

# Package ‘datefixR’

May 4, 2022

**Title** Fix Really Messy Dates

**Version** 0.1.6

**Description** Fixes messy dates in data frames such as those entered via text boxes. Standardizes / - and whitespace separation, month abbreviations, and year first or day first by converting to R's built-in Date class. Imputes missing date or month using user-provided values.

**License** GPL (>= 3)

**Depends** R (>= 4.0.0)

**Imports** stringr

**Language** en-US

**Encoding** UTF-8

**RoxygenNote** 7.1.2

**URL** <https://www.constantine-cooke.com/datefixR/>  
<https://github.com/nathansam/datefixR>

**BugReports** <https://github.com/nathansam/datefixR/issues>

**Suggests** rmarkdown, knitr, testthat (>= 3.0.0)

**Config/testthat/edition** 3

**VignetteBuilder** knitr

**NeedsCompilation** no

**Author** Nathan Constantine-Cooke [aut, cre]  
(<https://orcid.org/0000-0002-4437-8713>)

**Maintainer** Nathan Constantine-Cooke <nathan.constantine-cooke@ed.ac.uk>

**Repository** CRAN

**Date/Publication** 2022-05-03 23:00:05 UTC

## R topics documented:

fix_date . . . . .	2
fix_dates . . . . .	3

<b>Index</b>	<b>5</b>
--------------	----------

---

`fix_date`*Convert improperly formatted date to R's Date class*

---

**Description**

Converts a single improperly formatted date to R's Date class Supports numerous separators including /,- or white space. Supports all-numeric, abbreviation or long-hand month notation. Where day of the month has not been supplied, the first day of the month is imputed. When day, month and year is given either DMY or YMD is assumed; the US system of MDY is not supported.

**Usage**

```
fix_date(date, day.impute = 1, month.impute = 7, format = "dmy")
```

**Arguments**

<code>date</code>	Character to be converted to R's date class.
<code>day.impute</code>	Integer. Day of the month to be imputed if not available. defaults to 1. If <code>day.impute = NA</code> then NA will be imputed for the date instead and a warning will be raised. If <code>day.impute = NULL</code> then instead of imputing the day of the month, the function will fail
<code>month.impute</code>	Integer. Month to be be imputed if not available. Defaults to 7 (July). If <code>month.impute = NA</code> then NA will be imputed for the date instead and a warning will be raised. If <code>month.impute = NULL</code> then instead of imputing the month, the function will fail.
<code>format</code>	Character. The format which a date is mostly likely to be given in. Either "dmy" (default) or "mdy". If year appears to have been given first, then YMD is assumed for the subject (format argument is not used for these observations)

**Value**

An object belonging to R's built in Date class.

**See Also**

[fix\\_dates](#) Similar to `fix_date()` except is applicable to columns of a dataframe.

**Examples**

```
bad.date <- "02 03 2021"  
fixed.date <- fix_date(bad.date)
```

---

fix_dates	<i>Clean up messy date columns</i>
-----------	------------------------------------

---

## Description

Cleans up a dataframe object which has date columns entered via a free-text box (possibly by different users) and are therefore in a non-standardized format. Supports numerous separators including /, -, or space. Supports all-numeric, abbreviation, or long-hand month notation. Where day of the month has not been supplied, the first day of the month is imputed. Either DMY or YMD is assumed by default. However, the US system of MDY is supported via the format argument.

## Usage

```
fix_dates(  
  df,  
  col.names,  
  day.impute = 1,  
  month.impute = 7,  
  id = NULL,  
  format = "dmy"  
)
```

## Arguments

df	A dataframe or tibble object with messy date column(s)
col.names	Character vector of names of columns of messy date data
day.impute	Integer. Day of the month to be imputed if not available. defaults to 1. If day.impute = NA then NA will be imputed for the date instead and a warning will be raised. If day.impute = NULL then instead of imputing the day of the month, the function will fail
month.impute	Integer. Month to be be imputed if not available. Defaults to 7 (July). If month.impute = NA then NA will be imputed for the date instead and a warning will be raised. If month.impute = NULL then instead of imputing the month, the function will fail.
id	Name of column containing row IDs. By default, the first column is assumed.
format	Character. The format which a date is mostly likely to be given in. Either "dmy" (default) or "mdy". If year appears to have been given first, then YMD is assumed for the subject (format argument is not used for these observations)

## Value

A dataframe or tibble object. Dependent on the type of df. Selected columns are of type Date

## See Also

[fix\\_date](#) Similar to fix\_dates() except can only be applied to character objects.

**Examples**

```
bad.dates <- data.frame(  
  id = seq(5),  
  some.dates = c(  
    "02/05/92",  
    "01-04-2020",  
    "1996/05/01",  
    "2020-05-01",  
    "02-04-96"  
  ),  
  some.more.dates = c(  
    "2015",  
    "02/05/00",  
    "05/1990",  
    "2012-08",  
    "jan 2020"  
  )  
)  
fixed.df <- fix_dates(bad.dates, c("some.dates", "some.more.dates"))
```

# Index

fix\_date, [2](#), [3](#)  
fix\_dates, [2](#), [3](#)