6_R_Daten_aufbereiten_Uebung.R

Felix

2020-04-20

```r
### title: Übung 6 ####
### topic: Daten aufbereiten ###
### author: Felix Wohlgemuth ###
### date: 20.04.2020 ###


### Preliminaries
library(tidyverse) # Easily Install and Load the 'Tidyverse'
```

```
## -- Attaching packages ---------------------------------------------------------------

## v ggplot2 3.3.0      v purrr   0.3.3
## v tibble  3.0.0      v dplyr   0.8.5
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0

## -- Conflicts ------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
### task 1
# Laden Sie von moodle den OECD Datensatz "SOCX_AGG_31032020142101957.csv" herunter.
# Bearbeiten Sie den Datensatz, so dass Sie ein tidy dataset bekommen.

# load dataset
socx_data <- read_csv("_raw/SOCX_AGG_31032020142101957.csv")
```

```
## Parsed with column specification:
## cols(
##    .default = col_character(),
##    SOURCE = col_double(),
##    BRANCH = col_double(),
##    TYPEXP = col_double(),
##    TYPROG = col_double(),
##    YEAR = col_double(),
##    Year = col_double(),
##    `PowerCode Code` = col_double(),
##    `Reference Period Code` = col_logical(),
##    `Reference Period` = col_logical(),
##    Value = col_double(),
##    `Flag Codes` = col_logical(),
##    Flags = col_logical()
## )

## See spec(...) for full column specifications.
```

```r
# clean data
socx_data <- socx_data %>%
  select(COUNTRY, YEAR, Value) %>%     # select essential variables
  rename(fampol_exp_total_pct = Value) # rename Value Variable based in information in other variables



### task 2
# Berechnen Sie die durchschnittlichen Ausgaben für Familienpolitik in Österreich im Zeitraum 2000 - 20

# average public expenditure on family policy for Austria (2000 - 2015)
socx_data %>%                                            # select dataframe
  filter(COUNTRY == "AUS") %>%                           # select only Austrian data
  summarise(mean_fampol_exp_total_pct = mean(fampol_exp_total_pct)) # specify mean()

## # A tibble: 1 x 1
##   mean_fampol_exp_total_pct
##                       <dbl>
## 1                      2.78
# average public expenditure on family policy for all countries in sample (2000 - 2015)
mean_fampol_exp_by_country <-  socx_data %>%                     # select dataframe and save res
  group_by(COUNTRY) %>%                                         # group by countries, following
  summarise(mean_fampol_exp_total_pct = mean(fampol_exp_total_pct)) %>% # specify mean()
  print()                                                       # print results first 10 values

## # A tibble: 37 x 2
##    COUNTRY mean_fampol_exp_total_pct
##    <chr>                       <dbl>
##  1 AUS                          2.78
##  2 AUT                          2.80
##  3 BEL                          2.69
##  4 CAN                          1.17
##  5 CHE                          1.49
##  6 CHL                          1.21
##  7 CZE                          2.05
##  8 DEU                          2.05
##  9 DNK                          3.59
## 10 ESP                          1.19
## # ... with 27 more rows
### task 3
# Welche fünf Länder haben im Jahr 2010 am meisten für Familienpolitik ausgegeben (in % des BIPs)?
# Speichern Sie die Ergebnisse als # Kommentar im Skript.

socx_data %>%                            # select dataframe
  filter(YEAR == 2010) %>%               # select onle 2010 values
  arrange(-fampol_exp_total_pct) %>%     # arrange data ascending by family policy expenditure in 2010
  head( n = 5)                           # print first 5 values

## # A tibble: 5 x 3
##   COUNTRY  YEAR fampol_exp_total_pct
##   <chr>   <dbl>                <dbl>
## 1 GBR      2010                 4.00
## 2 LUX      2010                 3.94
```

```
## 3 DNK     2010            3.77
## 4 IRL     2010            3.68
## 5 ISL     2010            3.64
```