

DATA MANIPULATION WITH R

Ozan Bakış¹

¹Bahcesehir University, Department of Economics and BETAM

Filling missing values I

```
library(dplyr)
library(tidyr)
library(readxl)
f_url = "https://github.com/obakis/econ_data/raw/master/illere_gore_ihracat.xlsx"
download.file(url = f_url, destfile = "il_ihracat.xlsx", mode="wb")
dat = read_excel("il_ihracat.xlsx", col_names = TRUE,
                 range = "A5:P1458")
```

```
head(dat)
```

```
## # A tibble: 6 x 16
##   Year `Province code` `Province name` Total January February March
##   <chr> <chr>          <chr>          <chr> <chr>   <chr>   <chr>
## 1 <NA> <NA>          <NA>          <NA> <NA>   <NA>   <NA>
## 2 2018 <NA>          Toplam - Total 1245~ 124568~ <NA>   <NA>
## 3 <NA> <NA>          <NA>          <NA> <NA>   <NA>   <NA>
## 4 <NA> 0            Belirsiz- Nonspe~ 124.~ 124.199 <NA>   <NA>
## 5 <NA> 1            Adana          1503~ 150321~ <NA>   <NA>
## 6 <NA> 2            Adiyaman       1272~ 12722.~ <NA>   <NA>
## # i 9 more variables: April <chr>, May <chr>, June <chr>, July <chr>,
## #   August <chr>, September <chr>, October <chr>, November <chr>,
## #   December <chr>
```

Filling missing values II

```
dat = dat[,-c(3,4)] # drop prov names and total column
names(dat)[1:2] = c("year","province")
head(dat)
```

```
## # A tibble: 6 x 14
##   year province January February March April May June July August
##   <chr> <chr>   <chr>   <chr>   <chr> <chr> <chr> <chr> <chr> <chr>
## 1 <NA> <NA>   <NA>   <NA>   <NA> <NA> <NA> <NA> <NA> <NA>
## 2 2018 <NA>   124568~ <NA>   <NA>   <NA> <NA> <NA> <NA> <NA> <NA>
## 3 <NA> <NA>   <NA>   <NA>   <NA> <NA> <NA> <NA> <NA> <NA>
## 4 <NA> 0      124.199 <NA>   <NA>   <NA> <NA> <NA> <NA> <NA> <NA>
## 5 <NA> 1      150321~ <NA>   <NA>   <NA> <NA> <NA> <NA> <NA> <NA>
## 6 <NA> 2      12722.~ <NA>   <NA>   <NA> <NA> <NA> <NA> <NA> <NA>
## # i 4 more variables: September <chr>, October <chr>, November <chr>,
## #   December <chr>
```

```
str(dat)
```

Filling missing values III

```
## tibble [1,453 x 14] (S3: tbl_df/tbl/data.frame)
## $ year      : chr [1:1453] NA "2018" NA NA ...
## $ province  : chr [1:1453] NA NA NA "0" ...
## $ January   : chr [1:1453] NA "12456839.007999994" NA "124.199" ...
## $ February  : chr [1:1453] NA NA NA NA ...
## $ March     : chr [1:1453] NA NA NA NA ...
## $ April     : chr [1:1453] NA NA NA NA ...
## $ May       : chr [1:1453] NA NA NA NA ...
## $ June      : chr [1:1453] NA NA NA NA ...
## $ July      : chr [1:1453] NA NA NA NA ...
## $ August    : chr [1:1453] NA NA NA NA ...
## $ September: chr [1:1453] NA NA NA NA ...
## $ October   : chr [1:1453] NA NA NA NA ...
## $ November  : chr [1:1453] NA NA NA NA ...
## $ December  : chr [1:1453] NA NA NA NA ...
```

Filling missing values IV

```
dat = as_data_frame(dat)
# dat |>
# mutate_each(funs(extract_numeric), year:december) -> dat1
```

```
Nc = ncol(dat)
keep_rows = ifelse(rowSums(is.na(dat)) == Nc, FALSE, TRUE)
dat |>
  filter(keep_rows) |>
  transmute_all(extract_numeric) -> dat1
dat1[1:5,]
```

```
## # A tibble: 5 x 14
##   year province January February March April May June July August
##   <dbl>      <dbl> <dbl>      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  2018         NA  1.25e7      NA    NA    NA    NA    NA    NA    NA
## 2    NA          0  1.24e2      NA    NA    NA    NA    NA    NA    NA
## 3    NA          1  1.50e5      NA    NA    NA    NA    NA    NA    NA
## 4    NA          2  1.27e4      NA    NA    NA    NA    NA    NA    NA
## 5    NA          3  2.48e4      NA    NA    NA    NA    NA    NA    NA
## # i 4 more variables: September <dbl>, October <dbl>, November <dbl>,
## #   December <dbl>
```

Filling missing values V

```
dat1[83:89,]
```

```
## # A tibble: 7 x 14
##   year province January February March April May June
##   <dbl>   <dbl>   <dbl>   <dbl>   <dbl> <dbl> <dbl> <dbl>
## 1    NA      81    11204.      NA    NA      NA      NA      NA
## 2  2017     NA 11248475. 12090438. 1.45e7 1.29e7 1.36e7 1.31e7
## 3    NA      0     31.4      NA    5.98e0 1.37e1 1.33e1 5.27e1
## 4    NA      1  130337.  123554. 1.50e5 1.37e5 1.60e5 1.47e5
## 5    NA      2   12605.   8249. 1.13e4 6.44e3 9.50e3 6.43e3
## 6    NA      3   27496.  23421. 2.39e4 2.57e4 2.87e4 2.62e4
## 7    NA      4    3600.   4111. 3.36e3 3.23e3 3.18e3 2.47e3
## # i 6 more variables: July <dbl>, August <dbl>, September <dbl>,
## #   October <dbl>, November <dbl>, December <dbl>
```

```
dat2 = fill(dat1, year, .direction = "down")
head(dat2)
```

Filling missing values VI

```
## # A tibble: 6 x 14
##   year province January February March April May June July August
##   <dbl>   <dbl>   <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  2018     NA  1.25e7     NA   NA   NA   NA   NA   NA   NA   NA
## 2  2018      0  1.24e2     NA   NA   NA   NA   NA   NA   NA   NA
## 3  2018      1  1.50e5     NA   NA   NA   NA   NA   NA   NA   NA
## 4  2018      2  1.27e4     NA   NA   NA   NA   NA   NA   NA   NA
## 5  2018      3  2.48e4     NA   NA   NA   NA   NA   NA   NA   NA
## 6  2018      4  2.78e3     NA   NA   NA   NA   NA   NA   NA   NA
## # i 4 more variables: September <dbl>, October <dbl>, November <dbl>,
## #   December <dbl>
```

```
dat2 = dat2 |>
  filter(! province %in% c(0,NA))
head(dat2)
```

Filling missing values VII

```
## # A tibble: 6 x 14
##   year province January February March April May June July August
##   <dbl>   <dbl>   <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  2018     1 150322.     NA     NA     NA     NA     NA     NA     NA
## 2  2018     2  12722.     NA     NA     NA     NA     NA     NA     NA
## 3  2018     3  24786.     NA     NA     NA     NA     NA     NA     NA
## 4  2018     4   2776.     NA     NA     NA     NA     NA     NA     NA
## 5  2018     5   9008.     NA     NA     NA     NA     NA     NA     NA
## 6  2018     6 529935.     NA     NA     NA     NA     NA     NA     NA
## # i 4 more variables: September <dbl>, October <dbl>, November <dbl>,
## #   December <dbl>

dat_x1 = pivot_longer(data=dat2, cols = -c(province, year), names_to = "month", values_to = "export")
head(dat_x1)
```

Filling missing values VIII

```
## # A tibble: 6 x 4
##   year province month      export
##   <dbl>   <dbl> <chr>      <dbl>
## 1  2018         1 January  150322.
## 2  2018         1 February    NA
## 3  2018         1 March      NA
## 4  2018         1 April      NA
## 5  2018         1 May        NA
## 6  2018         1 June       NA
```

```
dat_x1 |>
  mutate(month = factor(month, levels = month.name)) |>
  arrange(year, month, province) -> dat_x
print(dat_x, n=3)
```

Filling missing values IX

```
## # A tibble: 16,452 x 4
##   year province month   export
##   <dbl>   <dbl> <fct>   <dbl>
## 1  2002         1 January 35247.
## 2  2002         2 January   740.
## 3  2002         3 January  3163.
## # i 16,449 more rows
```

```
saveRDS(dat_x, "tur_x.rds")
```

Data reshaping I

```
f_url = "https://github.com/obakis/econ_data/raw/master/illere_gore_ithalat.xlsx"
download.file(url = f_url, destfile = "il_ithalat.xlsx", mode="wb")
dat = read_excel("il_ithalat.xlsx", col_names = TRUE,
                 range = "A5:P1471")
```

```
head(dat)
```

```
## # A tibble: 6 x 16
##   Year `Province code` `Province name` Total January February March
##   <chr> <chr>          <chr>          <chr> <chr>   <chr>   <chr>
## 1 <NA> <NA>          <NA>          <NA> <NA>   <NA>   <NA>
## 2 2018 <NA>          Toplam - Total 2152~ 215239~ <NA>   <NA>
## 3 <NA> <NA>          <NA>          <NA> <NA>   <NA>   <NA>
## 4 <NA> 0            Belirsiz- Nonspe~ 160.~ 160.869 <NA>   <NA>
## 5 <NA> 1            Adana          2308~ 230840~ <NA>   <NA>
## 6 <NA> 2            Adiyaman       3082~ 3082.2~ <NA>   <NA>
## # i 9 more variables: April <chr>, May <chr>, June <chr>, July <chr>,
## #   August <chr>, September <chr>, October <chr>, November <chr>,
## #   December <chr>
```

Data reshaping II

```
dat = dat[,-c(3,4)]
names(dat)[1:2] = c("year","province")

dat = as_data_frame(dat)
Nc = ncol(dat)
keep_rows = ifelse(rowSums(is.na(dat)) == Nc, FALSE, TRUE)
dat |>
  filter(keep_rows) |>
  transmute_all(extract_numeric) -> dat1
head(dat1)

## # A tibble: 6 x 14
##   year province January February March April May June July August
##   <dbl>   <dbl>   <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  2018     NA  2.15e7     NA    NA    NA    NA    NA    NA    NA
## 2    NA      0  1.61e2     NA    NA    NA    NA    NA    NA    NA
## 3    NA      1  2.31e5     NA    NA    NA    NA    NA    NA    NA
## 4    NA      2  3.08e3     NA    NA    NA    NA    NA    NA    NA
## 5    NA      3  9.70e3     NA    NA    NA    NA    NA    NA    NA
## 6    NA      4  3.06e4     NA    NA    NA    NA    NA    NA    NA
## # i 4 more variables: September <dbl>, October <dbl>, November <dbl>,
```

Data reshaping III

```
## #   December <dbl>
```

```
dat2 = fill(dat1, year, .direction = "down")  
head(dat2)
```

```
## # A tibble: 6 x 14
```

```
##   year province January February March April May June July August  
##   <dbl>   <dbl>   <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1  2018     NA  2.15e7     NA    NA    NA    NA    NA    NA    NA  
## 2  2018     0  1.61e2     NA    NA    NA    NA    NA    NA    NA  
## 3  2018     1  2.31e5     NA    NA    NA    NA    NA    NA    NA  
## 4  2018     2  3.08e3     NA    NA    NA    NA    NA    NA    NA  
## 5  2018     3  9.70e3     NA    NA    NA    NA    NA    NA    NA  
## 6  2018     4  3.06e4     NA    NA    NA    NA    NA    NA    NA  
## # i 4 more variables: September <dbl>, October <dbl>, November <dbl>,  
## #   December <dbl>
```

```
dat2 = dat2 |>  
  filter(! province %in% c(0,NA,99))  
head(dat2)
```

Data reshaping IV

```
## # A tibble: 6 x 14
##   year province January February March April May June July August
##   <dbl>   <dbl>   <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  2018         1  2.31e5      NA    NA    NA    NA    NA    NA    NA
## 2  2018         2  3.08e3      NA    NA    NA    NA    NA    NA    NA
## 3  2018         3  9.70e3      NA    NA    NA    NA    NA    NA    NA
## 4  2018         4  3.06e4      NA    NA    NA    NA    NA    NA    NA
## 5  2018         5  4.43e3      NA    NA    NA    NA    NA    NA    NA
## 6  2018         6  1.24e6      NA    NA    NA    NA    NA    NA    NA
## # i 4 more variables: September <dbl>, October <dbl>, November <dbl>,
## #   December <dbl>

dat_m1 = pivot_longer(data=dat2, cols = -c(province, year), names_to = "month", values_to = "import")
head(dat_m1)
```

Data reshaping V

```
## # A tibble: 6 x 4
##   year province month      import
##   <dbl>   <dbl> <chr>      <dbl>
## 1  2018         1 January  230840.
## 2  2018         1 February    NA
## 3  2018         1 March      NA
## 4  2018         1 April      NA
## 5  2018         1 May        NA
## 6  2018         1 June       NA
```

```
dat_m1 |>
  mutate(month = factor(month, levels = month.name)) |>
  arrange(year, month, province) -> dat_m
print(dat_m, n=3)
```

Data reshaping VI

```
## # A tibble: 16,488 x 4
##   year province month   import
##   <dbl>   <dbl> <fct>   <dbl>
## 1  2002         1 January 44761.
## 2  2002         2 January  1868.
## 3  2002         3 January  1295.
## # i 16,485 more rows
```

```
saveRDS(dat_m, "tur_m.rds")
```

Data reshaping VII

```
f_url = "https://github.com/obakis/econ_data/raw/master/illere_gore_gsyh.xlsx"
download.file(url = f_url, destfile = "il_gsyh.xlsx", mode="wb")
dat = read_excel("il_gsyh.xlsx", col_names = FALSE,
                 range = "A9:BZ89")
```

```
head(dat)
```

```
## # A tibble: 6 x 78
##   ...1 ...2 ...3 ...4 ...5 ...6 ...7 ...8 ...9 ...10
##   <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <lg1> <dbl>
## 1 TR100 İstanb~ 5.30e5 4.71e7 1.04e8 1.51e8 2.18e7 1.73e8 NA 6.02e5
## 2 TR211 Tekird~ 8.33e5 3.51e6 2.41e6 6.75e6 9.74e5 7.72e6 NA 8.78e5
## 3 TR212 Edirne 8.47e5 4.60e5 1.23e6 2.54e6 3.66e5 2.90e6 NA 9.51e5
## 4 TR213 Kırkla~ 5.27e5 1.06e6 9.46e5 2.54e6 3.66e5 2.90e6 NA 5.63e5
## 5 TR221 Balıke~ 1.54e6 1.57e6 4.11e6 7.22e6 1.04e6 8.26e6 NA 1.81e6
## 6 TR222 Çanakk~ 9.65e5 7.99e5 1.71e6 3.47e6 5.01e5 3.97e6 NA 1.15e6
## # i 68 more variables: ...11 <dbl>, ...12 <dbl>, ...13 <dbl>,
## # ...14 <dbl>, ...15 <dbl>, ...16 <lg1>, ...17 <dbl>, ...18 <dbl>,
## # ...19 <dbl>, ...20 <dbl>, ...21 <dbl>, ...22 <dbl>, ...23 <lg1>,
## # ...24 <dbl>, ...25 <dbl>, ...26 <dbl>, ...27 <dbl>, ...28 <dbl>,
## # ...29 <dbl>, ...30 <lg1>, ...31 <dbl>, ...32 <dbl>, ...33 <dbl>,
## # ...34 <dbl>, ...35 <dbl>, ...36 <dbl>, ...37 <lg1>, ...38 <dbl>,
```

Data reshaping VIII

```
## # ...39 <dbl>, ...40 <dbl>, ...41 <dbl>, ...42 <dbl>, ...
```

```
keep_cols = colSums(is.na(dat)) < nrow(dat)  
keep_cols
```

```
## ...1 ...2 ...3 ...4 ...5 ...6 ...7 ...8 ...9 ...10 ...11  
## TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE FALSE TRUE TRUE  
## ...12 ...13 ...14 ...15 ...16 ...17 ...18 ...19 ...20 ...21 ...22  
## TRUE TRUE TRUE TRUE FALSE TRUE TRUE TRUE TRUE TRUE TRUE  
## ...23 ...24 ...25 ...26 ...27 ...28 ...29 ...30 ...31 ...32 ...33  
## FALSE TRUE TRUE TRUE TRUE TRUE TRUE FALSE TRUE TRUE TRUE  
## ...34 ...35 ...36 ...37 ...38 ...39 ...40 ...41 ...42 ...43 ...44  
## TRUE TRUE TRUE FALSE TRUE TRUE TRUE TRUE TRUE TRUE FALSE  
## ...45 ...46 ...47 ...48 ...49 ...50 ...51 ...52 ...53 ...54 ...55  
## TRUE TRUE TRUE TRUE TRUE TRUE FALSE TRUE TRUE TRUE TRUE  
## ...56 ...57 ...58 ...59 ...60 ...61 ...62 ...63 ...64 ...65 ...66  
## TRUE TRUE FALSE TRUE TRUE TRUE TRUE TRUE TRUE FALSE TRUE  
## ...67 ...68 ...69 ...70 ...71 ...72 ...73 ...74 ...75 ...76 ...77  
## TRUE TRUE TRUE TRUE TRUE FALSE TRUE TRUE TRUE TRUE TRUE  
## ...78  
## TRUE
```

Data reshaping IX

```
dat = dat[,keep_cols]
```

```
head(dat)
```

```
## # A tibble: 6 x 68
```

```
##   ...1 ...2 ...3 ...4 ...5 ...6 ...7 ...8 ...10 ...11
##   <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 TR100 İstan~ 5.30e5 4.71e7 1.04e8 1.51e8 2.18e7 1.73e8 6.02e5 5.42e7
## 2 TR211 Tekir~ 8.33e5 3.51e6 2.41e6 6.75e6 9.74e5 7.72e6 8.78e5 4.05e6
## 3 TR212 Edirne 8.47e5 4.60e5 1.23e6 2.54e6 3.66e5 2.90e6 9.51e5 5.36e5
## 4 TR213 Kırkl~ 5.27e5 1.06e6 9.46e5 2.54e6 3.66e5 2.90e6 5.63e5 1.25e6
## 5 TR221 Balık~ 1.54e6 1.57e6 4.11e6 7.22e6 1.04e6 8.26e6 1.81e6 1.88e6
## 6 TR222 Çanak~ 9.65e5 7.99e5 1.71e6 3.47e6 5.01e5 3.97e6 1.15e6 9.58e5
## # i 58 more variables: ...12 <dbl>, ...13 <dbl>, ...14 <dbl>,
## #   ...15 <dbl>, ...17 <dbl>, ...18 <dbl>, ...19 <dbl>, ...20 <dbl>,
## #   ...21 <dbl>, ...22 <dbl>, ...24 <dbl>, ...25 <dbl>, ...26 <dbl>,
## #   ...27 <dbl>, ...28 <dbl>, ...29 <dbl>, ...31 <dbl>, ...32 <dbl>,
## #   ...33 <dbl>, ...34 <dbl>, ...35 <dbl>, ...36 <dbl>, ...38 <dbl>,
## #   ...39 <dbl>, ...40 <dbl>, ...41 <dbl>, ...42 <dbl>, ...43 <dbl>,
## #   ...45 <dbl>, ...46 <dbl>, ...47 <dbl>, ...48 <dbl>, ...
```

Data reshaping X

```
years = 2004:2014
vars = c("agr", "ind", "ser", "sectot", "tax", "gdp")
Nyr = length(years)
Nvar = length(vars)

vec_var = rep(vars, Nyr)
vec_yr = rep(years, each=Nvar)
nms1 = paste(vec_var, vec_yr, sep="_")
nms = c("nuts3", "province", nms1)
colnames(dat)=nms
#head(as.data.frame(dat))
dat$province=NULL
dat1 = pivot_longer(data=dat, cols = -nuts3, names_to = "output",
                    values_to = "TL")

head(dat1)
```

Data reshaping XI

```
## # A tibble: 6 x 3
##   nuts3 output          TL
##   <chr> <chr>          <dbl>
## 1 TR100 agr_2004      530330.
## 2 TR100 ind_2004     47066568.
## 3 TR100 ser_2004    103603732.
## 4 TR100 sectot_2004 151200630.
## 5 TR100 tax_2004     21818414.
## 6 TR100 gdp_2004    173019044.
```

```
dat2 = dat1 |> separate(output, c("out", "year"))
head(dat2)
```

```
## # A tibble: 6 x 4
##   nuts3 out   year          TL
##   <chr> <chr> <chr>          <dbl>
## 1 TR100 agr   2004      530330.
## 2 TR100 ind   2004     47066568.
## 3 TR100 ser   2004    103603732.
## 4 TR100 sectot 2004   151200630.
## 5 TR100 tax   2004     21818414.
## 6 TR100 gdp   2004    173019044.
```

Data reshaping XII

```
dat3 = dat2 |>
  pivot_wider(names_from = "out", values_from = "TL")
print(dat3, n=3)

## # A tibble: 891 x 8
##   nuts3 year    agr      ind      ser    sectot    tax    gdp
##   <chr> <chr>  <dbl>   <dbl>   <dbl>   <dbl>   <dbl> <dbl>
## 1 TR100 2004  530330. 47066568. 103603732. 151200630. 21818414. 1.73e8
## 2 TR100 2005  601554. 54203103. 120996325. 175800982. 25517310. 2.01e8
## 3 TR100 2006  566282. 65537887. 141228209. 207332377. 29728690. 2.37e8
## # i 888 more rows

saveRDS(dat3, "tur_gdp.rds")
```

Data reshaping XIII

```
f_url = "https://github.com/obakis/econ_data/raw/master/illere_gore_isgucu.xlsx"
download.file(url = f_url, destfile = "il_isgucu.xlsx", mode="wb")
```

```
dat = read_excel("il_isgucu.xlsx", col_names = TRUE)
head(dat)
```

```
## # A tibble: 6 x 7
##   pr_no pr_name      lfp_rate un_rate emp_rate year nuts3
##   <dbl> <chr>      <dbl>   <dbl>   <dbl> <dbl> <chr>
## 1     1 Adana          49     26.5     36    2008 TR621
## 2     2 Adıyaman       38     17.9    31.2    2008 TRC12
## 3     3 Afyonkarahisar 44.7    10.8    39.9    2008 TR332
## 4     4 Ağrı           48     10.1    43.2    2008 TRA21
## 5     5 Amasya        56.2     6.9    52.4    2008 TR834
## 6     6 Ankara        44.9    13.6    38.8    2008 TR510
```

```
saveRDS(dat, "tur_labor.rds")
saveRDS(dat[1:81, c("pr_no", "nuts3")], "province-nuts3.rds")
```

Joining data frames I

```
##See http://dplyr.tidyverse.org/reference/join.html for more on joining
tur_m = readRDS("tur_m.rds")
tur_x = readRDS("tur_x.rds")
tur_xm = full_join(tur_m, tur_x, by=c("year", "province", "month"))
tur_xm |>
  arrange(year, month, province) -> tur_xm
print(tur_xm, n=3)

## # A tibble: 16,512 x 5
##   year province month   import export
##   <dbl>   <dbl> <fct>   <dbl> <dbl>
## 1  2002         1 January 44761. 35247.
## 2  2002         2 January  1868.   740.
## 3  2002         3 January  1295.  3163.
## # i 16,509 more rows

saveRDS(tur_xm, "tur_xm.rds")
```

Joining data frames II

```
# f_url = "https://github.com/obakis/econ_data/raw/master/tur_xm.rds"
# download.file(url = f_url, destfile = "tur_xm.rds", mode="wb")
# f_url = "https://github.com/obakis/econ_data/raw/master/tur_labor.rds"
# download.file(url = f_url, destfile = "tur_labor.rds", mode="wb")
xm = readRDS("tur_xm.rds")
lab = readRDS("tur_labor.rds")

ihs <- function(x){
  log(x + sqrt(x**2 + 1))
}

library(dplyr)
xm |>
  group_by(province, year) |>
  summarise(
    export = sum(export, na.rm=TRUE),
    import = sum(import, na.rm=TRUE)
  ) |>
  group_by(province) |>
  arrange(province, year) |>
  mutate(
```

Joining data frames III

```
ihs_x = ihs(export),
ihs_m = ihs(import)
) |>
mutate(
  gr_x = 100*(ihs_x - dplyr::lag(ihs_x))/dplyr::lag(ihs_x),
  gr_m = 100*(ihs_m - dplyr::lag(ihs_m))/dplyr::lag(ihs_m)
) |>
rename(pr_no = province) |>
mutate(
  gr_x = ifelse(is.na(gr_x) | is.infinite(gr_x), NA, gr_x),
  gr_m = ifelse(is.na(gr_m) | is.infinite(gr_m), NA, gr_m)
) -> xm_y
dat1 = inner_join(lab, xm_y, by=c("year", "pr_no"))
dat1 |> select(-pr_name) -> dat
head(dat, 3)
```

Joining data frames IV

```
## # A tibble: 3 x 12
##   pr_no lfp_rate un_rate emp_rate year nuts3 export import ihs_x
##   <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <dbl> <dbl> <dbl>
## 1     1     49    26.5    36   2008 TR621 1304024. 2151647. 14.8
## 2     2     38    17.9    31.2 2008 TRC12   59103.   36292.   11.7
## 3     3    44.7    10.8    39.9 2008 TR332 237839.   34370.   13.1
## # i 3 more variables: ihs_m <dbl>, gr_x <dbl>, gr_m <dbl>
```

```
saveRDS(dat, "tur_xmlab.rds")
```

```
xm |>
  filter(year %in% c(2009,2010)) |>
  group_by(province, year) |>
  summarise(
    export = sum(export, na.rm=TRUE),
    import = sum(import, na.rm=TRUE)
  ) |>
  group_by(province) |>
  arrange(province, year) |>
  mutate(
    ihs_x = ihs(export),
```

Joining data frames V

```
  ihs_m = ihs(import)
) |>
mutate(
  gr_x = 100*(ihs_x - dplyr::lag(ihs_x))/dplyr::lag(ihs_x),
  gr_m = 100*(ihs_m - dplyr::lag(ihs_m))/dplyr::lag(ihs_m)
) |>
rename(pr_no = province) |>
mutate(
  gr_x = ifelse(is.na(gr_x) | is.infinite(gr_x), NA, gr_x),
  gr_m = ifelse(is.na(gr_m) | is.infinite(gr_m), NA, gr_m)
) -> xm_2y
xm_2y

## # A tibble: 162 x 8
## # Groups:   pr_no [81]
##   pr_no year  export  import ihs_x ihs_m  gr_x  gr_m
##   <dbl> <dbl>   <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     1   2009 1135887. 1692782.  14.6  15.0  NA    NA
## 2     1   2010 1352306. 2229404.  14.8  15.3  1.19  1.83
## 3     2   2009  58091.  33336.  11.7  11.1  NA    NA
## 4     2   2010  71639.  85425.  11.9  12.0  1.80  8.47
```

Joining data frames VI

```
## 5      3  2009  208636.   40512.  12.9  11.3 NA      NA
## 6      3  2010  217496.   72668.  13.0  11.9 0.321  5.17
## 7      4  2009   44339.   45227.  11.4  11.4 NA      NA
## 8      4  2010   76904.   58973.  11.9  11.7 4.83   2.33
## 9      5  2009   21629.   13072.  10.7  10.2 NA      NA
## 10     5  2010   53018.   41629.  11.6  11.3 8.40   11.4
## # i 152 more rows
```

```
dat1 = inner_join(lab, xm_2y, by=c("year", "pr_no"))
dat1 |> select(-pr_name) -> dat2y
head(dat2y, 3)
```

```
## # A tibble: 3 x 12
##   pr_no lfp_rate un_rate emp_rate year nuts3 export import ihs_x
##   <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <dbl> <dbl> <dbl>
## 1     1    45.6   20.5   36.2  2009 TR621 1135887. 1692782.  14.6
## 2     2    42.1   16.5   35.1  2009 TRC12  58091.  33336.  11.7
## 3     3     44     7.7   40.6  2009 TR332  208636.  40512.  12.9
## # i 3 more variables: ihs_m <dbl>, gr_x <dbl>, gr_m <dbl>
```

```
saveRDS(dat2y, "tur_xmlab2y.rds")
```