annual monthly means for a given month over all years</td>
</tr>
</tbody>
</table>
**calc_longterm_percentile**

Calculates the long-term percentiles from a daily streamflow data set. Calculates statistics from all values, unless specified. Returns a tibble with statistics.

### Description

Calculates the long-term percentiles from a daily streamflow data set. Calculates statistics from all values, unless specified. Returns a tibble with statistics.

### Usage

```r
calc_longterm_percentile(
  data,
  dates = Date,
)```

### Examples

```r
# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

  # Calculate long-term monthly statistics using data argument with defaults
  flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
  calc_longterm_monthly_stats(data = flow_data,
                              start_year = 1980)

  # Calculate long-term monthly statistics using station_number argument with defaults
  calc_longterm_monthly_stats(station_number = "08NM116",
                              start_year = 1980)

  # Calculate long-term monthly statistics regardless if there is missing data for a given year
  calc_longterm_monthly_stats(station_number = "08NM116",
                              start_year = 1980,
                              ignore_missing = TRUE)

  # Calculate long-term monthly statistics and add custom stats for July-September
  calc_longterm_monthly_stats(station_number = "08NM116",
                              start_year = 1980,
                              custom_months = 7:9,
                              custom_months_label = "Summer")
}
```
values = Value,
groups = STATION_NUMBER,
station_number,
percentiles,
roll_days = 1,
roll_align = "right",
water_year_start = 1,
start_year,
end_year,
exclude_years,
complete_years = FALSE,
months = 1:12,
transpose = FALSE
)

Arguments

- **data**
  Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using `station_number` argument.

- **dates**
  Name of column in `data` that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using `station_number` argument.

- **values**
  Name of column in `data` that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using `station_number` argument.

- **groups**
  Name of column in `data` that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using `station_number` argument.

- **station_number**
  Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

- **percentiles**
  Numeric vector of percentiles (ex. c(5,10,25,75)) to calculate. Required.

- **roll_days**
  Numeric value of the number of days to apply a rolling mean. Default 1.

- **roll_align**
  Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.

- **water_year_start**
  Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

- **start_year**
  Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.
**Description**

Calculate cumulative monthly flow statistics for each month of the year of daily flow values from a daily streamflow data set. Calculates statistics from all values from complete years, unless specified. Defaults to volumetric cumulative flows, can use `use_yield` and `basin_area` to convert to area-based water yield. Returns a tibble with statistics.

**Value**

A tibble data frame of a long-term percentile of selected years and months.

**Examples**

```r
# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

  # Calculate the 20th percentile flow value from a flow record
calc_longterm_percentile(station_number = "08NM116",
                          percentile = 20)

  # Calculate the 90th percentile flow value with custom years
  calc_longterm_percentile(station_number = "08NM116",
                            start_year = 1980,
                            end_year = 2010,
                            percentile = 90)
}
```
Usage

calc_monthly_cumulative_stats(
    data,
    dates = Date,
    values = Value,
    groups = STATION_NUMBER,
    station_number,
    percentiles = c(5, 25, 75, 95),
    use_yield = FALSE,
    basin_area,
    water_year_start = 1,
    start_year,
    end_year,
    exclude_years,
    months = 1:12,
    transpose = FALSE
)

Arguments

data
    Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.

dates
    Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.

values
    Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.

groups
    Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.

station_number
    Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

percentiles
    Numeric vector of percentiles to calculate. Set to NA if none required. Default c(5, 25, 75, 95).

use_yield
    Logical value indicating whether to calculate area-based water yield, in mm, instead of volumetric discharge. Default FALSE.

basin_area
    Upstream drainage basin area, in square kilometres, to apply to observations. Three options:
    (1) Leave blank if groups is STATION_NUMBER with HYDAT station numbers to extract basin areas from HYDAT.
(2) A single numeric value to apply to all observations.

(3) List each basin area for each group/station in groups (can override HYDAT value if listed) as such c("08NM116" = 795, "08NM242" = 10). If group is not listed the HYDAT area will be applied if it exists, otherwise it will be NA.

water_year_start
Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year
Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year
Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years
Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months
Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12). Need to be consecutive months for given year/water year to work properly.

transpose
Logical value indicating whether to transpose rows and columns of results. Default FALSE.

Value
A tibble data frame with the following columns, default units in cubic metres, or millimetres if use_yield and basin_area provided:

Month
month (MMM-DD) of cumulative statistics

Mean
monthly mean of all cumulative flows for a given month of the year

Median
monthly mean of all cumulative flows for a given month of the year

Maximum
monthly mean of all cumulative flows for a given month of the year

Minimum
monthly mean of all cumulative flows for a given month of the year

P'n'
each monthly n-th percentile selected of all cumulative flows for a given month of the year

Default percentile columns:

P5
monthly 5th percentile of all cumulative flows for a given month of the year

P25
monthly 25th percentile of all cumulative flows for a given month of the year

P75
monthly 75th percentile of all cumulative flows for a given month of the year

P95
monthly 95th percentile of all cumulative flows for a given month of the year

Transposing data creates a column of "Statistics" and subsequent columns for each year selected.
Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Calculate annual monthly cumulative volume statistics
calc_monthly_cumulative_stats(station_number = "08NM116")

# Calculate annual monthly cumulative volume statistics with default HYDAT basin area
calc_monthly_cumulative_stats(station_number = "08NM116",
                           use_yield = TRUE)

# Calculate annual monthly cumulative volume statistics with custom basin area
calc_monthly_cumulative_stats(station_number = "08NM116",
                           use_yield = TRUE,
                           basin_area = 800)
}

calc_monthly_stats Calculate monthly summary statistics

Description

Calculates means, medians, maximums, minimums, and percentiles for each month of all years of
flow values from a daily streamflow data set. Calculates statistics from all values, unless specified.
Returns a tibble with statistics.

Usage
calc_monthly_stats(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  percentiles = c(10, 90),
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  transpose = FALSE,
  spread = FALSE,
  complete_years = FALSE,
  ignore_missing = FALSE,
  allowed_missing = ifelse(ignore_missing, 100, 0)
)
Arguments

data: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values: Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups: Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number: Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
percentiles: Numeric vector of percentiles to calculate. Set to NA if none required. Default c(10,90).
roll_days: Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align: Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
water_year_start: Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
start_year: Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.
end_year: Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.
exclude_years: Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.
months: Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).
transpose: Logical value indicating if each month statistic should be individual rows. Default FALSE.
spread: Logical value indicating if each month statistic should be the column name. Default FALSE.
complete_years: Logical values indicating whether to include only years with complete data in analysis. Default FALSE.
ignore_missing Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

allowed_missing Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used.

Value

A tibble data frame with the following columns:

- **Year**: calendar or water year selected
- **Month**: month of the year
- **Mean**: mean of all daily flows for a given month and year
- **Median**: median of all daily flows for a given month and year
- **Maximum**: maximum of all daily flows for a given month and year
- **Minimum**: minimum of all daily flows for a given month and year
- **P’n’**: each n-th percentile selected for a given month and year

Default percentile columns:

- **P10**: 10th percentile of all daily flows for a given month and year
- **P90**: 90th percentile of all daily flows for a given month and year

Transposing data creates a column of 'Statistics' for each month, labeled as 'Month-Statistic' (ex "Jan-Mean"), and subsequent columns for each year selected. Spreading data creates columns of Year and subsequent columns of Month-Statistics (ex 'Jan-Mean').

Examples

```r
# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Calculate statistics using a data frame and data argument with defaults
flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
calc_monthly_stats(data = flow_data,
                   start_year = 1980)

# Calculate statistics using station_number argument with defaults
calc_monthly_stats(station_number = "08NM116",
                   start_year = 1980)

# Calculate statistics regardless if there is missing data for a given year
calc_monthly_stats(station_number = "08NM116",
                   ignore_missing = TRUE)
```
# Calculate statistics for water years starting in October
calc_monthly_stats(station_number = "08NM116",
                     start_year = 1980,
                     water_year_start = 10)

# Calculate statistics with custom years and percentiles
calc_monthly_stats(station_number = "08NM116",
                     start_year = 1981,
                     end_year = 2010,
                     percentiles = c(25,75))

compute_annual_frequencies

Perform an annual low or high-flow frequency analysis

Description

Performs a flow volume frequency analysis on annual statistics from a daily streamflow data set. Defaults to a low flow frequency analysis using annual minimums. Set use_max = TRUE for annual high flow frequency analyses. Calculates statistics from all values, unless specified. Function will calculate using all values in `Values` column (no grouped analysis). Analysis methodology replicates that from HEC-SSP. Returns a list of tibbles and plots.

Usage

compute_annual_frequencies(
    data,
    dates = Date,
    values = Value,
    station_number,
    roll_days = c(1, 3, 7, 30),
    roll_align = "right",
    use_max = FALSE,
    use_log = FALSE,
    prob_plot_position = c("weibull", "median", "hazen"),
    prob_scale_points = c(0.9999, 0.999, 0.99, 0.9, 0.95, 0.2, 0.1, 0.02, 0.01, 0.001, 1e-04),
    fit_distr = c("PIII", "weibull"),
    fit_distr_method = ifelse(fit_distr == "PIII", "MOM", "MLE"),
    fit_quantiles = c(0.975, 0.99, 0.98, 0.95, 0.9, 0.8, 0.5, 0.2, 0.1, 0.05, 0.01),
    plot_curve = TRUE,
    water_year_start = 1,
    start_year,
    end_year,
compute_annual_frequencies

exclude_years,
months = 1:12,
complete_years = FALSE,
ignore_missing = FALSE,
allowed_missing = ifelse(ignore_missing, 100, 0)
)

Arguments

data
A data frame of daily data that contains columns of dates and flow values. Groupings and the groups argument are not used for this function (i.e. station numbers). Leave blank or set to NULL if using station_number argument.

dates
Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.

values
Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.

station_number
Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

roll_days
Numeric value of the number of days to apply a rolling mean. Default 1.

roll_align
Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.

use_max
Logical value to indicate using maximums rather than the minimums for analysis. Default FALSE.

use_log
Logical value to indicate log-scale transforming of flow data before analysis. Default FALSE.

prob_plot_position
Character string indicating the plotting positions used in the frequency plots, one of 'weibull', 'median', or 'hazen'. Points are plotted against (i-a)/(n+1-a-b) where i is the rank of the value; n is the sample size and a and b are defined as: (a=0, b=0) for Weibull plotting positions; (a=.2; b=.3) for Median plotting positions; and (a=.5; b=.5) for Hazen plotting positions. Default 'weibull'.

prob_scale_points
Numeric vector of probabilities to be plotted along the X axis in the frequency plot. Inverse of return period. Default c(.9999, .999, .99, .9, .5, .2, .1, .02, .01, .001, .0001).

fit_distr
Character string identifying the distribution to fit annual data, one of 'PIII' (Log Pearson Type III) or 'weibull' (Weibull) distributions. Default 'PIII'.

fit_distr_method
Character string identifying the method used to fit the distribution, one of 'MOM' (method of moments) or 'MLE' (maximum likelihood estimation). Selected as 'MOM' if fit_distr = 'PIII' (default) or 'MLE' if fit_distr = 'weibull'.

fit_quantiles Numeric vector of quantiles to be estimated from the fitted distribution. Default c(.975, .99, .98, .95, .90, .80, .50, .20, .10, .05, .01).

plot_curve Logical value to indicate plotting the computed curve on the probability plot. Default TRUE.

water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

complete_years Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

ignore_missing Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

allowed_missing Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used.

Value

A list with the following elements:

Freq_Analysis_Data Data frame with computed annual summary statistics used in analysis.

Freq_Plot_Data Data frame with co-ordinates used in frequency plot.

Freq_Plot ggplot2 object with frequency plot.

Freq_Fitting List of fitted objects from fitdistrplus.

Freq_Fitted_Quantiles Data frame with fitted quantiles.

See Also

compute_frequency_analysis
Examples

```r
## Not run:

# Working examples (see arguments for further analysis options):

# Compute an annual frequency analysis using default arguments
results <- compute_annual_frequencies(station_number = "08NM116",
                        start_year = 1980,
                        end_year = 2010)

# Compute an annual frequency analysis using default arguments (as listed)
results <- compute_annual_frequencies(station_number = "08NM116",
                        roll_days = c(1,3,7,30),
                        start_year = 1980,
                        end_year = 2010,
                        prob_plot_position = "weibull",
                        prob_scale_points = c(.9999, .999, .99, .9, .5,
                                                .2, .1, .02, .01, .001, .0001),
                        fit_distr = "PIII",
                        fit_distr_method = "MOM")

# Compute a 7-day annual frequency analysis with "median" plotting positions
# and fitting the data to a weibull distribution (not default PIII)
results <- compute_annual_frequencies(station_number = "08NM116",
                        roll_days = 7,
                        start_year = 1980,
                        end_year = 2010,
                        prob_plot_position = "median",
                        fit_distr = "weibull")

## End(Not run)
```

compute_annual_trends

**Calculate prewhitened nonlinear annual trends on streamflow data**

**Description**

Calculates prewhitened nonlinear trends on annual streamflow data. Uses the `zyp` package to calculate trends. Review `zyp` for more information. Calculates statistics from all values, unless specified. Returns a list of tibbles and plots. All annual statistics calculated using the `calc_all_annual_stats()` function which uses the following `fasstr` functions:

- `calc_annual_stats()`
- `calc_annual_lowflows()`
- `calc_annual_cumulative_stats()`
- `calc_annual_flow_timing()`
- `calc_monthly_stats()`
- `calc_annual_normal_days()`
Usage

```r
compute_annual_trends(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  zyp_method,
  basin_area,
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  annual_percentiles = c(10, 90),
  monthly_percentiles = c(10, 20),
  stats_days = 1,
  stats_align = "right",
  lowflow_days = c(1, 3, 7, 30),
  lowflow_align = "right",
  timing_percent = c(25, 33, 50, 75),
  normal_percentiles = c(25, 75),
  complete_years = FALSE,
  ignore_missing = FALSE,
  allowed_missing_annual = ifelse(ignore_missing, 100, 0),
  allowed_missing_monthly = ifelse(ignore_missing, 100, 0),
  include_plots = TRUE,
  zyp_alpha
)
```

Arguments

- **data**: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using `station_number` argument.
- **dates**: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using `station_number` argument.
- **values**: Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using `station_number` argument.
- **groups**: Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using `station_number` argument.
**compute_annual_trends**

**station_number**  Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

**zyp_method**  Character string identifying the prewhitened trend method to use from zyp, either 'zhang' or 'yuepilon'. 'zhang' is recommended over 'yuepilon' for hydrologic applications (Bürger 2017; Zhang and Zwiers 2004). Required.

**basin_area**  Upstream drainage basin area, in square kilometres, to apply to observations. Three options:

1. Leave blank if groups is STATION_NUMBER with HYDAT station numbers to extract basin areas from HYDAT.
2. A single numeric value to apply to all observations.
3. List each basin area for each group/station in groups (can override HYDAT value if listed) as such c("08NM116" = 795,"08NM242" = 10). If group is not listed the HYDAT area will be applied if it exists, otherwise it will be NA.

**water_year_start**  Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

**start_year**  Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

**end_year**  Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

**exclude_years**  Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

**months**  Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12). If not all months, seasonal total yield and volumetric flows will not be included.

**annual_percentiles**  Numeric vector of percentiles to calculate annually. Set to NA if none required. Used for calc_annual_stats() function. Default c(10,90).

**monthly_percentiles**  Numeric vector of percentiles to calculate monthly for each year. Set to NA if none required. Used for calc_monthly_stats() function. Default c(10,20).

**stats_days**  Numeric vector of the number of days to apply a rolling mean on basic stats. Default c(1). Used for calc_annual_stats() and calc_monthly_stats() functions.

**stats_align**  Character string identifying the direction of the rolling mean on basic stats from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'. Used for calc_annual_stats(), calc_monthly_stats(), and calc_annual_normal_days() functions.

**lowflow_days**  Numeric vector of the number of days to apply a rolling mean on low flow stats. Default c(1,3,7,30). Used for calc_lowflow_stats() function.
compute_annual_trends

lowflow_align Character string identifying the direction of the rolling mean on low flow stats from the specified date, either by the first (‘left’), last (‘right’), or middle (‘center’) day of the rolling n-day group of observations. Default ‘right’. Used for calc_lowflow_stats() function.

timing_percent Numeric vector of percents of annual total flows to determine dates. Used for calc_annual_flow_timing() function. Default c(25,33.3,50,75).

normal_percentiles Numeric vector of two values, lower and upper percentiles, respectively indicating the limits of the normal range. Default c(25,75).

complete_years Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

ignore_missing Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

allowed_missing_annual Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate an annual statistic (0 to 100 percent). If ‘ignore_missing = FALSE’ then it defaults to 0 (zero missing dates allowed), if ‘ignore_missing = TRUE’ then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used. Only for annual means, percentiles, minimums, and maximums.

allowed_missing_monthly Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a monthly statistic (0 to 100 percent). If ‘ignore_missing = FALSE’ then it defaults to 0 (zero missing dates allowed), if ‘ignore_missing = TRUE’ then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used. Only for monthly means, percentiles, minimums, and maximums.

include_plots Logical value indicating if annual trending plots should be included. Default TRUE.

zyp_alpha Numeric value of the significance level (ex. 0.05) of when to plot a trend line. Leave blank for no line.

Value

A list of tibbles and optional plots from the trending analysis including:

Annual_Trends_Data a tibble of the annual statistics used for trending

Annual_Trends_Results a tibble of the results of the zyp trending analysis

Annual_* each ggplot2 object for each annual trended statistic

References

References:
compute_annual_trends


See Also

zyppackage, calc_all_annual_stats

Examples

```r
## Not run:
# Working examples:

# Compute trends statistics using a data frame and data argument with defaults
flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
trends <- compute_annual_trends(data = flow_data,
                               zyp_method = "zhang")

# Compute trends statistics using station_number with defaults
trends <- compute_annual_trends(station_number = "08NM116",
                               zyp_method = "zhang")

# Compute trends statistics and plot a trend line if the significance is less than 0.05
trends <- compute_annual_trends(station_number = "08NM116",
                               zyp_method = "zhang",
                               zyp_alpha = 0.05)

# Compute trends statistics and do not plot the results
trends <- compute_annual_trends(station_number = "08NM116",
                               zyp_method = "zhang",
                               include_plots = FALSE)

## End(Not run)
```
**compute_frequency_analysis**

*Perform a custom volume frequency analysis*

**Description**

Performs a volume frequency analysis on custom data. Defaults to ranking by minimums; use `use_max` for to rank by maximum flows. Calculates the statistics from events and flow values provided. Columns of events (e.g. years), their values (minimums or maximums), and identifiers (low-flows, high-flows, etc.). Function will calculate using all values in the provided data (no grouped analysis). Analysis methodology replicates that from HEC-SSP. Returns a list of tibbles and plots.

**Usage**

```r
compute_frequency_analysis(
  data,
  events = Year,
  values = Value,
  measures = Measure,
  use_max = FALSE,
  use_log = FALSE,
  prob_plot_position = c("weibull", "median", "hazen"),
  prob_scale_points = c(0.9999, 0.999, 0.99, 0.9, 0.5, 0.2, 0.1, 0.02, 0.01, 0.001, 1e-04),
  compute_fitting = TRUE,
  fit_distr = c("PIII", "weibull"),
  fit_distr_method = ifelse(fit_distr == "PIII", "MOM", "MLE"),
  fit_quantiles = c(0.975, 0.99, 0.98, 0.95, 0.9, 0.8, 0.5, 0.2, 0.1, 0.05, 0.01),
  plot_curve = TRUE,
  plot_axis_title = "Discharge (cms)"
)
```

**Arguments**

- **data**
  A data frame of data that contains columns of events, flow values, and measures (data type).

- **events**
  Column in data that contains event identifiers, typically year values. Default 'Year'.

- **values**
  Column in data that contains numeric flow values, in units of cubic metres per second. Default 'Value'.

- **measures**
  Column in data that contains measure identifiers (example data: '7-day low' or 'Annual Max'). Can have multiple measures (e.g. '7-day low' and '30-day low') in column if multiple statistics are desired. Default 'Measure'.

- **use_max**
  Logical value to indicate using maximums rather than the minimums for analysis. Default FALSE.
compute_frequency_analysis

use_log Logical value to indicate log-scale transforming of flow data before analysis. Default FALSE.

prob_plot_position Character string indicating the plotting positions used in the frequency plots, one of 'weibull', 'median', or 'hazen'. Points are plotted against (i-a)/(n+1-a-b) where i is the rank of the value; n is the sample size and a and b are defined as: (a=0, b=0) for Weibull plotting positions; (a=.2; b=.3) for Median plotting positions; and (a=.5; b=.5) for Hazen plotting positions. Default 'weibull'.

prob_scale_points Numeric vector of probabilities to be plotted along the X axis in the frequency plot. Inverse of return period. Default c(.9999, .999, .99, .9, .5, .2, .1, .02, .01, .001, .0001).

compute_fitting Logical value to indicate whether to fit plotting positions to a distribution. If 'FALSE' the output will return only the data, plotting positions, and plot. Default TRUE.

fit_distr Character string identifying the distribution to fit annual data, one of 'PIII' (Log Pearson Type III) or 'weibull' (Weibull) distributions. Default 'PIII'.

fit_distr_method Character string identifying the method used to fit the distribution, one of 'MOM' (method of moments) or 'MLE' (maximum likelihood estimation). Selected as 'MOM' if fit_distr = 'PIII' (default) or 'MLE' if fit_distr = 'weibull'.

fit_quantiles Numeric vector of quantiles to be estimated from the fitted distribution. Default c(.975, .99, .98, .95, .90, .80, .50, .20, .10, .05, .01).

plot_curve Logical value to indicate plotting the computed curve on the probability plot. Default TRUE.

plot_axis_title Character string of the plot y-axis title. Default 'Discharge (cms)'.

Value

A list with the following elements:

Freq_Analysis_Data Data frame with provided data for analysis.
Freq_Plot_Data Data frame with plotting positions used in frequency plot.
Freq_Plot ggplot2 object with plotting positions and (optional) fitted curve.
Freq_Fitting List of fitted objects from fitdistrplus.
Freq_Fitted_Quantiles Data frame with fitted quantiles.

Examples

## Not run:

# Working example:

```r
```
# Calculate some values to use for a frequency analysis
# (requires years, values for those years, and the name of the measure/metric)
low_flows <- calc_annual_lowflows(station_number = "08NM116",
                                   start_year = 1980,
                                   end_year = 2000,
                                   roll_days = 7)
low_flows <- dplyr::select(low_flows, Year, Value = Min_7_Day)
low_flows <- dplyr::mutate(low_flows, Measure = "7-Day")

# Compute the frequency analysis using the default parameters
results <- compute_frequency_analysis(data = low_flows,
                                       events = Year,
                                       values = Value,
                                       measure = Measure)

## End(Not run)

---

**compute_frequency_quantile**

*Calculate an annual frequency analysis quantile*

**Description**

Performs a volume frequency analysis on annual statistics from a daily streamflow data set and calculates a statistic based on the provided mean n-days and return period of the statistic, defaults to minimum flows. For example, to determine the 7Q10 of a data set, set the `roll_days` to 7 and the `return_period` to 10. Function will calculate using all values in 'Values' column (no grouped analysis), unless specified. Analysis methodology replicates that from HEC-SSP. Returns a tibble with statistics.

**Usage**

```r
compute_frequency_quantile(
  data,
  dates = Date,
  values = Value,
  station_number,
  roll_days = NA,
  roll_align = "right",
  return_period = NA,
  use_max = FALSE,
  use_log = FALSE,
  fit_distr = c("PIII", "weibull"),
  fit_distr_method = ifelse(fit_distr == "PIII", "MOM", "MLE"),
  water_year_start = 1,
  start_year,
  end_year,
)```
compute_frequency_quantile

```r
eclude_years,
months = 1:12,
ignore_missing = FALSE,
allowed_missing = ifelse(ignore_missing, 100, 0)
)
```

**Arguments**

- **data**
  - A data frame of data that contains columns of events, flow values, and measures (data type).

- **dates**
  - Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.

- **values**
  - Column in data that contains numeric flow values, in units of cubic metres per second. Default 'Value'.

- **station_number**
  - Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

- **roll_days**
  - Numeric value of the number of days to apply a rolling mean. Required.

- **roll_align**
  - Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.

- **return_period**
  - Numeric vector of the estimated time interval, in years, between flow events of a similar size, inverse of probability, used to estimate the frequency statistic. Required.

- **use_max**
  - Logical value to indicate using maximums rather than the minimums for analysis. Default FALSE.

- **use_log**
  - Logical value to indicate log-scale transforming of flow data before analysis. Default FALSE.

- **fit_distr**
  - Character string identifying the distribution to fit annual data, one of 'PIII' (Log Pearson Type III) or 'weibull' (Weibull) distributions. Default 'PIII'.

- **fit_distr_method**
  - Character string identifying the method used to fit the distribution, one of 'MOM' (method of moments) or 'MLE' (maximum likelihood estimation). Selected as 'MOM' if fit_distr = 'PIII' (default) or 'MLE' if fit_distr = 'weibull'.

- **water_year_start**
  - Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

- **start_year**
  - Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

- **end_year**
  - Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

- **exclude_years**
  - Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.
compute_full_analysis

months Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when `water_year_start = 10` (Oct). Default summarizes all months (1:12).

ignore_missing Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

allowed_missing Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used.

Value

A numeric value of the frequency analysis quantile, given the roll_days and return_period.

See Also

compute_frequency_analysis

Examples

## Not run:

# Working example:

# Compute the annual 7-day flow value with a 1 in 10 year return interval
compute_frequency_quantile(station_number = "08NM116",
                           roll_days = 7,
                           return_period = 10)

## End(Not run)

compute_full_analysis  Compute a suite of tables and plots from various fasstr functions

Description

Calculates tables and plots from a suite of statistics from fasstr functions. Calculates statistics from all values, unless specified. The statistics are grouped into 7 analysis groups (see analyses argument) which are stored in lists in the object. Due to the number of tables and plots to be made, this function may take several minutes to complete. If ignore_missing = FALSE (default) and there is missing data, some tables and plots may be empty and produce warnings. Use ignore_missing = TRUE to ignore the missing values or filter your data to complete years. Returns a list of tibbles and plots.
Usage

compute_full_analysis(
    data, 
    dates = Date, 
    values = Value, 
    groups = station_number, 
    analyses = 1:7, 
    basin_area, 
    water_year_start = 1, 
    start_year, 
    end_year, 
    exclude_years, 
    months = 1:12, 
    complete_years = FALSE, 
    ignore_missing = FALSE, 
    allowed_missing_annual = ifelse(ignore_missing, 100, 0), 
    allowed_missing_monthly = ifelse(ignore_missing, 100, 0), 
    zyp_method = "zhang", 
    zyp_alpha
)

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
analyses Numeric vector of analyses to run (default is all (1:7)):
  • 1: Screening
  • 2: Long-term
  • 3: Annual
• 4: Monthly
• 5: Daily
• 6: Annual Trends
• 7: Low-flow Frequencies

**basin_area**
Upstream drainage basin area, in square kilometres, to apply to observations. Three options:
1. Leave blank if groups is STATION_NUMBER with HYDAT station numbers to extract basin areas from HYDAT.
2. A single numeric value to apply to all observations.
3. List each basin area for each group/station in groups (can override HYDAT value if listed) as such c("08NM116" = 795, "08NM242" = 10). If group is not listed the HYDAT area will be applied if it exists, otherwise it will be NA.

**water_year_start**
Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

**start_year**
Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

**end_year**
Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

**exclude_years**
Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

**months**
Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12). If not all months, seasonal total yield and volumetric flows will not be included.

**complete_years**
Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

**ignore_missing**
Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

**allowed_missing_annual**
Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate an annual statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used. Only for annual means, percentiles, minimums, and maximums.

**allowed_missing_monthly**
Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a monthly statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used. Only for monthly means, percentiles, minimums, and maximums.
compute_hydat_peak_frequencies

zyp_method  Character string identifying the prewhitened trend method to use from 'zyp', either 'zhang' or 'yuepilon'. 'zhang' is recommended over 'yuepilon' for hydrologic applications (see compute_annual_trends(); Bürger 2017; Zhang and Zwiers 2004). Only required if analysis group 6 is included. Default 'zhang'.

zyp_alpha  Numeric value of the significance level (ex. 0.05) of when to plot a trend line. Leave blank for no line.

Value

A list of lists of tibble data frames and ggplot2 objects from various fasstr functions organized by the analysis groups as listed above.

See Also

```
plot_flow_data, screen_flow_data, plot_data_screening, plot_missing_dates, calc_longterm_monthly_stats, plot_longterm_monthly_stats, calc_longterm_daily_stats, plot_longterm_daily_stats, plot_monthly_means, plot_flow_duration, calc_annual_stats, plot_annual_stats, calc_annual_cumulative_stats, plot_annual_cumulative_stats, calc_annual_flow_timing, plot_annual_flow_timing, calc_annual_normal_days, plot_annual_normal_days, calc_annual_lowflows, plot_annual_lowflows, plot_annual_means, calc_monthly_stats, plot_monthly_stats, calc_monthly_cumulative_stats, plot_monthly_cumulative_stats, calc_daily_stats, plot_daily_stats, calc_daily_cumulative_stats, plot_daily_cumulative_stats, compute_annual_trends, compute_annual_frequencies, write_flow_data, write_plots
```

Examples

```
## Not run:

# Working examples:

# Compute a full analysis will all the analyses
results <- compute_full_analysis(station_number = "08NM116",
                                  start_year = 1980,
                                  end_year = 2010)

# Compute a full analysis with only Annual (3) and Daily (5) analyses
results <- compute_full_analysis(station_number = "08NM116",
                                  start_year = 1980,
                                  end_year = 2010,
                                  analyses = c(3,5))

## End(Not run)
```

---

**compute_hydat_peak_frequencies**

*Perform a frequency analysis on annual peak statistics from HYDAT*
compute_hydat_peak_frequencies

Description

Performs a volume frequency analysis on annual peak statistics (instantaneous minimums or maximums) extracted from HYDAT. Calculates statistics from all years, unless specified. The data argument is not available. Analysis methodology replicates that from HEC-SSP. Returns a list of tibbles and plots.

Usage

compute_hydat_peak_frequencies(
  station_number,
  use_max = FALSE,
  use_log = FALSE,
  prob_plot_position = c("weibull", "median", "hazen"),
  prob_scale_points = c(0.9999, 0.999, 0.99, 0.9, 0.5, 0.2, 0.1, 0.02, 0.01, 0.001, 1e-04),
  fit_distr = c("PIII", "weibull"),
  fit_distr_method = ifelse(fit_distr == "PIII", "MOM", "MLE"),
  fit_quantiles = c(0.975, 0.99, 0.98, 0.95, 0.9, 0.8, 0.5, 0.2, 0.1, 0.05, 0.01),
  start_year,
  end_year,
  exclude_years,
  plot_curve = TRUE
)

Arguments

station_number A character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract annual peak minimum or maximum instantaneous streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database.

use_max Logical value to indicate using maximums rather than the minimums for analysis. Default FALSE.

use_log Logical value to indicate log-scale transforming of flow data before analysis. Default FALSE.

prob_plot_position Character string indicating the plotting positions used in the frequency plots, one of 'weibull', 'median', or 'hazen'. Points are plotted against \((i-a)/(n+1-a-b)\) where \(i\) is the rank of the value; \(n\) is the sample size and \(a\) and \(b\) are defined as: (a=0, b=0) for Weibull plotting positions; (a=0.2; b=0.3) for Median plotting positions; and (a=0.5; b=0.5) for Hazen plotting positions. Default 'weibull'.

prob_scale_points Numeric vector of probabilities to be plotted along the X axis in the frequency plot. Inverse of return period. Default \(c(0.9999, 0.999, 0.99, 0.9, 0.5, 0.2, 0.1, 0.02, 0.01, 0.001, 0.0001)\).

fit_distr Character string identifying the distribution to fit annual data, one of 'PIII' (Log Pearson Type III) or 'weibull' (Weibull) distributions. Default 'PIII'.

fit_distr_method Character string identifying the distribution fitting method to use, one of 'MOM' or 'MLE'. Default 'MOM' for PIII distributions. For Weibull distributions, 'MLE' is used by default.
fit_distr_method

Character string identifying the method used to fit the distribution, one of 'MOM' (method of moments) or 'MLE' (maximum likelihood estimation). Selected as 'MOM' if fit_distr = 'PIII' (default) or 'MLE' if fit_distr = 'weibull'.

fit_quantiles

Numeric vector of quantiles to be estimated from the fitted distribution. Default c(.975, .99, .98, .95, .90, .80, .50, .20, .10, .05, .01).

start_year

Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year

Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years

Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

plot_curve

Logical value to indicate plotting the computed curve on the probability plot. Default TRUE.

Value

A list with the following elements:

Freq_Analysis_Data

Data frame with computed annual summary statistics used in analysis.

Freq_Plot_Data

Data frame with co-ordinates used in frequency plot.

Freq_Plot

ggplot2 object with frequency plot

Freq_Fitting

List of fitted objects from fitdistrplus.

Freq_Fitted_Quantiles

Data frame with fitted quantiles.

See Also

compute_frequency_analysis

Examples

## Not run:

# Working examples (see arguments for further analysis options):

# Compute an annual peak frequency analysis using default arguments (instantaneous lows)
results <- compute_hydat_peak_frequencies(station_number = "08NM116",
start_year = 1980,
end_year = 2010)

# Compute an annual peak frequency analysis using default arguments (instantaneous highs)
results <- compute_hydat_peak_frequencies(station_number = "08NM116",
start_year = 1980,
end_year = 2010,
use_max = TRUE)

## End(Not run)
fill_missing_dates  Fills data gaps of missing dates

Description

Fills data gaps of missing dates of the data provided. Builds a continuous data set from the start date to the end date. Only missing dates are filled, columns not specified as dates or groups will be filled with NA. Will completely fill first and last years, unless specified using pad_ends = FALSE.

Usage

```r
fill_missing_dates(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  water_year_start = 1,
  pad_ends = TRUE
)
```

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
<td>Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.</td>
</tr>
<tr>
<td>dates</td>
<td>Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.</td>
</tr>
<tr>
<td>values</td>
<td>Name of column in data that contains numeric flow values, in units of cubic metres per second. Not required as of fasstr 0.3.3 as all other columns are filled with NA.</td>
</tr>
<tr>
<td>groups</td>
<td>Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.</td>
</tr>
<tr>
<td>station_number</td>
<td>Character string vector of seven digit Water Survey of Canada station numbers (e.g. &quot;08NM116&quot;) of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.</td>
</tr>
<tr>
<td>water_year_start</td>
<td>Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.</td>
</tr>
<tr>
<td>pad_ends</td>
<td>Logical value indicating whether to fill incomplete start and end years with rows of dates. If FALSE then only missing dates between the provided start and end dates will be filled. Default TRUE.</td>
</tr>
</tbody>
</table>
plot_annual_cumulative_stats

Description

Plots annual and seasonal (if include_seasons = TRUE) total cumulative flows, volumetric discharge or water yields, from a daily streamflow data set. Calculates statistics from all values, unless specified. Data calculated from plot_annual_cumulative_stats() function. For water year and seasonal data, the designated year is the year in which the year or season ends. Returns a list of plots.

Usage

plot_annual_cumulative_stats(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  use_yield = FALSE,
  basin_area,
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  include_seasons = FALSE,
  include_title = FALSE,
  complete_years = FALSE,
  plot_type = "bar"
)
Arguments

**data**
Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.

**dates**
Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.

**values**
Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.

**groups**
Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.

**station_number**
Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

**use_yield**
Logical value indicating whether to calculate area-based water yield, in mm, instead of volumetric discharge. Default FALSE.

**basin_area**
Upstream drainage basin area, in square kilometres, to apply to observations. Three options:

1. Leave blank if groups is STATION_NUMBER with HYDAT station numbers to extract basin areas from HYDAT.
2. A single numeric value to apply to all observations.
3. List each basin area for each group/station in groups (can override HYDAT value if listed) as such c("08NM116" = 795, "08NM242" = 10). If group is not listed the HYDAT area will be applied if it exists, otherwise it will be NA.

**water_year_start**
Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

**start_year**
Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

**end_year**
Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

**exclude_years**
Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

**months**
Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12). If not all months, seasonal total yield and volumetric flows will not be included.

**include_seasons**
Logical value indication whether to include seasonal yields or volumetric discharges. Default TRUE.
plot_annual_cumulative_stats

include_title Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

complete_years Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

plot_type Type of plot, either "bar" or "line" styles. Default "bar". Use "line" for previous version of plot.

Value

A list of ggplot2 objects with the following for each station provided:

Annual_Total_Volume
annual total volumetric discharge, in cubic metres

Two_Seasons_Total_Volume
if include_seasons = TRUE, two seasons total volumetric discharges, in cubic metres

Four_Seasons_Total_Volume
if include_seasons = TRUE, four seasons total volumetric discharges, in cubic metres

If use_yield argument is used the list will contain the following objects:

Annual_Yield annual water yield, in millimetres

Two_Seasons_Yield
if include_seasons = TRUE, two seasons water yield, in millimetres

Four_Seasons_Yield
if include_seasons = TRUE, four seasons water yield, in millimetres

See Also

calc_annual_cumulative_stats

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Plot annual cumulative yield statistics with default HYDAT basin area
plot_annual_cumulative_stats(station_number = "08NM116",
                            use_yield = TRUE)

# Plot annual cumulative yield statistics with custom basin area
plot_annual_cumulative_stats(station_number = "08NM116",
                            use_yield = TRUE,
                            basin_area = 800)
}
plot_annual_extremes

Plot annual high and low flows

Description

Plots annual n-day minimum and maximum values and the day of year of occurrence of daily flow values from a daily streamflow data set. Calculates statistics from all values, unless specified. Returns a tibble with statistics.

Usage

plot_annual_extremes(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  roll_days = 1,
  roll_days_min = NA,
  roll_days_max = NA,
  roll-align = "right",
  water-year-start = 1,
  start-year,
  end-year,
  exclude-years,
  months = 1:12,
  months_min = NA,
  months_max = NA,
  complete-years = FALSE,
  ignore-missing = FALSE,
  allowed-missing = ifelse(ignore-missing, 100, 0),
  include-title = FALSE
)

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
plot_annual_extremes

groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.

station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

roll_days Numeric value of the number of days to apply a rolling mean. Default 1.

roll_days_min Numeric value of the number of days to apply a rolling mean for low flows. Will override 'roll_days' argument for low flows. Default NA.

roll_days_max Numeric value of the number of days to apply a rolling mean for high flows. Will override 'roll_days' argument for high flows. Default NA.

roll_align Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.

water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

months_min Numeric vector of specified months for window of low flows (3 for March, 6:8 for Jun-Aug). Will override 'months' argument for low flows. Default NA.

months_max Numeric vector of specified months for window of high flows (3 for March, 6:8 for Jun-Aug). Will override 'months' argument for high flows. Default NA.

complete_years Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

ignore_missing Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

allowed_missing Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used.
include_title Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

Value
A list of ggplot2 objects with the following for each station provided:

Annual_Extreme_Flows ggplot2 object of annual minimum and maximum flows of selected n-day rolling means
Annual_Extreme_Flows_Dates ggplot2 object of the day of years of annual minimum and maximum flows of selected n-day rolling means

See Also
calc_annual_extremes

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {
  # Plot annual 1-day (default) max/min flow data with
  # default alignment ('right')
  plot_annual_extremes(station_number = "08NM116")

  # Plot custom annual 3-day max and 7-min flow data with 'center' alignment
  plot_annual_extremes(station_number = "08NM116",
                      roll_days_max = 3,
                      roll_days_min = 7,
                      roll_align = "center")
}

plot_annual_extremes_year

Plot annual high and low flows for a specific year

Description
Plots an annual hydrograph for a specific year with the values and timing of annual n-day low and high flows. The ‘normal’ range of percentiles also plotted for reference and are calculated from only years of complete data. Shows the values and dates of max/mins for a specific year from the calc_annual_extremes() and plot_annual_extremes() functions. Can remove either low or high flows using plot_min = FALSE() or plot_max = FALSE(), respectively. Returns a list of plots.
Usage

plot_annual_extremes_year(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  year_to_plot = NA,
  roll_days = 1,
  roll_days_min = NA,
  roll_days_max = NA,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  months_min = NA,
  months_max = NA,
  log_discharge = TRUE,
  log_ticks = FALSE,
  include_title = FALSE,
  plot_normal_percentiles = TRUE,
  normal_percentiles = c(25, 75),
  plot_min = TRUE,
  plot_max = TRUE,
  complete_years = FALSE,
  ignore_missing = FALSE,
  allowed_missing = ifelse(ignore_missing, 100, 0)
)

Arguments

data

Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.

dates

Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.

values

Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.

groups

Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
plot_annual_extremes_year

station_number: Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

year_to_plot: Numeric value indicating the year/water year to plot flow data with normal category colours. Default NA.

roll_days: Numeric value of the number of days to apply a rolling mean. Default 1.

roll_days_min: Numeric value of the number of days to apply a rolling mean for low flows. Will override 'roll_days' argument for low flows. Default NA.

roll_days_max: Numeric value of the number of days to apply a rolling mean for high flows. Will override 'roll_days' argument for high flows. Default NA.

roll_align: Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.

water_year_start: Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year: Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year: Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years: Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months: Numeric vector of specific months to plot. For example, 3 for March, 6:8 for Jun-Aug. Will be overridden for low or high flow statistics if months_min or months_max set, but will still define the date limits on the x-axis. Default plots all months (1:12).

months_min: Numeric vector of specified months for window of low flows (3 for March, 6:8 for Jun-Aug). Will override 'months' argument for low flows. Default NA.

months_max: Numeric vector of specified months for window of high flows (3 for March, 6:8 for Jun-Aug). Will override 'months' argument for high flows. Default NA.

log_discharge: Logical value to indicate plotting the discharge axis (Y-axis) on a logarithmic scale. Default FALSE.

log_ticks: Logical value to indicate plotting logarithmic scale ticks when log_discharge = TRUE. Ticks will not appear when log_discharge = FALSE. Default to TRUE when log_discharge = TRUE.

include_title: Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

plot_normal_percentiles: Logical value indicating whether to plot the normal percentiles ribbon. Default TRUE.

normal_percentiles: Numeric vector of two values, lower and upper percentiles, respectively indicating the limits of the normal range. Default c(25,75).
plot_annual_flow_timing

plot_min Logical value indicating whether to plot annual low flows. Default TRUE.
plot_max Logical value indicating whether to plot annual high flows. Default TRUE.
complete_years Logical values indicating whether to include only years with complete data in analysis. Default FALSE.
ignore_missing Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.
allowed_missing Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used.

Value
A list of ggplot2 objects with the following for each station provided:

Annual_Extremes_Year
a plot that contains the an annual hydrograph and identified low and high flow periods

See Also
calc_annual_extremes
plot_annual_extremes

Examples
# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Plot the year 2000 with the annual maximum and minimums
plot_annual_extremes_year(station_number = "08NM116",
                     roll_days_max = 3,
                     roll_days_min = 7,
                     year_to_plot = 2001)
}

plot_annual_flow_timing
Plot annual timing of flows
Description

Plots the timing (day of year and date) of portions of total annual flow of daily flow values from a daily streamflow data set. Calculates statistics from all values from complete years, unless specified. Data calculated using `calc_annual_flow_timing()` function. Returns a list of plots.

Usage

```r
plot_annual_flow_timing(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  percent_total = c(25, 33.3, 50, 75),
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  include_title = FALSE
)
```

Arguments

data: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g., station numbers). Leave blank or set to NULL if using `station_number` argument.
dates: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using `station_number` argument.
values: Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using `station_number` argument.
groups: Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using `station_number` argument.
station_number: Character string vector of seven digit Water Survey of Canada station numbers (e.g., "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
percent_total: Numeric vector of percents of total annual flows to determine dates. Default `c(25, 33.3, 50, 75)`.
water_year_start: Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
plot_annual_flow_timing

start_year  Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year  Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years  Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months  Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

include_title  Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

Value

A list of ggplot2 objects with the following for each station provided:

Annual_Flow_Timing

a plot that contains each n-percent of total volumetric discharge

Default plots on each object:

DoY_25pct_TotalQ  day of year of 25-percent of total volumetric discharge

DoY_33.3pct_TotalQ  day of year of 33.3-percent of total volumetric discharge

DoY_50pct_TotalQ  day of year of 50-percent of total volumetric discharge

DoY_75pct_TotalQ  day of year of 75-percent of total volumetric discharge

References


See Also

calc_annual_flow_timing

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Plot annual flow timing statistics with default percent totals
plot_annual_flow_timing(station_number = "08NM116")

# Plot annual flow timing with custom percent totals
plot_annual_flow_timing_year

plot_annual_flow_timing(station_number = "08NM116",
percent_total = 50,
start_year = 1980)

plot_annual_flow_timing_year

Plot annual timing of flows for a specific year

Description

Plots an annual hydrograph for a specific year with the dates of flow timing of portions of total annual flow identified. The ‘normal’ range of percentiles also plotted for reference and are calculated from only years of complete data. Shows the dates of flow timing for a specific year from the counts from the \texttt{plot_annual_flow_timing()} function. Returns a list of plots.

Usage

\begin{verbatim}
plot_annual_flow_timing_year(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  percent_total = c(25, 33.3, 50, 75),
  year_to_plot = NA,
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  log_discharge = TRUE,
  log_ticks = FALSE,
  include_title = FALSE,
  plot_vlines = TRUE,
  plot_normal_percentiles = TRUE,
  normal_percentiles = c(25, 75)
)
\end{verbatim}

Arguments

\begin{itemize}
  \item \texttt{data} Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using \texttt{station_number} argument.
  \item \texttt{dates} Name of column in \texttt{data} that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using \texttt{station_number} argument.
\end{itemize}
values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.

groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.

station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

percent_total Numeric vector of percents of total annual flows to determine dates. Default c(25, 33.3, 50, 75).

year_to_plot Numeric value indicating the year/water year to plot flow data with normal category colours. Default NA.

water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

log_discharge Logical value to indicate plotting the discharge axis (Y-axis) on a logarithmic scale. Default FALSE.

log_ticks Logical value to indicate plotting logarithmic scale ticks when log_discharge = TRUE. Ticks will not appear when log_discharge = FALSE. Default to TRUE when log_discharge = TRUE.

include_title Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

plot_vlines Logical value indicating whether to plot the vertical lines indicating dates of flow timing. Default TRUE.

plot_normal_percentiles Logical value indicating whether to plot the normal percentiles ribbon. Default TRUE.

normal_percentiles Numeric vector of two values, lower and upper percentiles, respectively indicating the limits of the normal range. Default c(25, 75).
plot_annual_highflows

Value

A list of ggplot2 objects with the following for each station provided:

Annual_Normal_Days_Year

- a plot that contains the above, below, and normal colour daily flow points

See Also

calc_annual_flow_timing
plot_annual_flow_timing

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {
  # Plot the year 2000 and change the flow timing percent totals
  plot_annual_flow_timing_year(station_number = "08NM116",
                               percent_total = 50,
                               year_to_plot = 2000)
}

plot_annual_highflows  Plot annual high flows and dates

Description

Plot annual n-day maximum values, and the day of year and date of occurrence of daily flow values from a daily streamflow data set. Calculates statistics from all values, unless specified. Data calculated from calc_annual_highflows() function. Returns a list of plots.

Usage

plot_annual_highflows(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  roll_days = c(1, 3, 7, 30),
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  complete_years = FALSE,
ignore_missing = FALSE,
allowed_missing = ifelse(ignore_missing, 100, 0),
include_title = FALSE)
}

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
roll_days Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
start_year Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.
end_year Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.
exclude_years Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.
months Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).
complete_years Logical values indicating whether to include only years with complete data in analysis. Default FALSE.
ignore_missing Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing
plot_annual_lowflows

dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

allowed_missing

Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used.

include_title

Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

Value

A list of ggplot2 objects with the following for each station provided:

Annual_Maximums

ggplot2 object of annual maximums of selected n-day rolling means

Annual_Maximums_Days

ggplot2 object of the day of years of annual maximums of selected n-day rolling means

See Also

calc_annual_highflows

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

    # Plot annual 1, 3, 7, and 30-day (default) high flow statistics with default alignment
    plot_annual_highflows(station_number = "08NM116")

    # Plot annual custom 3 and 7-day high flow statistics with "center" alignment
    plot_annual_highflows(station_number = "08NM116",
                          roll_days = c(3,7),
                          roll_align = "center")

}

plot_annual_lowflows

Plot annual low flows and dates

Description

Plot annual n-day minimum values, and the day of year and date of occurrence of daily flow values from a daily streamflow data set. Calculates statistics from all values, unless specified. Data calculated from calc_annual_lowflows() function. Returns a list of plots.
Usage

plot_annual_lowflows(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  roll_days = c(1, 3, 7, 30),
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  complete_years = FALSE,
  ignore_missing = FALSE,
  allowed_missing = ifelse(ignore_missing, 100, 0),
  include_title = FALSE
)

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
roll_days Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
start_year Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

complete_years Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

ignore_missing Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

allowed_missing Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used.

include_title Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

Value

A list of ggplot2 objects with the following for each station provided:

Annual_Minimums ggplot2 object of annual minimums of selected n-day rolling means

Annual_Minimums_Days ggplot2 object of the day of years of annual minimums of selected n-day rolling means

See Also

calc_annual_lowflows

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Plot annual 1, 3, 7, and 30-day (default) low flow statistics with default alignment
plot_annual_lowflows(station_number = "08NM116")

# Plot annual custom 3 and 7-day low flow statistics with "center" alignment
plot_annual_lowflows(station_number = "08NM116", ignored missing = TRUE)
Plot annual means compared to the long-term mean

Description

Plot annual means using the long-term annual mean as the point of reference for annual means. Calculates statistics from all values, unless specified. Data calculated using calc_annual_stats() function. Returns a list of plots.

Usage

```
plot_annual_means(
  data,                      # Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
  dates = Date,              # Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
  values = Value,            # Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
  groups = STATION_NUMBER,   # (Optional) groups (e.g. station numbers).
  station_number = NULL,     # Data calculated using calc_annual_stats() function. Returns a list of plots.
  roll_days = c(3,7),        # Roll days for calculation.
  roll_align = "center",     # Roll alignment.
  water_year_start = 1,      # Water year start.
  start_year,                # Start year.
  end_year,                  # End year.
  exclude_years,             # Exclude years.
  months = 1:12,             # Months.
  complete_years = FALSE,    # Complete years.
  ignore_missing = FALSE,    # Ignore missing.
  allowed_missing = ifelse(ignore_missing, 100, 0),
  include_title = FALSE,     # Include title.
  percentiles_mad = c(10, 90) # Percentiles MAD.
)
```

Arguments

- **data**: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
- **dates**: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
- **values**: Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.

Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

Numeric value of the number of days to apply a rolling mean. Default 1.

Character string identifying the direction of the rolling mean from the specified date, either by the first ("left"), last ("right"), or middle ("center") day of the rolling n-day group of observations. Default "right".

Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used.

Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

Numeric vector of percentiles of annual means to plot, up to two values. Set to NA if none required. Default c(10, 90).

A list of ggplot2 objects for with the following plots for each station provided:

Annual_Means a plot that contains annual means with the long-term mean as the x-axis intercept
plot_annual_normal_days

See Also
calc_annual_stats

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Plot annual means
plot_annual_means(station_number = "08NM116")

# Plot mean flows from July-September
plot_annual_means(station_number = "08NM116",
                   months = 7:9)
}

plot_annual_normal_days

Plot annual count of normal days and days above and below normal

Description

Plots the number of days per year within, above and below the 'normal' range (typically between 25 and 75th percentiles) for each day of the year. Upper and lower-range percentiles are calculated for each day of the year of from all years, and then each daily flow value for each year is compared. Calculates statistics from all values from complete years, unless specified. Data calculated using calc_annual_normal_days() function. Returns a list of plots.

Usage

plot_annual_normal_days(
  data,     
  dates = Date,       
  values = Value,     
  groups = STATION_NUMBER,  
  station_number,     
  normal_percentiles = c(25, 75),           
  roll_days = 1,       
  roll_align = "right",      
  water_year_start = 1,     
  start_year,              
  end_year,               
  exclude_years,          
  months = 1:12,           
  include_title = FALSE)

Arguments

data  Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates  Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values  Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not ‘Value’ (default). Leave blank if using station_number argument.
groups  Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not ‘STATION_NUMBER’. Function will automatically group by a column named ‘STATION_NUMBER’ if present. Remove the ‘STATION_NUMBER’ column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number  Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
normal_percentiles  Numeric vector of two values, lower and upper percentiles, respectively indicating the limits of the normal range. Default c(25, 75).
roll_days  Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align  Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
water_year_start  Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
start_year  Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.
end_year  Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.
exclude_years  Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.
months  Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).
include_title  Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

Value

A list of ggplot2 objects with the following for each station provided:
Annual_Normal_Days

a plot that contains the number of days outside normal

Default plots on each object:

Normal_Days number of days per year below and above the daily normal (default 25/75th percentile)

Below_Normal_Days number of days per year below the daily normal (default 25th percentile)

Above_Normal_Days number of days per year above the daily normal (default 75th percentile)

See Also

calc_annual_normal_days

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

  # Plot annual statistics with default limits of normal (25 and 75th percentiles)
  plot_annual_normal_days(station_number = "08NM116")

  # Plot annual statistics with custom limits of normal
  plot_annual_normal_days(station_number = "08NM116", normal_percentiles = c(10,90))

}

plot_annual_normal_days_year

Plot days above normal, below normal and normal for a specific year

Description

Plots an annual hydrograph for a specific year with daily flow values coloured by whether the daily values are normal, above normal, or below normal, overlaying the normals range. The normal range is typically between 25 and 75th percentiles for each day of the year. Upper and lower-range percentiles are calculated for each day of the year of from all years, and then each daily flow value for each year is compared. Normals calculated from only years of complete data, although incomplete years can be plotted. Shows the annual values for a specific year from the counts from the plot_annual_normal_days() function. Returns a list of plots.
Usage

```r
plot_annual_normal_days_year(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  normal_percentiles = c(25, 75),
  year_to_plot = NA,
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  log_discharge = TRUE,
  log_ticks = FALSE,
  include_title = FALSE,
  plot_flow_line = TRUE,
  plot_normal_percentiles = TRUE
)
```

Arguments

data: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values: Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups: Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number: Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
normal_percentiles: Numeric vector of two values, lower and upper percentiles, respectively indicating the limits of the normal range. Default c(25, 75).
**year_to_plot**  Numeric value indicating the year/water year to plot flow data with normal category colours. Default NA.

**roll_days**  Numeric value of the number of days to apply a rolling mean. Default 1.

**roll_align**  Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.

**water_year_start**  Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

**start_year**  Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

**end_year**  Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

**exclude_years**  Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

**months**  Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

**log_discharge**  Logical value to indicate plotting the discharge axis (Y-axis) on a logarithmic scale. Default FALSE.

**log_ticks**  Logical value to indicate plotting logarithmic scale ticks when log_discharge = TRUE. Ticks will not appear when log_discharge = FALSE. Default to TRUE when log_discharge = TRUE.

**include_title**  Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

**plot_flow_line**  Logical value indicating whether to connect flow data coloured points with lines. Default TRUE.

**plot_normal_percentiles**  Logical value indicating whether to plot the normal percentiles ribbon. Default TRUE.

**Value**

A list of ggplot2 objects with the following for each station provided:

*Annual_Normal_Days_Year*

a plot that contains the above, below, and normal colour daily flow points

**See Also**

calc_annual_normal_days

plot_annual_normal_days
Examples

```r
# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

  # Plot the year 2000 using a data frame and data argument with defaults
  flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
  plot_annual_normal_days_year(data = flow_data,
                               year_to_plot = 2000)

  # Plot the year 2000 using the station_number argument
  plot_annual_normal_days_year(station_number = "08NM116",
                               year_to_plot = 2000)

  # Plot the year 2000 and change the normal percentiles range
  plot_annual_normal_days_year(station_number = "08NM116",
                               normal_percentiles = c(20, 80),
                               year_to_plot = 2000)
}
```

Description

This function has been superseded by the `plot_annual_normal_days()` function.

Plots the number of days per year outside of the ‘normal’ range (typically between 25 and 75th percentiles) for each day of the year. Upper and lower-range percentiles are calculated for each day of the year of from all years, and then each daily flow value for each year is compared. All days above or below the normal range are included. Calculates statistics from all values from complete years, unless specified. Data calculated using `calc_annual_outside_normal()` function. Returns a list of plots.

Usage

```r
plot_annual_outside_normal(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  normal_percentiles = c(25, 75),
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
)```
Arguments

data          Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates         Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values        Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups        Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
normal_percentiles Numeric vector of two values, lower and upper percentiles, respectively indicating the limits of the normal range. Default c(25, 75).
roll_days     Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align    Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
start_year    Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.
end_year      Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.
exclude_years Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.
months        Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).
include_title Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.
plot_annual_stats

Value

A list of ggplot2 objects with the following for each station provided:

Annual_Days_Outside_Normal
   a plot that contains the number of days outside normal

Default plots on each object:

Days_Below_Normal
   number of days per year below the daily normal (default 25th percentile)

Days_Above_Normal
   number of days per year above the daily normal (default 75th percentile)

Days_Outside_Normal
   number of days per year below and above the daily normal (default 25/75th percentile)

See Also

calc_annual_outside_normal

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Plot annual statistics with default limits of normal (25 and 75th percentiles)
plot_annual_outside_normal(station_number = "08NM116")

# Plot annual statistics with custom limits of normal
plot_annual_outside_normal(station_number = "08NM116",
                           normal_percentiles = c(10,90))
}

plot_annual_stats          Plot annual summary statistics (as lines)

Description

Plots means, medians, maximums, minimums, and percentiles for each year from all years of a
daily streamflow data set. Calculates statistics from all values, unless specified. Data calculated
using calc_annual_stats() function. Returns a list of plots.
Usage

plot_annual_stats(
  data,  
  dates = Date,  
  values = Value,  
  groups = STATION_NUMBER,  
  station_number,  
  percentiles,  
  roll_days = 1,  
  roll_align = "right",  
  water_year_start = 1,  
  start_year,  
  end_year,  
  exclude_years,  
  months = 1:12,  
  complete_years = FALSE,  
  ignore_missing = FALSE,  
  allowed_missing = ifelse(ignore_missing, 100, 0),  
  log_discharge = FALSE,  
  log_ticks = ifelse(log_discharge, TRUE, FALSE),  
  include_title = FALSE  
)

Arguments

data        Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates       Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values      Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups      Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
percentiles Numeric vector of percentiles to calculate. Set to NA if none required. Default NA.
roll_days   Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align  Character string identifying the direction of the rolling mean from the specified
date, either by the first ('left'), last ('right'), or middle ('center') day of
the rolling n-day group of observations. Default 'right'.

water_year_start  Numeric value indicating the month (1 through 12) of the start of water year for
analysis. Default 1.

start_year  Numeric value of the first year to consider for analysis. Leave blank or set well
before start date (i.e. 1800) to use from the first year of the source data.

day_end  Numeric value of the last year to consider for analysis. Leave blank or set well
after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years  Numeric vector of years to exclude from analysis. Leave blank or set to
NULL to include all years.

months  Numeric vector of months to include in analysis. For example, 3 for March, 6:8
for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start
= 10 (Oct). Default summarizes all months (1:12).

complete_years  Logical values indicating whether to include only years with complete data in
analysis. Default FALSE.

ignore_missing  Logical value indicating whether dates with missing values should be included
in the calculation. If TRUE then a statistic will be calculated regardless of missing
dates. If FALSE then only those statistics from time periods with no missing dates
will be returned. Default FALSE.

allowed_missing  Numeric value between 0 and 100 indicating the percentage of missing dates al-
lowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing
= FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing
= TRUE' then it defaults to 100 (any missing dates allowed); consistent with
ignore_missing usage. Supersedes ignore_missing when used.

log_discharge  Logical value to indicate plotting the discharge axis (Y-axis) on a logarithmic
scale. Default FALSE.

log_ticks  Logical value to indicate plotting logarithmic scale ticks when log_discharge
= TRUE. Ticks will not appear when log_discharge = FALSE. Default to TRUE
when log_discharge = TRUE.

include_title  Logical value to indicate adding the group/station number to the plot, if pro-
vided. Default FALSE.

Value

A list of ggplot2 objects for with the following plots (percentile plots optional) for each station
provided:

Annual_Stats  a plot that contains annual statistics

Default plots on each object:

Mean  annual mean of all daily flows

Median  annual median of all daily flows

Maximum  annual maximum of all daily flows

Minimum  annual minimum of all daily flows
plot_annual_stats2

See Also
calc_annual_stats

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Plot annual statistics using a data frame and data argument with defaults
flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
plot_annual_stats(data = flow_data)

# Plot annual statistics using station_number argument with defaults
plot_annual_stats(station_number = "08NM116")

# Plot annual statistics regardless if there is missing data for a given year
plot_annual_stats(station_number = "08NM116",
                 ignore_missing = TRUE)

# Plot annual statistics for water years starting in October
plot_annual_stats(station_number = "08NM116",
                 water_year_start = 10)

# Plot annual statistics with custom years and percentiles
plot_annual_stats(station_number = "08NM116",
                 start_year = 1981,
                 end_year = 2010,
                 percentiles = c(25,75))

}

plot_annual_stats2 Plot annual summary statistics (as ribbons)

Description

Plots means, medians, maximums, minimums, and percentiles as ribbons for each year from all years of a daily streamflow data set. Calculates statistics from all values, unless specified. Data calculated using calc_annual_stats() function. Returns a list of plots.

Usage

plot_annual_stats2(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
plot_annual_stats2

station_number,
roll_days = 1,
roll_align = "right",
water_year_start = 1,
start_year,
end_year,
exclude_years,
months = 1:12,
complete_years = FALSE,
ignore_missing = FALSE,
allowed_missing = ifelse(ignore_missing, 100, 0),
plot_extremes = TRUE,
plot_inner_percentiles = TRUE,
plot_outer_percentiles = TRUE,
inner_percentiles = c(25, 75),
outer_percentiles = c(5, 95),
log_discharge = TRUE,
log_ticks = ifelse(log_discharge, TRUE, FALSE),
include_title = FALSE
)

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
roll_days Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
start_year  Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year  Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years  Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months  Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

complete_years  Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

ignore_missing  Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

allowed_missing  Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used.

plot_extremes  Logical value to indicate plotting a ribbon with the range of daily minimum and maximum flows. Default TRUE.

plot_inner_percentiles  Logical value indicating whether to plot the inner percentiles ribbon. Default TRUE.

plot_outer_percentiles  Logical value indicating whether to plot the outer percentiles ribbon. Default TRUE.

inner_percentiles  Numeric vector of two percentile values indicating the lower and upper limits of the inner percentiles ribbon for plotting. Default c(25, 75), set to NULL for no inner ribbon.

outer_percentiles  Numeric vector of two percentile values indicating the lower and upper limits of the outer percentiles ribbon for plotting. Default c(5, 95), set to NULL for no outer ribbon.

log_discharge  Logical value to indicate plotting the discharge axis (Y-axis) on a logarithmic scale. Default FALSE.

log_ticks  Logical value to indicate plotting logarithmic scale ticks when log_discharge = TRUE. Ticks will not appear when log_discharge = FALSE. Default to TRUE when log_discharge = TRUE.

include_title  Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.
Value

A list of ggplot2 objects for with the following plots (percentile plots optional) for each station provided:

Annual_Stats    a plot that contains annual statistics

Default plots on each object:

Mean           annual mean
Median         annual median
25-75 Percentiles            a ribbon showing the range of data between the annual 25th and 75th percentiles
5-95 Percentiles            a ribbon showing the range of data between the annual 5th and 95th percentiles
Minimum-Maximum             a ribbon showing the range of data between the annual minimum and maximums

See Also

calc_annual_stats

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Plot annual statistics using a data frame and data argument with defaults
flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
plot_annual_stats2(data = flow_data)

# Plot annual statistics using station_number argument with defaults
plot_annual_stats2(station_number = "08NM116")

# Plot annual statistics regardless if there is missing data for a given year
plot_annual_stats2(station_number = "08NM116",
                   ignore_missing = TRUE)

# Plot annual statistics for water years starting in October
plot_annual_stats2(station_number = "08NM116",
                   water_year_start = 10)

}
plot_annual_symbols  Plot daily streamflow data symbols by year

Description

Plots data symbols for a daily data set by year, either by day of year, total days, or percent of year (see plot_type argument. A column of symbols is required, default symbols = 'Symbol'. For HYDAT data, symbols include: 'E' Estimate, 'A' Partial Day, 'B' Ice Conditions, 'D' Dry, and 'R' Revised. Other symbols or categories may be used to colour points of plot. Returns a list of plots.

Usage

plot_annual_symbols(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  symbols = Symbol,
  station_number,
  water_year_start = 1,
  start_year,
  end_year,
  months = 1:12,
  include_title = FALSE,
  plot_type = "dayofyear"
)

Arguments

data  Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.

dates  Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.

values  Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.

groups  Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.

symbols  Name of column in data that contains symbols. Only required if symbols column name is not 'Symbol' (default). Leave blank or set to NULL if using station_number argument.
`plot_annual_symbols`  

- **station_number**  
  Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

- **water_year_start**  
  Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

- **start_year**  
  Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

- **end_year**  
  Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

- **months**  
  Numeric vector of months to include in plotting. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default plots all months (1:12).

- **include_title**  
  Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

- **plot_type**  
  Character. One of c('dayofyear', 'count', 'percent'). With 'dayofyear' plot (default), the day of year for each year of data are coloured by symbols or missing dates are colours for each flow day of year. For 'count' and 'percent' plots, the total count or percent of all symbols or missing dates per year are displayed.

**Value**

A list of ggplot2 objects with the following for each station provided:

- **Annual_Symbols**  
  a plot that contains data symbols and missing dates

**Examples**

```r
# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

  # Plot annual symbol counts from a data frame and data argument
  flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
  plot_annual_symbols(data = flow_data)

  # Plot annual symbol counts using station_number argument with defaults
  plot_annual_symbols(station_number = "08NM116")

  # Plot annual symbol percentages using station_number argument and plot by annual counts
  plot_annual_symbols(station_number = "08NM116",
                      plot_type = "count")

} 
```
plot_daily_cumulative_stats

Plot cumulative daily flow statistics

Description

Plot the daily cumulative mean, median, maximum, minimum, and 5, 25, 75, 95th percentiles for each day of the year from a daily streamflow data set. Calculates statistics from all values from complete, unless specified. Data calculated using calc_daily_cumulative_stats() function. Can plot individual years for comparison using the add_year argument. Defaults to volumetric cumulative flows, can use use_yield and basin_area to convert to water yield. Returns a list of plots.

Usage

plot_daily_cumulative_stats(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  use_yield = FALSE,
  basin_area,
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  log_discharge = FALSE,
  log_ticks = ifelse(log_discharge, TRUE, FALSE),
  include_title = FALSE,
  add_year
)

Arguments

data: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values: Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.

station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

use_yield Logical value indicating whether to calculate area-based water yield, in mm, instead of volumetric discharge. Default FALSE.

basin_area Upstream drainage basin area, in square kilometres, to apply to observations. Three options:
(1) Leave blank if groups is STATION_NUMBER with HYDAT station numbers to extract basin areas from HYDAT.
(2) A single numeric value to apply to all observations.
(3) List each basin area for each group/station in groups (can override HYDAT value if listed) as such c("08NM116" = 795,"08NM242" = 10). If group is not listed the HYDAT area will be applied if it exists, otherwise it will be NA.

water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12). Need to be consecutive months for given year/water year to work properly.

log_discharge Logical value to indicate plotting the discharge axis (Y-axis) on a logarithmic scale. Default FALSE.

log_ticks Logical value to indicate plotting logarithmic scale ticks when log_discharge = TRUE. Ticks will not appear when log_discharge = FALSE. Default to TRUE when log_discharge = TRUE.

include_title Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

add_year Numeric value indicating a year of daily flows to add to the daily statistics plot. Leave blank or set to NULL for no years.

Value
A list of ggplot2 objects with the following for each station provided:
plot_daily_cumulative_stats

Daily_Cumulative_Stats

a plot that contains daily cumulative flow statistics

Default plots on each object:

Mean
daily cumulative mean

Median
daily cumulative median

Min-5 Percentile Range

a ribbon showing the range of data between the daily cumulative minimum and 5th percentile

5-25 Percentiles Range

a ribbon showing the range of data between the daily cumulative 5th and 25th percentiles

25-75 Percentiles Range

a ribbon showing the range of data between the daily cumulative 25th and 75th percentiles

75-95 Percentiles Range

a ribbon showing the range of data between the daily cumulative 75th and 95th percentiles

95 Percentile-Max Range

a ribbon showing the range of data between the daily cumulative 95th percentile and the maximum

‘Year’ Flows

(optional) the daily cumulative flows for the designated year

See Also

calc_daily_cumulative_stats

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Plot annual daily yield statistics with default HYDAT basin area
plot_daily_cumulative_stats(station_number = "08NM116",
                           use_yield = TRUE)

# Plot annual daily yield statistics with custom basin area
plot_daily_cumulative_stats(station_number = "08NM116",
                           use_yield = TRUE,
                           basin_area = 800)
}

plot_daily_stats

Plot daily summary statistics

Description

Plots means, medians, maximums, minimums, and percentiles for each day of the year of flow values from a daily streamflow data set. Can determine statistics of rolling mean days (e.g. 7-day flows) using the roll_days argument. Calculates statistics from all values, unless specified. The Maximum-Minimum band can be removed using the plot_extremes argument and the percentile bands can be customized using the inner_percentiles and outer_percentiles arguments. Data calculated using calc_daily_stats() function. Returns a list of plots.

Usage

plot_daily_stats(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  complete_years = FALSE,
  months = 1:12,
  ignore_missing = FALSE,
  plot_extremes = TRUE,
  plot_inner_percentiles = TRUE,
  plot_outer_percentiles = TRUE,
  inner_percentiles = c(25, 75),
  outer_percentiles = c(5, 95),
  add_year,
  log_discharge = TRUE,
  log_ticks = ifelse(log_discharge, TRUE, FALSE),
  include_title = FALSE
)

Arguments

data: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
### plot_daily_stats

**dates**  
Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.

**values**  
Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.

**groups**  
Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.

**station_number**  
Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

**roll_days**  
Numeric value of the number of days to apply a rolling mean. Default 1.

**roll_align**  
Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.

**water_year_start**  
Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

**start_year**  
Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

**end_year**  
Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

**exclude_years**  
Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

**complete_years**  
Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

**months**  
Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

**ignore_missing**  
Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

**plot_extremes**  
Logical value to indicate plotting a ribbon with the range of daily minimum and maximum flows. Default TRUE.

**plot_inner_percentiles**  
Logical value indicating whether to plot the inner percentiles ribbon. Default TRUE.

**plot_outer_percentiles**  
Logical value indicating whether to plot the outer percentiles ribbon. Default TRUE.
inner_percentiles
Numeric vector of two percentile values indicating the lower and upper limits of the inner percentiles ribbon for plotting. Default c(25, 75), set to NULL for no inner ribbon.

outer_percentiles
Numeric vector of two percentile values indicating the lower and upper limits of the outer percentiles ribbon for plotting. Default c(5, 95), set to NULL for no outer ribbon.

add_year
Numeric value indicating a year of daily flows to add to the daily statistics plot. Leave blank or set to NULL for no years.

log_discharge
Logical value to indicate plotting the discharge axis (Y-axis) on a logarithmic scale. Default FALSE.

log_ticks
Logical value to indicate plotting logarithmic scale ticks when log_discharge = TRUE. Ticks will not appear when log_discharge = FALSE. Default to TRUE when log_discharge = TRUE.

include_title
Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

Value
A list of ggplot2 objects with the following for each station provided:

Daily_Stats  a plot that contains daily flow statistics

Default plots on each object:

Mean  daily mean
Median  daily median

25-75 Percentiles  a ribbon showing the range of data between the daily 25th and 75th percentiles

5-95 Percentiles  a ribbon showing the range of data between the daily 5th and 95th percentiles

Minimum-Maximum  a ribbon showing the range of data between the daily minimum and maximums

'Year'  (on annual plots) the daily flows for the designated year

See Also
calc_daily_stats

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Plot daily statistics using a data frame and data argument with defaults
flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
plot_daily_stats(data = flow_data,
# Plot daily statistics using only years with no missing data
plot_daily_stats(station_number = "08NM116",
     complete_years = TRUE)

# Plot daily statistics and add a specific year's daily flows
plot_daily_stats(station_number = "08NM116",
     start_year = 1980,
     add_year = 1985)

# Plot daily statistics for 7-day flows for July-September months only
plot_daily_stats(station_number = "08NM116",
     start_year = 1980,
     roll_days = 7,
     months = 7:9)

plot_data_screening

Plot annual summary statistics for data screening

Description

Plots the mean, median, maximum, minimum, standard deviation of annual flows and indicates data availability. Calculates statistics from all values, unless specified. Data calculated using screen_flow_data() function. Returns a list of plots.

Usage

plot_data_screening(
    data,
    dates = Date,
    values = Value,
    groups = STATION_NUMBER,
    station_number,
    roll_days = 1,
    roll_align = "right",
    water_year_start = 1,
    months = 1:12,
    start_year,
    end_year,
    include_title = FALSE,
    plot_availability = TRUE,
    include_stats = c("Mean", "Median", "Minimum", "Maximum", "Standard Deviation")
)
Arguments

- **data**: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
- **dates**: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
- **values**: Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
- **groups**: Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
- **station_number**: Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
- **roll_days**: Numeric value of the number of days to apply a rolling mean. Default 1.
- **roll_align**: Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
- **water_year_start**: Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
- **months**: Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).
- **start_year**: Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.
- **end_year**: Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.
- **include_title**: Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.
- **plot_availability**: Logical value specifying whether to indicate if years contain complete data or missing values. Default TRUE. Use FALSE for original fasstr version.
- **include_stats**: Vector of one or all of c("Mean", "Median", "Minimum", "Maximum", "Standard Deviation") to list annual summary statistics to plot for screening. Default all.

Value

A list of ggplot2 objects with the following for each station provided:
Data_Screening  a plot that contains annual summary statistics for screening

Default plots on each object:

Minimum  annual minimum of all daily flows for a given year
Maximum  annual maximum of all daily flows for a given year
Mean  annual mean of all daily flows for a given year
StandardDeviation  annual 1 standard deviation of all daily flows for a given year

See Also

screen_flow_data

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

  # Plot screening statistics using a data frame and data argument with defaults
  flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
  plot_data_screening(data = flow_data)

  # Plot screening statistics using station_number argument with defaults
  plot_data_screening(station_number = "08NM116")

  # Plot screening statistics for water years starting in October
  plot_data_screening(station_number = "08NM116",
                      water_year_start = 10)

  # Plot screening statistics for 7-day flows for July-September months only
  plot_data_screening(station_number = "08NM116",
                      roll_days = 7,
                      months = 7:9)
}

plot_flow_data  

Plot a daily streamflow data set

Description

Plot the daily mean flow values from a streamflow data set. Plots daily discharge values from all
years, unless specified. Can choose specific dates to start and end plotting. Can choose to plot out
each year separately. Multiple groups/stations can be plotted if provided with the groups argument.
Returns a list of plots.
plot_flow_data

Usage

plot_flow_data(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  start_date,
  end_date,
  log_discharge = FALSE,
  log_ticks = ifelse(log_discharge, TRUE, FALSE),
  plot_by_year = FALSE,
  one_plot = FALSE,
  include_title = FALSE
)

Arguments

data          Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates         Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values        Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups        Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
roll_days     Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align    Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
water_year_start
Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year
Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year
Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years
Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months
Numeric vector of months to include in plotting. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default plots all months (1:12).

start_date
Date (YYYY-MM-DD) of first date to consider for plotting. Leave blank if all years are required.

end_date
Date (YYYY-MM-DD) of last date to consider for plotting. Leave blank if all years are required.

log_discharge
Logical value to indicate plotting the discharge axis (Y-axis) on a logarithmic scale. Default TRUE.

log_ticks
Logical value to indicate plotting logarithmic scale ticks when using a log-scale discharge axis. Default to FALSE when log_discharge = FALSE and TRUE when log_discharge = TRUE.

plot_by_year
Logical value to indicate whether to plot each year of data individually. Default FALSE.

one_plot
Logical value to indicate whether to plot all groups/stations on one plot. Default FALSE.

include_title
Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

Value
A ggplot2 object of daily flows from flow_data or HYDAT flow data provided

Examples
# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Plot data from a data frame and data argument
flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
plot_flow_data(data = flow_data)

# Plot data directly from HYDAT
plot_flow_data(station_number = "08NM116")

# Plot statistics with custom years
plot_flow_data(station_number = "08NM116",
start_year = 1981,
end_year = 2010,

# Plot data multiple groups on one plot
plot_flow_data(station_number = c("08NM241", "08NM242"),
one_plot = TRUE)

# Plot data between specific dates
plot_flow_data(station_number = "08NM116",
start_date = "1990-01-01",
end_date = "1990-06-01")

plot_flow_data_symbols

*Plot daily streamflow data with their symbols*

**Description**

Plots data symbols for a daily data set. A column of symbols is required, default `symbols = 'Symbol'`. For HYDAT data, symbols include: 'E' Estimate, 'A' Partial Day, 'B' Ice Conditions, 'D' Dry, and 'R' Revised. Other symbols or categories may be used to colour points of plot. Returns a list of plots.

**Usage**

```r
plot_flow_data_symbols(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  symbols = Symbol,
  station_number,
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  start_date,
  end_date,
  log_discharge = FALSE,
  include_title = FALSE
)
```
Arguments

data: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g., station numbers). Leave blank or set to NULL if using station_number argument.

dates: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.

values: Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.

groups: Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.

symbols: Name of column in data that contains symbols. Only required if symbols column name is not 'Symbol' (default). Leave blank or set to NULL if using station_number argument.

station_number: Character string vector of seven digit Water Survey of Canada station numbers (e.g., "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

water_year_start: Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year: Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e., 1800) to use from the first year of the source data.

end_year: Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e., 2100) to use up to the last year of the source data.

exclude_years: Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months: Numeric vector of months to include in plotting. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default plots all months (1:12).

start_date: Date (YYYY-MM-DD) of first date to consider for plotting. Leave blank if all years are required.

end_date: Date (YYYY-MM-DD) of last date to consider for plotting. Leave blank if all years are required.

log_discharge: Logical value to indicate plotting the discharge axis (Y-axis) on a logarithmic scale. Default TRUE.

include_title: Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.
plot_flow_duration

Value

A list of ggplot2 objects with the following for each station provided:

Flow_Data_Symbols
   a plot that contains the flow data with symbol categories

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

   # Plot data and symbols from a data frame and data argument
   flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
   plot_flow_data_symbols(data = flow_data)

   # Plot data and symbols using station_number argument with defaults
   plot_flow_data_symbols(station_number = "08NM116")

}

plot_flow_duration      Plot flow duration curves

Description

Plots flow duration curves of flow data from a daily streamflow data set. Plots the percent time flows are equalled or exceeded. Calculates statistics from all values, unless specified. Data calculated using calc_longterm_stats() function then converted for plotting. Returns a list of plots.

Usage

plot_flow_duration(
   data,
   dates = Date,
   values = Value,
   groups = STATION_NUMBER,
   station_number,
   roll_days = 1,
   roll_align = "right",
   water_year_start = 1,
   start_year,
   end_year,
   exclude_years,
   custom_months,
   custom_months_label,
   complete_years = FALSE,
   ignore_missing = FALSE,
   months = 1:12,
include_longterm = TRUE,
log_discharge = TRUE,
log_ticks = ifelse(log_discharge, TRUE, FALSE),
include_title = FALSE
)

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
roll_days Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
start_year Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.
end_year Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.
exclude_years Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.
custom_months Numeric vector of months to combine to summarize (ex. 6:8 for Jun-Aug). Adds results to the end of table. If wanting months that overlap calendar years (ex. Oct-Mar), choose water_year_start that begins before the first month listed. Leave blank for no custom month summary.
custom_months_label Character string to label custom months. For example, if months = 7:9 you may choose "Summer" or "Jul-Sep". Default "Custom-Months".
plot_flow_duration

complete_years Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

ignore_missing Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

months Numeric vector of month curves to plot. NA if no months required. Default 1:12.

include_longterm Logical value indicating whether to include long-term curve of all data. Default TRUE.

log_discharge Logical value to indicate plotting the discharge axis (Y-axis) on a logarithmic scale. Default FALSE.

log_ticks Logical value to indicate plotting logarithmic scale ticks when log_discharge = TRUE. Ticks will not appear when log_discharge = FALSE. Default to TRUE when log_discharge = TRUE.

include_title Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

Value

A list of ggplot2 objects with the following for each station provided:

Flow_Duration a plot that contains flow duration curves for each month, long-term, and (option) customized months

See Also
calc_longterm_daily_stats

Examples

```r
## Not run:

# Working examples:

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {
  # Plot flow durations using a data frame and data argument with defaults
  flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
  plot_flow_duration(data = flow_data, start_year = 1980)

  # Plot flow durations using station_number argument with defaults
  plot_flow_duration(station_number = "08NM116",
                     start_year = 1980)

  # Plot flow durations and add custom stats for July-September
  plot_flow_duration(station_number = "08NM116",
                     start_year = 1980, include_title = TRUE)
```
plot_longterm_daily_stats

Plot long-term summary statistics from daily mean flows

Description

Plots the long-term mean, median, maximum, minimum, and percentiles of daily flow values for over all months and all data (Long-term) from a daily streamflow data set. Calculates statistics from all values, unless specified. The Maximum-Minimum band can be removed using the plot_extremes argument and the percentile bands can be customized using the inner_percentiles and outer_percentiles arguments. Data calculated using the calc_longterm_daily_stats() function. Returns a list of plots.

Usage

```r
plot_longterm_daily_stats(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  complete_years = FALSE,
  ignore_missing = FALSE,
  plot_extremes = TRUE,
  plot_inner_percentiles = TRUE,
  plot_outer_percentiles = TRUE,
  inner_percentiles = c(25, 75),
  outer_percentiles = c(5, 95),
  add_year,
  log_discharge = TRUE,
  log_ticks = ifelse(log_discharge, TRUE, FALSE),
  include_title = FALSE
)
```
Arguments

- **data**: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using `station_number` argument.

- **dates**: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using `station_number` argument.

- **values**: Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using `station_number` argument.

- **groups**: Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using `station_number` argument.

- **station_number**: Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

- **roll_days**: Numeric value of the number of days to apply a rolling mean. Default 1.

- **roll_align**: Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.

- **water_year_start**: Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

- **start_year**: Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

- **end_year**: Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

- **exclude_years**: Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

- **months**: Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when `water_year_start` = 10 (Oct). Default summarizes all months (1:12).

- **complete_years**: Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

- **ignore_missing**: Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

- **plot_extremes**: Logical value to indicate plotting a ribbon with the range of daily minimum and maximum flows. Default TRUE.
plot_longterm_daily_stats

plot_inner_percentiles
Logical value indicating whether to plot the inner percentiles ribbon. Default TRUE.

plot_outer_percentiles
Logical value indicating whether to plot the outer percentiles ribbon. Default TRUE.

inner_percentiles
Numeric vector of two percentile values indicating the lower and upper limits of the inner percentiles ribbon for plotting. Default c(25,75), set to NULL for no inner ribbon.

outer_percentiles
Numeric vector of two percentile values indicating the lower and upper limits of the outer percentiles ribbon for plotting. Default c(5,95), set to NULL for no outer ribbon.

add_year
Numeric value indicating a year of daily flows to add to the daily statistics plot. Leave blank or set to NULL for no years.

log_discharge
Logical value to indicate plotting the discharge axis (Y-axis) on a logarithmic scale. Default FALSE.

log_ticks
Logical value to indicate plotting logarithmic scale ticks when log_discharge = TRUE. Ticks will not appear when log_discharge = FALSE. Default to TRUE when log_discharge = TRUE.

include_title
Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

Value
A list of ggplot2 objects with the following for each station provided:

Long-term_Monthly_Statistics
a plot that contains long-term flow statistics

Default plots on each object:

Monthly Mean
mean of all annual monthly means for a given month over all years

Monthly Median
median of all annual monthly means for a given month over all years

25-75 Percentiles Range
a ribbon showing the range of data between the monthly 25th and 75th percentiles

5-95 Percentiles Range
a ribbon showing the range of data between the monthly 5th and 95th percentiles

Max-Min Range
a ribbon showing the range of data between the monthly minimum and maximums

See Also
calc_longterm_daily_stats
Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

  # Plot longterm daily statistics using data argument with defaults
  flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
  plot_longterm_daily_stats(data = flow_data,
                             start_year = 1980)

  # Plot longterm daily statistics for water years starting in October
  plot_longterm_daily_stats(station_number = "08NM116",
                             start_year = 1980,
                             end_year = 2010,
                             water_year_start = 10)

}

plot_longterm_monthly_stats

Plot long-term summary statistics from annual monthly mean flows

Description

Plots the long-term mean, median, maximum, minimum, and percentiles of annual monthly mean
flow values for all months and all data (Long-term) from a daily streamflow data set. Calculates
statistics from all values, unless specified. The Maximum-Minimum band can be removed using the
plot_extremes argument and the percentile bands can be customized using the inner_percentiles
and outer_percentiles arguments. Data calculated using the calc_longterm_monthly_stats() function. Returns a list of plots.

Usage

plot_longterm_monthly_stats(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  complete_years = FALSE,
  ignore_missing = FALSE,
  plot_extremes = TRUE,
plot_longterm_monthly_stats

plot_inner_percentiles = TRUE,
plot_outer_percentiles = TRUE,
inner_percentiles = c(25, 75),
outer_percentiles = c(5, 95),
add_year,
log_discharge = TRUE,
log_ticks = ifelse(log_discharge, TRUE, FALSE),
include_title = FALSE

Arguments

data          Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates         Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values        Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups        Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "00NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
roll_days     Numeric value of the number of days to apply a rolling mean. Default 1.
roll_align    Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.
water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
start_year    Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.
end_year      Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.
exclude_years Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.
months        Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).
**plot_longterm_monthly_stats**

- **complete_years**: Logical values indicating whether to include only years with complete data in analysis. Default FALSE.
- **ignore_missing**: Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.
- **plot_extremes**: Logical value to indicate plotting a ribbon with the range of daily minimum and maximum flows. Default TRUE.
- **plot_inner_percentiles**: Logical value indicating whether to plot the inner percentiles ribbon. Default TRUE.
- **plot_outer_percentiles**: Logical value indicating whether to plot the outer percentiles ribbon. Default TRUE.
- **inner_percentiles**: Numeric vector of two percentile values indicating the lower and upper limits of the inner percentiles ribbon for plotting. Default c(25, 75), set to NULL for no inner ribbon.
- **outer_percentiles**: Numeric vector of two percentile values indicating the lower and upper limits of the outer percentiles ribbon for plotting. Default c(5, 95), set to NULL for no outer ribbon.
- **add_year**: Numeric value indicating a year of daily flows to add to the daily statistics plot. Leave blank or set to NULL for no years.
- **log_discharge**: Logical value to indicate plotting the discharge axis (Y-axis) on a logarithmic scale. Default FALSE.
- **log_ticks**: Logical value to indicate plotting logarithmic scale ticks when log_discharge = TRUE. Ticks will not appear when log_discharge = FALSE. Default to TRUE when log_discharge = TRUE.
- **include_title**: Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

**Value**

A list of ggplot2 objects with the following for each station provided:

- **Long-term_Monthly_Statistics**: a plot that contains long-term flow statistics

Default plots on each object:

- **Monthly Mean**: mean of all annual monthly means for a given month over all years
- **Monthly Median**: median of all annual monthly means for a given month over all years
- **25-75 Percentiles Range**: a ribbon showing the range of data between the monthly 25th and 75th percentiles
- **5-95 Percentiles Range**: a ribbon showing the range of data between the monthly 5th and 95th percentiles
Max-Min Range  
a ribbon showing the range of data between the monthly minimum and maximums

See Also

calc_longterm_monthly_stats

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Plot longterm monthly statistics using station_number argument with defaults
plot_longterm_monthly_stats(station_number = "08NM116",
start_year = 1980)

# Plot longterm monthly statistics and add a specific year's daily flows
plot_longterm_monthly_stats(station_number = "08NM116",
start_year = 1980,
add_year = 1985)

}

plot_missing_dates  
Plot annual and monthly missing dates

Description

Plots the data availability for each month of each year. Calculates statistics from all values, unless specified. Data calculated using screen_flow_data() function. Returns a list of plots.

Usage

plot_missing_dates(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  months = 1:12,
  include_title = FALSE,
  plot_type = "tile"
)
### Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
<td>Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.</td>
</tr>
<tr>
<td>dates</td>
<td>Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.</td>
</tr>
<tr>
<td>values</td>
<td>Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.</td>
</tr>
<tr>
<td>groups</td>
<td>Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.</td>
</tr>
<tr>
<td>station_number</td>
<td>Character string vector of seven digit Water Survey of Canada station numbers (e.g. &quot;08NM116&quot;) of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.</td>
</tr>
<tr>
<td>roll_days</td>
<td>Numeric value of the number of days to apply a rolling mean. Default 1.</td>
</tr>
<tr>
<td>roll_align</td>
<td>Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.</td>
</tr>
<tr>
<td>water_year_start</td>
<td>Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.</td>
</tr>
<tr>
<td>start_year</td>
<td>Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.</td>
</tr>
<tr>
<td>end_year</td>
<td>Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.</td>
</tr>
<tr>
<td>months</td>
<td>Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).</td>
</tr>
<tr>
<td>include_title</td>
<td>Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.</td>
</tr>
<tr>
<td>plot_type</td>
<td>Type of missing data plot, either &quot;tile&quot; or &quot;bar&quot; styles. Default &quot;tile&quot;. Use &quot;bar&quot; for previous version of plot.</td>
</tr>
</tbody>
</table>

### Value

A list of ggplot2 objects with the following for each station provided:

- **Missing_Dates** a plot that contains the data availability for each year and month
plot_monthly_cumulative_stats

Plot cumulative monthly flow statistics

Description

Plot the monthly cumulative mean, median, maximum, minimum, and 5, 25, 75, 95th percentiles for each month of the year from a daily streamflow data set. Calculates statistics from all values from complete years, unless specified. Data calculated using calc_monthly_cumulative_stats() function. Can plot individual years for comparison using the add_year argument. Defaults to volumetric cumulative flows, can use use_yield and basin_area to convert to water yield. Returns a list of plots.

Usage

plot_monthly_cumulative_stats(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  use_yield = FALSE,
basin_area,
water_year_start = 1,
start_year,
end_year,
elude_years,
months = 1:12,
log_discharge = FALSE,
log_ticks = ifelse(log_discharge, TRUE, FALSE),
include_title = FALSE,
add_year
)

## Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
groups Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.
station_number Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.
use_yield Logical value indicating whether to calculate area-based water yield, in mm, instead of volumetric discharge. Default FALSE.
basin_area Upstream drainage basin area, in square kilometres, to apply to observations. Three options:
(1) Leave blank if groups is STATION_NUMBER with HYDAT station numbers to extract basin areas from HYDAT.
(2) A single numeric value to apply to all observations.
(3) List each basin area for each group/station in groups (can override HYDAT value if listed) as such c("08NM116" = 795,"08NM242" = 10). If group is not listed the HYDAT area will be applied if it exists, otherwise it will be NA.
water_year_start Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.
start_year Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.
plot_monthly_cumulative_stats

end_year  
Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years  
Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months  
Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12). Need to be consecutive months for given year/water year to work properly.

log_discharge  
Logical value to indicate plotting the discharge axis (Y-axis) on a logarithmic scale. Default FALSE.

log_ticks  
Logical value to indicate plotting logarithmic scale ticks when log_discharge = TRUE. Ticks will not appear when log_discharge = FALSE. Default to TRUE when log_discharge = TRUE.

include_title  
Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

add_year  
Numeric value indicating a year of daily flows to add to the daily statistics plot. Leave blank or set to NULL for no years.

Value
A list of ggplot2 objects with the following for each station provided:

Monthly_Cumulative_Stats  
a plot that contains monthly cumulative flow statistics

Default plots on each object:

Mean  
monthly cumulative mean

Median  
monthly cumulative median

Min-5 Percentile Range  
a ribbon showing the range of data between the monthly cumulative minimum and 5th percentile

5-25 Percentiles Range  
a ribbon showing the range of data between the monthly cumulative 5th and 25th percentiles

25-75 Percentiles Range  
a ribbon showing the range of data between the monthly cumulative 25th and 75th percentiles

75-95 Percentiles Range  
a ribbon showing the range of data between the monthly cumulative 75th and 95th percentiles

95 Percentile-Max Range  
a ribbon showing the range of data between the monthly cumulative 95th percentile and the maximum

'Year' Flows  
(optional) the monthly cumulative flows for the designated year
plot_monthly_means

Plot monthly means and percent LTMADs

Description

Plot monthly means and add long-term mean annual discharge percentages. Calculates statistics from all values, unless specified. Mean data calculated using calc_longterm_daily_stats() function. Returns a list of plots.

Usage

plot_monthly_means(
    data,
    dates = Date,
    values = Value,
    groups = STATION_NUMBER,
    station_number,
    roll_days = 1,
    roll_align = "right",
    water_year_start = 1,
    start_year,
    end_year,
    exclude_years,
    months = 1:12,
    plot_months = 1:12,
    complete_years = FALSE,
ignore_missing = FALSE,
include_title = FALSE,
percent_MAD = c(10, 20, 100)
)

Arguments

**data**
Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using `station_number` argument.

**dates**
Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using `station_number` argument.

**values**
Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using `station_number` argument.

**groups**
Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using `station_number` argument.

**station_number**
Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using `data` argument.

**roll_days**
Numeric value of the number of days to apply a rolling mean. Default 1.

**roll_align**
Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.

**water_year_start**
Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

**start_year**
Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

**end_year**
Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

**exclude_years**
Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

**months**
Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12, 1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).

**plot_months**
Numeric vector of months to include on the plot after calculating statistics. For example, 3 for March or 6:8 for Jun-Aug. differs from 'months' argument where that argument filters for specific months, this one just chooses which months to plot. Default 1:12.
plot_monthly_means

complete_years Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

group_missing Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.

include_title Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

percent_MAD Numeric vector of percentages of long-term mean annual discharge to add to the plot (ex. 20 for 20 percent MAD or c(5, 10, 20) for multiple percentages). Set to NA for none. Default c(10, 20, 100).

Value

A list of ggplot2 objects for with the following plots for each station provided:

Annual_Means a plot that contains annual means with the long-term mean as the x-axis intercept

See Also
calc_longterm_daily_stats
calc_longterm_mean

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {
  
  # Plot monthly means
  plot_monthly_means(station_number = "08NM116",
                     complete_years = TRUE)

  # Plot mean flows with custom LTMADs
  plot_monthly_means(station_number = "08NM116",
                     complete_years = TRUE,
                     percent_MAD = c(5, 10, 20, 100))

  # Plot mean flows and plot just summer months
  plot_monthly_means(station_number = "08NM116",
                     complete_years = TRUE,
                     plot_months = 6:9)

}
plot_monthly_stats  

Plot monthly summary statistics

Description

Plots means, medians, maximums, minimums, and percentiles for each month of all years of flow values from a daily streamflow data set. Calculates statistics from all values, unless specified. Data calculated using the calc_monthly_stats() function. Produces a list containing a plot for each statistic. Returns a list of plots.

Usage

```r
plot_monthly_stats(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  percentiles,
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  complete_years = FALSE,
  ignore_missing = FALSE,
  allowed_missing = ifelse(ignore_missing, 100, 0),
  log_discharge = FALSE,
  log_ticks = ifelse(log_discharge, TRUE, FALSE),
  scales_discharge = "fixed",
  include_title = FALSE
)
```

Arguments

- **data**: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
- **dates**: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
- **values**: Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.
**plot_monthly_stats**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>groups</code></td>
<td>Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using <code>station_number</code> argument.</td>
</tr>
<tr>
<td><code>station_number</code></td>
<td>Character string vector of seven digit Water Survey of Canada station numbers (e.g. &quot;08NM116&quot;) of which to extract daily streamflow data from a HYDAT database. Requires <code>tidyhydat</code> package and a HYDAT database. Leave blank if using data argument.</td>
</tr>
<tr>
<td><code>percentiles</code></td>
<td>Numeric vector of percentiles to calculate. Set to NA if none required. Default NA.</td>
</tr>
<tr>
<td><code>roll_days</code></td>
<td>Numeric value of the number of days to apply a rolling mean. Default 1.</td>
</tr>
<tr>
<td><code>roll_align</code></td>
<td>Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.</td>
</tr>
<tr>
<td><code>water_year_start</code></td>
<td>Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.</td>
</tr>
<tr>
<td><code>start_year</code></td>
<td>Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.</td>
</tr>
<tr>
<td><code>end_year</code></td>
<td>Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.</td>
</tr>
<tr>
<td><code>exclude_years</code></td>
<td>Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.</td>
</tr>
<tr>
<td><code>months</code></td>
<td>Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when <code>water_year_start</code> = 10 (Oct). Default summarizes all months (1:12).</td>
</tr>
<tr>
<td><code>complete_years</code></td>
<td>Logical values indicating whether to include only years with complete data in analysis. Default FALSE.</td>
</tr>
<tr>
<td><code>ignore_missing</code></td>
<td>Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.</td>
</tr>
<tr>
<td><code>allowed_missing</code></td>
<td>Numeric value between 0 and 100 indicating the percentage of missing dates allowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing' = FALSE then it defaults to 0 (zero missing dates allowed), if 'ignore_missing' = TRUE then it defaults to 100 (any missing dates allowed); consistent with <code>ignore_missing</code> usage. Supersedes <code>ignore_missing</code> when used.</td>
</tr>
<tr>
<td><code>log_discharge</code></td>
<td>Logical value to indicate plotting the discharge axis (Y-axis) on a logarithmic scale. Default FALSE.</td>
</tr>
<tr>
<td><code>log_ticks</code></td>
<td>Logical value to indicate plotting logarithmic scale ticks when <code>log_discharge</code> = TRUE. Ticks will not appear when <code>log_discharge</code> = FALSE. Default to TRUE when <code>log_discharge</code> = TRUE.</td>
</tr>
</tbody>
</table>
scales_discharge String, either 'fixed' (all y-axis scales the same) or 'free' (each plot has their own scale). Default 'fixed'.

include_title Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

Value
A list of ggplot2 objects for each monthly statistic for each station provided that contain:

Monthly Mean Flows
mean of all daily flows for a given month and year

Monthly Median Flows
median of all daily flows for a given month and year

Monthly Maximum Flows
maximum of all daily flows for a given month and year

Monthly Minimum Flows
minimum of all daily flows for a given month and year

Monthly P’n’ Flows
(optional) each n-th percentile selected for a given month and year

See Also
calc_monthly_stats

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

  # Plot monthly statistics using a data frame and data argument with defaults
  flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
  results <- plot_monthly_stats(data = flow_data,
                               start_year = 1980,
                               percentiles = 10)

  # Plot monthly statistics for water years starting in October
  results <- plot_monthly_stats(station_number = "08NM116",
                                start_year = 1980,
                                end_year = 2010,
                                water_year_start = 10,
                                percentiles = 10)
}

}
Description

Plots means, medians, maximums, minimums, and percentiles as ribbons for each month of all years of flow values from a daily streamflow data set. Calculates statistics from all values, unless specified. Data calculated using the `calc_monthly_stats()` function. Produces a list containing a plot for each statistic. Returns a list of plots.

Usage

```r
plot_monthly_stats2(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  complete_years = FALSE,
  ignore_missing = FALSE,
  allowed_missing = ifelse(ignore_missing, 100, 0),
  plot_extremes = TRUE,
  plot_inner_percentiles = TRUE,
  plot_outer_percentiles = TRUE,
  inner_percentiles = c(25, 75),
  outer_percentiles = c(5, 95),
  log_discharge = TRUE,
  log_ticks = ifelse(log_discharge, TRUE, FALSE),
  scales_discharge = "fixed",
  include_title = FALSE
)
```

Arguments

- **data**: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using `station_number` argument.
- **dates**: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using `station_number` argument.
values
Name of column in data that contains numeric flow values, in units of cubic
metres per second. Only required if values column name is not 'Value' (default).
Leave blank if using station_number argument.

groups
Name of column in data that contains unique identifiers for different data sets, if
applicable. Only required if groups column name is not 'STATION_NUMBER'.
Function will automatically group by a column named 'STATION_NUMBER'
if present. Remove the 'STATION_NUMBER' column beforehand to remove
this grouping. Leave blank if using station_number argument.

station_number
Character string vector of seven digit Water Survey of Canada station numbers
(e.g. "08NM116") of which to extract daily streamflow data from a HYDAT
database. Requires tidyhydat package and a HYDAT database. Leave blank if
using data argument.

roll_days
Numeric value of the number of days to apply a rolling mean. Default 1.

roll_align
Character string identifying the direction of the rolling mean from the specified
date, either by the first ('left'), last ('right'), or middle ('center') day of
the rolling n-day group of observations. Default 'right'.

water_year_start
Numeric value indicating the month (1 through 12) of the start of water year for
analysis. Default 1.

start_year
Numeric value of the first year to consider for analysis. Leave blank or set well
before start date (i.e. 1800) to use from the first year of the source data.

end_year
Numeric value of the last year to consider for analysis. Leave blank or set well
after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years
Numeric vector of years to exclude from analysis. Leave blank or set to NULL to
include all years.

months
Numeric vector of months to include in analysis. For example, 3 for March, 6:8
for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start
= 10 (Oct). Default summarizes all months (1:12).

complete_years
Logical values indicating whether to include only years with complete data in
analysis. Default FALSE.

ignore_missing
Logical value indicating whether dates with missing values should be included
in the calculation. If TRUE then a statistic will be calculated regardless of missing
dates. If FALSE then only those statistics from time periods with no missing dates
will be returned. Default FALSE.

allowed_missing
Numeric value between 0 and 100 indicating the percentage of missing dates al-
lowed to be included to calculate a statistic (0 to 100 percent). If 'ignore_missing
= FALSE ' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing
= TRUE' then it defaults to 100 (any missing dates allowed); consistent with
ignore_missing usage. Supersedes ignore_missing when used.

plot_extremes
Logical value to indicate plotting a ribbon with the range of daily minimum and
maximum flows. Default TRUE.

plot_inner_percentiles
Logical value indicating whether to plot the inner percentiles ribbon. Default
TRUE.
plot_monthly_stats2

plot_outer_percentiles
   Logical value indicating whether to plot the outer percentiles ribbon. Default TRUE.

inner_percentiles
   Numeric vector of two percentile values indicating the lower and upper limits of the inner percentiles ribbon for plotting. Default c(25,75), set to NULL for no inner ribbon.

outer_percentiles
   Numeric vector of two percentile values indicating the lower and upper limits of the outer percentiles ribbon for plotting. Default c(5,95), set to NULL for no outer ribbon.

log_discharge
   Logical value to indicate plotting the discharge axis (Y-axis) on a logarithmic scale. Default FALSE.

log_ticks
   Logical value to indicate plotting logarithmic scale ticks when log_discharge = TRUE. Ticks will not appear when log_discharge = FALSE. Default to TRUE when log_discharge = TRUE.

scales_discharge
   String, either 'fixed' (all y-axis scales the same) or 'free' (each plot has their own scale). Default 'fixed'.

include_title
   Logical value to indicate adding the group/station number to the plot, if provided. Default FALSE.

Value

A list of ggplot2 objects for each monthly statistic for each station provided that contain:

Monthly Mean Flows
   mean of all daily flows for a given month and year

Monthly Median Flows
   median of all daily flows for a given month and year

Monthly Maximum Flows
   maximum of all daily flows for a given month and year

Monthly Minimum Flows
   minimum of all daily flows for a given month and year

Monthly P’n’ Flows
   (optional) each n-th percentile selected for a given month and year

See Also

calc_monthly_stats

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Plot monthly statistics using a data frame and data argument with defaults
flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")

results <- plot_monthly_stats2(data = flow_data,
                           start_year = 1980)

# Plot monthly statistics for water years starting in October
results <- plot_monthly_stats2(station_number = "08NM116",
                           start_year = 1980,
                           end_year = 2010,
                           water_year_start = 10)

---

**screen_flow_data**

*Calculate annual summary and missing data statistics for screening data*

**Description**

Calculates means, medians, maximums, minimums, standard deviations of annual flows and data availability and missing data statistics, and symbol counts (if column exists) for each year and month of each year. Calculates the statistics from all daily discharge values from all years, unless specified. Returns a tibble with statistics.

**Usage**

```r
screen_flow_data(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  symbols = "Symbol",
  station_number,
  roll_days = 1,
  roll_align = "right",
  water_year_start = 1,
  start_year,
  end_year,
  months = 1:12,
  transpose = FALSE,
  include_symbols = TRUE
)
```

**Arguments**

- **data**: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
- **dates**: Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
**screen_flow_data**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>values</strong></td>
<td>Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.</td>
</tr>
<tr>
<td><strong>groups</strong></td>
<td>Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.</td>
</tr>
<tr>
<td><strong>symbols</strong></td>
<td>Name of column in data that contains symbols. Only required if symbols column name is not 'Symbol' (default). Leave blank or set to NULL if using station_number argument.</td>
</tr>
<tr>
<td><strong>station_number</strong></td>
<td>Character string vector of seven digit Water Survey of Canada station numbers (e.g. &quot;08NM116&quot;) of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.</td>
</tr>
<tr>
<td><strong>roll_days</strong></td>
<td>Numeric value of the number of days to apply a rolling mean. Default 1.</td>
</tr>
<tr>
<td><strong>roll_align</strong></td>
<td>Character string identifying the direction of the rolling mean from the specified date, either by the first ('left'), last ('right'), or middle ('center') day of the rolling n-day group of observations. Default 'right'.</td>
</tr>
<tr>
<td><strong>water_year_start</strong></td>
<td>Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.</td>
</tr>
<tr>
<td><strong>start_year</strong></td>
<td>Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.</td>
</tr>
<tr>
<td><strong>end_year</strong></td>
<td>Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.</td>
</tr>
<tr>
<td><strong>months</strong></td>
<td>Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12).</td>
</tr>
<tr>
<td><strong>transpose</strong></td>
<td>Logical value indicating whether to transpose rows and columns of results. Default FALSE.</td>
</tr>
<tr>
<td><strong>include_symbols</strong></td>
<td>Logical. Include columns of counts of symbol categories from the symbols column.</td>
</tr>
</tbody>
</table>

**Value**

A tibble data frame with the following columns:

- **Year**: calendar or water year selected
- **n_days**: number of days per year
- **n_Q**: number of days per year with flow data
- **n_missing_Q**: number of days per year with no flow data
- **No_Symbol**: number of days with no symbol category, if include_symbol=TRUE
screen_flow_data

x_Symbol
number of days with a specific symbol category (x) from symbols column, if include_symbol=TRUE

Maximum
annual maximum of all daily flows for a given year

Mean
annual mean of all daily flows for a given year

Median
annual median of all daily flows for a given year

StandardDeviation
annual 1 standard deviation of all daily flows for a given year

and the following monthly missing columns (order will depend on water_year_month):

Jan_missing_Q
number of Jan days per year with no flow data

Feb_missing_Q
number of Feb days per year with no flow data

Mar_missing_Q
number of Mar days per year with no flow data

Apr_missing_Q
number of Apr days per year with no flow data

May_missing_Q
number of May days per year with no flow data

Jun_missing_Q
number of Jun days per year with no flow data

Jul_missing_Q
number of Jul days per year with no flow data

Aug_missing_Q
number of Aug days per year with no flow data

Sep_missing_Q
number of Sep days per year with no flow data

Oct_missing_Q
number of Oct days per year with no flow data

Nov_missing_Q
number of Nov days per year with no flow data

Dec_missing_Q
number of Dec days per year with no flow data

Transposing data creates a column of "Statistics" and subsequent columns for each year selected.

Examples

# Run if HYDAT database has been downloaded (using tidyhydat::download_hydat())
if (file.exists(tidyhydat::hy_downloaded_db())) {

# Calculate screening statistics using data frame and data argument with defaults
flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
screen_flow_data(data = flow_data)

# Calculate screening statistics using station_number argument with defaults
screen_flow_data(station_number = "08NM116")

# Calculate screening statistics for water years starting in October
screen_flow_data(station_number = "08NM116",
               water_year_start = 9)

# Calculate screening statistics for 7-day flows for July-September months only
screen_flow_data(station_number = "08NM116",
               roll_days = 7,
               months = 7:9)
}
write_flow_data  Write a streamflow dataset as a .xlsx, .xls, or .csv file

Description

Write a daily streamflow data set to a directory. Can fill missing dates or filter data by years or dates before writing using given arguments. List data frame or HYDAT station number to write its entirety. Can write as .xls, .xlsx, or .csv file types. Writing as Excel file type uses the writexl package.

Usage

```r
write_flow_data(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  water_year_start = 1,
  start_year,
  end_year,
  start_date,
  end_date,
  file_name,
  fill_missing = FALSE,
  digits
)
```

Arguments

- `data`: Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using `station_number` argument.
- `dates`: Name of column in `data` that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using `station_number` argument.
- `values`: Name of column in `data` that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using `station_number` argument.
- `groups`: Name of column in `data` that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using `station_number` argument.
station_number  Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

water_year_start  Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year  Numeric value of the first year of data to write. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year  Numeric value of the last year of data to write. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

start_date  Date (YYYY-MM-DD) of first date of data to write. Leave blank or set well before start date (i.e. 1800-01-01) if all dates required.

end_date  Date (YYYY-MM-DD) of last date of data to write. Leave blank or set well after end date (i.e. 2100-12-31) if all dates required.

file_name  Character string naming the output file. If none provided, a default file name (with .xlsx) is provided (see "Successfully created" message when using function for file name).

fill_missing  Logical value indicating whether to fill dates with missing flow data with NA. Default FALSE.

digits  Integer indicating the number of decimal places or significant digits used to round flow values. Use follows that of base::round() digits argument.

Examples

## Not run:

# Working examples:

# Write data from a data frame
flow_data <- tidyhydat::hy_daily_flows(station_number = "08NM116")
write_flow_data(data = flow_data,
       file_name = "Mission_Creek_daily_flows.xlsx")

# Write data directly from HYDAT
write_flow_data(station_number = "08NM116",
       file_name = "Mission_Creek_daily_flows.xlsx")

# Write data directly from HYDAT and fill missing dates with NA
write_flow_data(station_number = "08NM116",
       file_name = "Mission_Creek_daily_flows.xlsx",
       fill_missing = TRUE)

## End(Not run)
Write a suite of tables and plots from various fasstr functions into a directory

Description

Calculates and writes tables and plots from a suite of statistics from fasstr functions into an Excel workbook, and accompanying plot files for certain analyses. Due to the number of tables and plots to be made, this function may take several minutes to complete. If ignore_missing = FALSE (default) and there is missing data, some tables and plots may be empty and produce warnings. Use ignore_missing = TRUE to ignore the missing values or filter your data to complete years. Calculates statistics from all values, unless specified. Returns a list of tibbles and plots, along with saving the Excel and image files in a directory.

Usage

write_full_analysis(
  data,
  dates = Date,
  values = Value,
  groups = STATION_NUMBER,
  station_number,
  analyses = 1:7,
  basin_area,
  water_year_start = 1,
  start_year,
  end_year,
  exclude_years,
  months = 1:12,
  ignore_missing = FALSE,
  complete_years = FALSE,
  allowed_missing_annual = ifelse(ignore_missing, 100, 0),
  allowed_missing_monthly = ifelse(ignore_missing, 100, 0),
  zyp_method = "zhang",
  zyp_alpha,
  file_name,
  plot_filetype = "pdf"
)

Arguments

data Data frame of daily data that contains columns of dates, flow values, and (optional) groups (e.g. station numbers). Leave blank or set to NULL if using station_number argument.
dates Name of column in data that contains dates formatted YYYY-MM-DD. Only required if dates column name is not 'Date' (default). Leave blank or set to NULL if using station_number argument.
values
Name of column in data that contains numeric flow values, in units of cubic metres per second. Only required if values column name is not 'Value' (default). Leave blank if using station_number argument.

groups
Name of column in data that contains unique identifiers for different data sets, if applicable. Only required if groups column name is not 'STATION_NUMBER'. Function will automatically group by a column named 'STATION_NUMBER' if present. Remove the 'STATION_NUMBER' column beforehand to remove this grouping. Leave blank if using station_number argument.

station_number
Character string vector of seven digit Water Survey of Canada station numbers (e.g. "08NM116") of which to extract daily streamflow data from a HYDAT database. Requires tidyhydat package and a HYDAT database. Leave blank if using data argument.

analyses
Numeric vector of analyses to run (default is all (1:7)):
• 1: Screening
• 2: Long-term
• 3: Annual
• 4: Monthly
• 5: Daily
• 6: Annual Trends
• 7: Low-flow Frequencies

basin_area
Upstream drainage basin area, in square kilometres, to apply to observations. Three options:
(1) Leave blank if groups is STATION_NUMBER with HYDAT station numbers to extract basin areas from HYDAT.
(2) A single numeric value to apply to all observations.
(3) List each basin area for each group/station in groups (can override HYDAT value if listed) as such c("08NM116" = 795,"08NM242" = 10). If group is not listed the HYDAT area will be applied if it exists, otherwise it will be NA.

water_year_start
Numeric value indicating the month (1 through 12) of the start of water year for analysis. Default 1.

start_year
Numeric value of the first year to consider for analysis. Leave blank or set well before start date (i.e. 1800) to use from the first year of the source data.

end_year
Numeric value of the last year to consider for analysis. Leave blank or set well after end date (i.e. 2100) to use up to the last year of the source data.

exclude_years
Numeric vector of years to exclude from analysis. Leave blank or set to NULL to include all years.

months
Numeric vector of months to include in analysis. For example, 3 for March, 6:8 for Jun-Aug or c(10:12,1) for first four months (Oct-Jan) when water_year_start = 10 (Oct). Default summarizes all months (1:12). If not all months, seasonal total yield and volumetric flows will not be included.

ignore_missing
Logical value indicating whether dates with missing values should be included in the calculation. If TRUE then a statistic will be calculated regardless of missing dates. If FALSE then only those statistics from time periods with no missing dates will be returned. Default FALSE.
complete_years  Logical values indicating whether to include only years with complete data in analysis. Default FALSE.

allowed_missing_annual  
Numeric value between 0 and 100 indicating the **percentage** of missing dates allowed to be included to calculate an annual statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used. Only for annual means, percentiles, minimums, and maximums.

allowed_missing_monthly  
Numeric value between 0 and 100 indicating the **percentage** of missing dates allowed to be included to calculate a monthly statistic (0 to 100 percent). If 'ignore_missing = FALSE' then it defaults to 0 (zero missing dates allowed), if 'ignore_missing = TRUE' then it defaults to 100 (any missing dates allowed); consistent with ignore_missing usage. Supersedes ignore_missing when used. Only for monthly means, percentiles, minimums, and maximums.

zyp_method  
Character string identifying the prewhitened trend method to use from 'zyp', either 'zhang' or 'yuepilon'. 'zhang' is recommended over 'yuepilon' for hydrologic applications (see compute_annual_trends(); Bürger 2017; Zhang and Zwiers 2004). Only required if analysis group 6 is included. Default 'zhang'.

zyp_alpha  
Numeric value of the significance level (ex. 0.05) of when to plot a trend line. Leave blank for no line.

file_name  
Character string of the name of the Excel Workbook (and folder for plots if necessary) to create on drive to write all results.

plot_filetype  
Image type to write. One of 'png', 'eps', 'ps', 'tex', 'pdf', 'jpeg', 'tiff', 'bmp', or 'svg'. If not 'pdf' then individual plots will be created instead of a combined PDF. Default 'pdf'.

**See Also**

compute_full_analysis, screen_flow_data, plot_data_screening, plot_missing_dates, calc_longterm_monthly_stats, plot_longterm_monthly_stats, calc_longterm_daily_stats, plot_longterm_daily_stats, plot_monthly_means, plot_flow_duration, calc_annual_stats, plot_annual_stats, calc_annual_cumulative_stats, plot_annual_cumulative_stats, calc_annual_flow_timing, plot_annual_flow_timing, calc_annual_normal_days, plot_annual_normal_days, calc_annual_lowflows, plot_annual_lowflows, plot_annual_means, calc_monthly_stats, plot_monthly_stats, calc_monthly_cumulative_stats, plot_monthly_cumulative_stats, calc_daily_stats, plot_daily_stats, calc_daily_cumulative_stats, plot_daily_cumulative_stats, compute_annual_trends, compute_annual_frequencies, write_flow_data, write_plots

**Examples**

```r
## Not run:
#
# Working examples:
#
# Save a full analysis will all the analyses
write_full_analysis(station_number = "08BM116",
```
write_objects_list

Write all data frames and plots from a list of objects into a directory

Description
Write a list of tables (data frames) and plots (ggplots; as used by fasstr) into a directory. Objects that are not class "data.frame" or "gg" will not be saved. Each table and plot will be named by the object name in the list.

Usage
write_objects_list(
  list, 
  folder_name, 
  table_filetype, 
  plot_filetype, 
  width, 
  height, 
  units = "in", 
  dpi = 300 
)

Arguments

list List of data frames and plots to write to disk.
folder_name Name of folder to create on disk (if it does not exist) to write each plot from list. If using combined_pdf argument, then it will be the name of the PDF document.
table_filetype Table file type to write. One of 'csv', 'xls', or 'xlsx'.
plot_filetype Image type to write. One of 'png', 'eps', 'ps', 'tex', 'pdf', 'jpeg', 'tiff', 'bmp', or 'svg'. Image type will be overwritten if using combined_pdf is used.
width Numeric plot width in units. If not supplied, uses the size of current graphics device.
Write plots from a list into a directory or PDF document

Description

Write a list of plots (ggplots; as used by fasstr) into a directory or PDF document. When writing into a named directory each plot will be named by the plot name listed in the list; uses ggplot2::ggsave function. When writing into a PDF document (combined_pdf == TRUE) the plot names will not appear; uses grDevices::pdf function.

Usage

```r
write_plots(
  plots,
  folder_name,
  plot_filetype,
  width,
  height,
  units = "in",
  dpi = 300,
  combined_pdf = FALSE
)
```
Arguments

plots List of plots to write to disk.
folder_name Name of folder to create on disk (if it does not exist) to write each plot from list. If using combined_pdf argument, then it will be the name of the PDF document.
plot_filetype Image type to write. One of 'png', 'eps', 'ps', 'pdf', 'jpeg', 'tiff', 'bmp', or 'svg'. Image type will be overwritten if using combined_pdf is used.
width Numeric plot width in units. If not supplied, uses the size of current graphics device.
height Numeric plot height in units. If not supplied, uses the size of current graphics device.
units Character string plot height and width units, one of 'in', 'cm', or 'mm'. Default 'in'.
dpi Numeric resolution of plots. Default 300.
combined_pdf Logical value indicating whether to combine list of plots into one PDF document. Default FALSE.

Examples

## Not run:

# Working examples:

# Example plots to save
plots <- plot_annual_lowflows(station_number = "08NM116")

# Write the plots as "png" files
write_plots(plots = plots,
            folder_name = "Low Flow Plots",
            plot_filetype = "png")

# Write the plots as a combined "pdf" document
write_plots(plots = plots,
            folder_name = "Low Flow Plots",
            combined_pdf = TRUE)

## End(Not run)

write_results Write a data frame as a .xlsx, .xls, or .csv file

Description

Write a data frame to a directory with all numbers rounded to specified digits. Can write as .xls, .xlsx, or .csv file types. Writing as .xlsx or .xls uses the writexl package.
Usage

write_results(data, file_name, digits)

Arguments

data Data frame to be written to a directory.
file_name Character string naming the output file. Required.
digits Integer indicating the number of decimal places or significant digits used to round flow values. Use follows that of base::round() digits argument.

Examples

## Not run:

# Working examples:

data_results <- calc_longterm_daily_stats(station_number = c("08HA002", "08HA011"),
                                           start_year = 1971, end_year = 2000)

# Write the data and round numbers to 1 decimal place
write_results(data = data_results,
              digits = 1)

## End(Not run)
add_basin_area, 3
add_cumulative_volume, 4
add_cumulative_yield, 6
add_daily_volume, 8
add_daily_yield, 9
add_date_variables, 10
add_rolling_means, 11
add_seasons, 13

calc_all_annual_stats, 14, 72
calc_annual_cumulative_stats, 17, 18, 80, 86, 163
calc_annual_extremes, 21, 89, 92
calc_annual_flow_timing, 17, 24, 80, 94, 97, 163
calc_annual_highflows, 26, 99
calc_annual_lowflows, 17, 29, 80, 101, 163
calc_annual_normal_days, 17, 32, 80, 106, 108, 163
calc_annual_outside_normal, 34, 111
calc_annual_peaks, 36
calc_annual_stats, 17, 39, 80, 104, 114, 117, 163
calc_daily_cumulative_stats, 42, 80, 122, 163
calc_daily_stats, 45, 80, 125, 163
calc_flow_percentile, 47
calc_longterm_daily_stats, 49, 80, 135, 138, 149, 163
calc_longterm_mean, 52, 149
calc_longterm_monthly_stats, 54, 80, 142, 163
calc_longterm_percentile, 57
calc_monthly_cumulative_stats, 59, 80, 147, 163
calc_monthly_stats, 17, 62, 80, 152, 155, 163
compute_annual_frequencies, 65, 80, 163
compute_annual_stats, 77, 163
compute_annual_trends, 68, 80, 163
compute_frequency_analysis, 67, 73, 77, 82
compute_frequency_quantile, 75
compute_full_analysis, 77, 163
compute_hydat_peak_frequencies, 80
fill_missing_dates, 83
plot_annual_cumulative_stats, 80, 84, 163
plot_annual_extremes, 87, 92
plot_annual_extremes_year, 89
plot_annual_flow_timing, 80, 92, 97, 163
plot_annual_flow_timing_year, 95
plot_annual_highflows, 97
plot_annual_lowflows, 80, 99, 163
plot_annual_means, 80, 102, 163
plot_annual_normal_days, 80, 104, 108, 163
plot_annual_normal_days_year, 106
plot_annual_outside_normal, 109
plot_annual_stats, 80, 111, 163
plot_annual_stats2, 114
plot_annual_symbols, 118
plot_daily_cumulative_stats, 80, 120, 163
plot_daily_stats, 80, 123, 163
plot_data_screening, 80, 126, 163
plot_flow_data, 80, 128
plot_flow_data_symbols, 131
plot_flow_duration, 80, 133, 163
plot_longterm_daily_stats, 80, 136, 163
plot_longterm_monthly_stats, 80, 139, 163
plot_missing_dates, 80, 142, 163
plot_monthly_cumulative_stats, 80, 144, 163
plot_monthly_means, 80, 147, 163
plot_monthly_stats, 80, 150, 163
plot_monthly_stats2, 153
screen_flow_data, 80, 128, 144, 156, 163

write_flow_data, 80, 159, 163
write_full_analysis, 161
write_objects_list, 164
write_plots, 80, 163, 165
write_results, 166