Package ‘harmonizer’

May 23, 2022

Type Package
Date 2022-05-22
Title Harmonizing CN8 and PC8 Product Codes
Version 0.3.0
Author Christoph Baumgartner [cre, aut] (<https://orcid.org/0000-0002-2681-071X>), Stjepan Srhoj [aut] (<https://orcid.org/0000-0002-7105-8088>), Janette Walde [aut] (<https://orcid.org/0000-0002-1241-5292>)
Maintainer Christoph Baumgartner <Christoph.Baumgartner@uibk.ac.at>
Description Several functions are provided to harmonize CN8 (Combined Nomenclature 8 digits) and PC8 (Production Communautaire 8 digits) product codes over time and the classification systems HS6 and BEC. Harmonization of CN8 codes are possible by default from 1995 to 2022 and of PC8 from 2001 to 2021, respectively.
License GPL (>= 3)
Imports stats, utils, methods
Encoding UTF-8
LazyLoad yes
Depends R (>= 3.5.0)
RoxygenNote 7.2.0
Suggests rmarkdown, knitr
VignetteBuilder knitr
NeedsCompilation no
Repository CRAN
Date/Publication 2022-05-23 15:20:02 UTC

R topics documented:

  cn8_to_bec .......................................................... 2
  get_data_directory .............................................. 3
**Description**

Provides a dataframe which contains all CN8 product codes and related BEC codes in a given time period.

**Usage**

\[
\text{cn8\_to\_bec}(b, e, \text{historymatrix} = \text{NULL}, \text{progress} = \text{TRUE})
\]

**Arguments**

- **b**: first year of interest
- **e**: last year of interest
- **historymatrix**: History matrix of CN8 product codes. Provided by `history\_matrix\_cn8()`.
- **progress**: logical, determines whether progress is printed in console or not.

**Value**

A data frame that contains all CN8 product codes and related BEC and HS6 codes in a given time period. The following table offers an overview of all provided variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN8</td>
<td>character; a specific CN8 code</td>
</tr>
<tr>
<td>HS6</td>
<td>character; provides the HS6 classification of the CN8plus code</td>
</tr>
<tr>
<td>BEC</td>
<td>character; provides the BEC classification on a high aggregation level (1 digit)</td>
</tr>
<tr>
<td>BEC_agr</td>
<td>character; provides the BEC classification on a lower aggregation level (up to 3 digits)</td>
</tr>
</tbody>
</table>

**Examples**

\[
\text{cn8\_to\_bec}(b = 2008, e = 2010)
\]
**get_data_directory**

*Data path for custom data*

**Description**

Provides the directory where custom data must be stored and the used data (e.g., concordance lists, list of codes) can be edited.

**Usage**

```
get_data_directory(path = TRUE, open_explorer = FALSE, 
                    show_data = NULL)
```

**Arguments**

- `path` logical, determines whether the path is printed in the console
- `open_explorer` logical, determines whether an explorer is opened in addition. Only executable if the directory path does not contain any blanks.
- `show_data` character string, which must take one of the following values: "CN8", "HS6", "PC8" or "HS6toBEC". All available data in the given directory is printed in the console. Only executable if the directory path does not contain any blanks.

**Value**

Returns the path (character), of the directory where custom data must be stored and the used data (e.g., concordance lists, list of codes) can be edited.

**Examples**

```
get_data_directory()
get_data_directory(path = FALSE, show_data = "CN8")
```

---

**harmonize_cn8**

*Harmonization of CN8 product codes*

**Description**

Provides a dataframe which contains all CN8 product codes and their history in the demanded time period, as well as harmonized CN8plus code, harmonized HS6plus code and BEC classification.

**Usage**

```
harmonize_cn8(b, e, historymatrix, harmonize.to = "e", 
               progress = TRUE)
```
Arguments

b  first year of interest
e  last year of interest

historymatrix  History matrix of CN8 product codes. Provided by `history_matrix_cn8()`. By default `NULL`; the function computes the needed harmonized data.

harmonize.to  Defines which year for harmonization is used. It may take the following values:
  • "e", harmonizes product codes towards year e
  • "b", harmonizes product codes towards year b

HS6breaks  Vector of years, where HS6 codes were changed. Do not edit, unless additional break is needed.

progress  logical, determines whether progress is printed in console or not.

Value

A data frame that contains all CN8 product codes and their history, harmonized CN8plus codes, harmonized HS6plus codes, and BEC classification. The 'plus-codes' are the main outcome of the function. They provide harmonized information of the product codes, i.e. comparable codes. Every harmonization refers to the last year of interest. The following table offers an overview of all provided variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN8_xxxx</td>
<td>character; a specific CN8 code in a given year</td>
</tr>
<tr>
<td>CN8plus</td>
<td>character; the harmonization code for CN8, which refers to the last/first year of the time period</td>
</tr>
<tr>
<td>HS6plus</td>
<td>character; the harmonization code of HS6, which refers to the last/first year of the time period</td>
</tr>
<tr>
<td>BEC</td>
<td>character; provides the BEC classification at a high aggregation level (1 digit)</td>
</tr>
<tr>
<td>BEC_agr</td>
<td>character; provides the BEC classification at a lower aggregation level (up to 3 digits)</td>
</tr>
<tr>
<td>SNA_basic_class</td>
<td>character; provides information if the code is classified as consumption, capital or intermediate good in SNA</td>
</tr>
<tr>
<td>flag</td>
<td>numeric; either 0 or 1; 1 indicates that this code remained the same in notation over the whole time period</td>
</tr>
<tr>
<td>flagyear</td>
<td>numeric; indicates the first year in which the flag was set to 1</td>
</tr>
</tbody>
</table>

Examples

```
harmonize_cn8(b = 2008, e = 2010)

mydata <- history_matrix_cn8(b = 2016, e = 2018)
myharmonization <- harmonize_cn8(b = 2016, e = 2018,
                                 historymatrix = mydata)
```
**Description**

Provides a dataframe which contains all PC8 product codes and their history in the demanded time period, as well as harmonized PC8plus code, harmonized HS6plus code and BEC classification.

**Usage**

```r
harmonize_pc8(b, e, historymatrix = NULL, harmonize.to = "e",
              progress = TRUE)
```

**Arguments**

- `b` first year of interest
- `e` last year of interest
- `historymatrix` History matrix of PC8 product codes. Provided by `history_matrix_pc8()`. By default `NULL`; the function computes the needed harmonized data.
- `harmonize.to` Defines which year for harmonization is used. It may take the following values:
  - "e", harmonizes product codes towards year e
  - "b", harmonizes product codes towards year b
- `HS6breaks` Vector of years, where HS6 codes were changed.
- `progress` logical, determines whether progress is printed in console or not.

**Value**

A data frame that contains all PC8 product codes and their history, harmonized PC8plus codes, harmonized HS6plus codes, and BEC classification. The 'plus-codes' are the main outcome of the function. They provide harmonized information of the product codes, i.e. comparable codes. Every harmonization refers to the last year of interest. The following table offers an overview of all provided variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC8_xxxx</td>
<td>character; a specific PC8 code in a given year</td>
</tr>
<tr>
<td>PC8plus</td>
<td>character; the harmonization code for PC8, which refers to the last/first year of the time period</td>
</tr>
<tr>
<td>HS6plus</td>
<td>character; the harmonization code of HS6, which refers to the last/first year of the time period</td>
</tr>
<tr>
<td>BEC</td>
<td>character; provides the BEC classification at a high aggregation level (1 digit)</td>
</tr>
<tr>
<td>BEC_agr</td>
<td>character; provides the BEC classification at a lower aggregation level (up to 3 digits)</td>
</tr>
<tr>
<td>SNA_basic_class</td>
<td>character; provides information if the code is classified as consumption, capital or intermediate good in BEC</td>
</tr>
<tr>
<td>flag</td>
<td>numeric; either 0 or 1; 1 indicates that this code remained the same in notation over the whole time period</td>
</tr>
<tr>
<td>flagyear</td>
<td>numeric; indicates the first year in which the flag was set to 1</td>
</tr>
</tbody>
</table>

**Examples**

```r
harmonize_pc8(b = 2009, e = 2011)
mydata <- history_matrix_pc8(b = 2015, e = 2017)
myharmonization <- harmonize_pc8(b = 2015, e = 2017,
```
**Description**

Provides a dataframe which contains all CN8 product codes and their history in a given time period.

**Usage**

```r
history_matrix_cn8(b, e, c1 = 1988, c2 = 2022, progress = TRUE)
```

**Arguments**

- `b`: first year of interest
- `e`: last year of interest
- `c1`: first year of the concordance list
- `c2`: last year of the concordance list
- `progress`: logical, determines whether progress is printed in console or not.

**Value**

A data frame that contains all CN8 product codes and their history over time for the demanded time period. This dataset is the basis for the main function `harmonize_cn8()` and can be obtained therewith as well. The following table offers an overview of all provided variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN8_xxxx</td>
<td>character; a specific CN8 code in a given year</td>
</tr>
<tr>
<td>flag</td>
<td>numeric; either 0 or 1; 1 indicates that this code remained the same in notation over the whole time period but was split or merged in addition</td>
</tr>
<tr>
<td>flagyear</td>
<td>numeric; indicates the first year in which the flag was set to 1</td>
</tr>
</tbody>
</table>

**Examples**

```r
history_matrix_cn8(b = 2008, e = 2010)
```

---

**Description**

Provides a dataframe which contains all PC8 product codes and their history in a given time period.
**Usage**

```r
history_matrix_pc8(b, e, progress = TRUE)
```

**Arguments**

- `b`: first year of interest
- `e`: last year of interest
- `progress`: logical, determines whether progress is printed in console or not.

**Value**

A data frame that contains all PC8 product codes and their history over time for the demanded time period. This dataset is the basis for the main function `harmonize_pc8()` and can be obtained therewith as well. The following table offers an overview of all provided variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC8_xxxx</td>
<td>character; a specific PC8 code in a given year</td>
</tr>
<tr>
<td>flag</td>
<td>numeric; either 0 or 1; 1 indicates that this code remained the same in notation over the whole time period but was split or merged in addition</td>
</tr>
<tr>
<td>flagyear</td>
<td>numeric; indicates the first year in which the flag was set to 1</td>
</tr>
</tbody>
</table>

**Examples**

```r
history_matrix_pc8(b = 2008, e = 2010)
```

---

**pc8_to_bec**

*concordance list between PC8 and BEC*

**Description**

Provides a dataframe which contains all PC8 product codes and related BEC codes in the demanded time period.

**Usage**

```r
pc8_to_bec(b, e, historymatrix = NULL, progress = TRUE)
```

**Arguments**

- `b`: first year of interest
- `e`: last year of interest
- `historymatrix`: History matrix of PC8 product codes. Provided by `history_matrix_pc8()`.
- `progress`: logical, determines whether progress is printed in console or not.
Value
A data frame that contains all PC8 product codes and related BEC and HS6 codes in a given time period. The following table offers an overview of all provided variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC8</td>
<td>character; a specific PC8 code</td>
</tr>
<tr>
<td>HS6</td>
<td>character; provides the HS6 classification of the PC8plus code</td>
</tr>
<tr>
<td>BEC</td>
<td>character; provides the BEC classification on a high aggregation level (1 digit)</td>
</tr>
<tr>
<td>BEC_agr</td>
<td>character; provides the BEC classification on a lower aggregation level (up to 3 digits)</td>
</tr>
</tbody>
</table>

Examples

```r
cp8_to_bec(b = 2008, e = 2010)
```

---

**utilize_cn8**

A possible utilization of harmonized CN8 products codes

Description
Provide an application of the data frames obtained by the main function, `harmonize_cn8`. To use these additional functions, data on firm-level is required, which is data that is not provided by the package.

Usage

```r
utilize_cn8(b, e, firm_data, harmonized_data = NULL, progress = TRUE, output = "merged.firm.data", value = FALSE, base = "CN8")
```

Arguments

- `b` first year of interest
- `e` last year of interest
- `firm_data` Data on firm level which must provide the following columns: "firmID", "year" and "CN8".
- `harmonized_data` Harmonized data of CN8 product codes. Provided by `harmonize_cn8()`. By default NULL; the function computes the needed harmonized data.
- `progress` logical, determines whether progress is printed in console or not.
- `output` Defines which dataframe is returned. It may take the following values:
  - "product.changes", returns all changed CN8 product codes per firm per year (see description of (a) below)
  - "merged.firm.data", returns entered firm data, extended by harmonized data (see description of (b) below)
- "all", returns both dataframes as a list

**value** logical, determines whether value is calculated for same/new/dropped products. Only possible if data contains a column: "value". Value may contain different quantities (e.g. sales [Euro] or weight [kg]).

**base** Defines which plus-codes are used as a base for calculating added/dropped/same products and their corresponding values. It may take the following values:

- "CN8", uses CN8plus codes for computation.
- "HS6", uses HS6plus codes for computation.

## Value

Provides two possible data frames:

(a) One dataframe that contains all changed CN8 product codes per firm per year. In more detail, this means how many products remained the same, were added or dropped - the value of the same/added/dropped products - how many products were produced by a certain firm in a given year, and how many products were produced in the year after. As a base of this computation CN8plus codes or HS6plus codes can be used.

(b) One dataframe that is based on the entered firm data. The entered firm data is extended by harmonized data (that is "CN8plus", "flag", "flagyear", "HS6plus", "BEC", "BEC_agr", "SNA_basic_class").

Table that summarizes the output, described by the notation (a) above:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>firmID</td>
<td>character; specific code that describes a firm over the years (this code does not change over the time period)</td>
</tr>
<tr>
<td>period_UL</td>
<td>character; lower limit of the time period</td>
</tr>
<tr>
<td>period</td>
<td>character; time period in which the product was produced</td>
</tr>
<tr>
<td>gap</td>
<td>numeric; indicating if the time period is greater than one (i.e. upper limit - lower limit &gt; 1)</td>
</tr>
<tr>
<td>same_products</td>
<td>numeric; number of products that were produced in both years (i.e. remained in the product portfolio)</td>
</tr>
<tr>
<td>value_same_products</td>
<td>numeric; value of products that were produced in both years (i.e. remained in the product portfolio)</td>
</tr>
<tr>
<td>new_products</td>
<td>numeric; number of added products in the upper limit of the time period (i.e. added to the product portfolio)</td>
</tr>
<tr>
<td>value_new_products</td>
<td>numeric; value of added products in the upper limit of the time period (i.e. added to the product portfolio)</td>
</tr>
<tr>
<td>dropped_products</td>
<td>numeric; number of dropped products in the upper limit of the time period (i.e. removed from the product portfolio)</td>
</tr>
<tr>
<td>value_dropped_products</td>
<td>numeric; value of dropped products in the upper limit of the time period (i.e. removed from the product portfolio)</td>
</tr>
<tr>
<td>nbr_of_products_period_LL</td>
<td>numeric; number of all products produced in the lower limit of the time period (i.e. entire product portfolio)</td>
</tr>
<tr>
<td>nbr_of_products_period_UL</td>
<td>numeric; number of all products produced in the upper limit of the time period (i.e. entire product portfolio)</td>
</tr>
</tbody>
</table>

Table that summarizes the output, described by the notation (b) above:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>firmID</td>
<td>character; specific code that describes a firm over the years (this code does not change over the time period)</td>
</tr>
<tr>
<td>year</td>
<td>numeric; year in which the firm produced a product (provided by user)</td>
</tr>
<tr>
<td>CN8</td>
<td>character; CN8 code of firm product (provided by user)</td>
</tr>
<tr>
<td>(value)</td>
<td>numeric; value of the corresponding product code (may be provided by user)</td>
</tr>
<tr>
<td>...</td>
<td>character; additional columns from original firm data (provided by user)</td>
</tr>
<tr>
<td>CN8plus</td>
<td>character; final harmonization, which refers to the last year of the time period</td>
</tr>
<tr>
<td>flag</td>
<td>character; either 0 or 1; 1 indicates that this code remained the same in notation over the whole time period</td>
</tr>
</tbody>
</table>
**utilize_pc8**

- **HS6**: character; provides the HS6 classification of the CN8plus code
- **HS6plus**: character; also adjusts for the change lists of HS6
- **BEC**: character; provides the BEC classification on a high aggregated level (1 digit)
- **BEC_agr**: character; provides the BEC classification on a less aggregated level (up to 3 digits)
- **SNA_basic_class**: character; provides information if the code is classified as consumption, capital or intermediate good in BEC

**Examples**

```r
sampledata <- read.table(paste0(system.file("extdata", package = "harmonizer"), 
"/sampledata/cn8sample.txt"), sep = ";", 
header = TRUE, colClasses = "character")

utilize_cn8(b = 2008, e = 2010, firm_data = sampledata)

newdata <- utilize_cn8(b = 2008, e = 2010, firm_data = sampledata, 
output = "all")

changes <- newdata[[1]]

merged_data <- newdata[[2]]
```

**Description**

Provide an application of the data frames obtained by the main function, `harmonize_pc8`. To use these additional functions, data on firm-level is required, which is data that is not provided by the package.

**Usage**

```r
utilize_pc8(b, e, firm_data, harmonized_data = NULL, 
progress = TRUE, output = "merged.firm.data", 
value = FALSE, base = "PC8")
```

**Arguments**

- **b**: first year of interest
- **e**: last year of interest
- **firm_data**: Data on firm level which must provide the following columns: "firmID", "year" and "PC8".
- **harmonized_data**: Harmonized data of PC8 product codes. Provided by `harmonize_pc8()`. By default `NULL`; the function computes the needed harmonized data.
- **progress**: logical, determines whether progress is printed in console or not.
- **output**: Defines which dataframe is returned. It may take the following values:
• "product.changes", returns all changed PC8 product codes per firm per year (see description of (a) below)
• "merged.firm.data", returns entered firm data, extended by harmonized data (see description of (b) below)
• "all", returns both dataframes as a list

value
logical, determines whether value is calculated for same/new/dropped products. Only possible if data contains a column: "value". Value may contain different quantities (e.g. sales [Euro] or weight [kg]).

base
Defines which plus-codes are used as a base for calculating added/dropped/same products and their corresponding values. It may take the following values:
• "PC8", uses CN8plus codes for computation.
• "HS6", uses HS6plus codes for computation.

Value
Provides two possible data frames:

(a) One dataframe that contains all changed PC8 product codes per firm per year. In more detail, this means how many products remained the same, were added or dropped - the value of the same/added/dropped products - how many products were produced by a certain firm in a given year, and how many products were produced in the year after. As a base of this computation PC8plus codes or HS6plus codes can be used.

(b) One dataframe that is based on the entered firm data. The entered firm data is extended by harmonized data (that is "PC8plus", "flag", "flagyear", "HS6plus", "BEC", "BEC_agr", "SNA_basic_class").

Table that summarizes the output, described by the notation (a) above:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>firmID</td>
<td>character; specific code that describes a firm over the years (this code does not change over time)</td>
</tr>
<tr>
<td>period_UL</td>
<td>character; lower limit of the time period</td>
</tr>
<tr>
<td>period</td>
<td>character; time period in which the product was produced</td>
</tr>
<tr>
<td>gap</td>
<td>numeric; indicating if the time period is greater than one (i.e. upper limit - lower limit &gt; 1)</td>
</tr>
<tr>
<td>same_products</td>
<td>numeric; number of products that were produced in both years (i.e. remained in the product portfolio)</td>
</tr>
<tr>
<td>value_same_products</td>
<td>numeric; value of products that were produced in both years (i.e. remained in the product portfolio)</td>
</tr>
<tr>
<td>new_products</td>
<td>numeric; number of added products in the upper limit of the time period (i.e. added to the product portfolio)</td>
</tr>
<tr>
<td>value_new_products</td>
<td>numeric; value of added products in the upper limit of the time period (i.e. added to the product portfolio)</td>
</tr>
<tr>
<td>dropped_products</td>
<td>numeric; number of dropped products in the upper limit of the time period (i.e. removed from the product portfolio)</td>
</tr>
<tr>
<td>value_dropped_products</td>
<td>numeric; value of dropped products in the upper limit of the time period (i.e. removed from the product portfolio)</td>
</tr>
<tr>
<td>nbr_of_products_period_LL</td>
<td>numeric; number of all products produced in the lower limit of the time period (i.e. entire product portfolio)</td>
</tr>
<tr>
<td>nbr_of_products_period_UL</td>
<td>numeric; number of all products produced in the upper limit of the time period (i.e. entire product portfolio)</td>
</tr>
</tbody>
</table>

Table that summarizes the output, described by the notation (b) above:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>firmID</td>
<td>character; specific code that describes a firm over the years (this code does not change over time, provided by user)</td>
</tr>
<tr>
<td>year</td>
<td>numeric; year in which the firm produced a product (provided by user)</td>
</tr>
<tr>
<td>PC8</td>
<td>character; PC8 code of firm product (provided by user)</td>
</tr>
</tbody>
</table>
(value) numeric; value of the corresponding product code (may be provided by user)
...
PC8plus character; additional columns from original firm data (provided by user)
flag character; final harmonization, which refers to the last year of the time period
HS6 character; either 0 or 1; 1 indicates that this code remained the same in notation over the whole time period
HS6plus character; provides the HS6 classification of the PC8plus code
BEC character; also adjusts for the change lists of HS6
BEC_agr character; provides the BEC classification on a high aggregated level (1 digit)
SNA_basic_class character; provides the BEC classification on a less aggregated level (up to 3 digits)

Examples

```r
sampledata <- read.table(paste0(system.file("extdata", package = "harmonizer"),
"/sampledata/pc8sample.txt"), sep = ";",
header = TRUE, colClasses = "character")

utilize_pc8(b = 2011, e = 2013, firm_data = sampledata)

newdata <- utilize_pc8(b = 2011, e = 2013, firm_data = sampledata,
output = "all")
changes <- newdata[[1]]
merged_data <- newdata[[2]]
```
Index

cn8_to_bec, 2
get_data_directory, 3
harmonize_cn8, 3
harmonize_pc8, 4
history_matrix_cn8, 6
history_matrix_pc8, 6
pc8_to_bec, 7
utilize_cn8, 8
utilize_pc8, 10