Bureau of Labor Statistics (BLS)

R Statistics

François Geerolf

Load packages:

Instructions are taken from the following url:

```
url <- "https://download.bls.gov/pub/time.series/in/"
```

Scrapping Data

Elements of the scrapping data:

```
read_html("https://download.bls.gov/pub/time.series/in/") %>%
  html_nodes("a") %>%
  html_text(trim = TRUE) %>%
  as.data.frame %>%
  rename(X0 = ".") %>%
  as.tibble
```

```
# # A tibble: 14 x 1
#    X0
#    <fct>
#  1 [To Parent Directory]
#  2 in.contacts
#  3 in.country
#  4 in.data.0.Current
#  5 in.data.1.AllData
#  6 in.economicgroup
#  7 in.economicseries
#  8 in.footnote
#  9 in.industry
# 10 in.period
# 11 in.periodicity
# 12 in.seasonal
# 13 in.series
# 14 in.txt
```

```
read_html("https://download.bls.gov/pub/time.series/in/") %>%
  str_match_all("<a href=\"(.*?)\"") %>%
  as.data.frame %>%
          mutate(X2 = paste0("https://download.bls.gov", X2)) %>%
  as.tibble
```

```
# Warning in stri_match_all_regex(string, pattern, omit_no_match = TRUE, opts_regex = opts(pattern)): a
```

```
# # A tibble: 14 x 2
#    X1                                                  X2
#    <fct>                                               <chr>
#  1 "<a href=\"/pub/time.series/\""                     https://download.bls.gov/pub/time.series/
#  2 "<a href=\"/pub/time.series/in/in.contacts\""       https://download.bls.gov/pub/time.series/in/in
#  3 "<a href=\"/pub/time.series/in/in.country\""        https://download.bls.gov/pub/time.series/in/in
#  4 "<a href=\"/pub/time.series/in/in.data.0.Current\"" https://download.bls.gov/pub/time.series/in/in
#  5 "<a href=\"/pub/time.series/in/in.data.1.AllData\"" https://download.bls.gov/pub/time.series/in/in
#  6 "<a href=\"/pub/time.series/in/in.economicgroup\""  https://download.bls.gov/pub/time.series/in/in
#  7 "<a href=\"/pub/time.series/in/in.economicseries\"" https://download.bls.gov/pub/time.series/in/in
#  8 "<a href=\"/pub/time.series/in/in.footnote\""       https://download.bls.gov/pub/time.series/in/in
#  9 "<a href=\"/pub/time.series/in/in.industry\""       https://download.bls.gov/pub/time.series/in/in
# 10 "<a href=\"/pub/time.series/in/in.period\""         https://download.bls.gov/pub/time.series/in/in
# 11 "<a href=\"/pub/time.series/in/in.periodicity\""    https://download.bls.gov/pub/time.series/in/in
# 12 "<a href=\"/pub/time.series/in/in.seasonal\""       https://download.bls.gov/pub/time.series/in/in
# 13 "<a href=\"/pub/time.series/in/in.series\""         https://download.bls.gov/pub/time.series/in/in
# 14 "<a href=\"/pub/time.series/in/in.txt\""            https://download.bls.gov/pub/time.series/in/in
```

```r
datasets <- read_html("https://download.bls.gov/pub/time.series/in/") %>%
  html_nodes("a") %>%
  html_text(trim = TRUE) %>%
  as.data.frame %>%
  rename(X0 = ".") %>%
  cbind(read_html("https://download.bls.gov/pub/time.series/in/") %>%
          str_match_all("<a href=\"(.*?)\"") %>%
          as.data.frame %>%
          mutate(X2 = paste0("https://download.bls.gov", X2))) %>%
  mutate_all(paste)
```

```
# Warning in stri_match_all_regex(string, pattern, omit_no_match = TRUE, opts_regex = opts(pattern)): a
```

```r
datasets %>%
  as.tibble
```

```
# # A tibble: 14 x 3
#    X0                    X1                                                  X2
#    <chr>                 <chr>                                               <chr>
#  1 [To Parent Directory] "<a href=\"/pub/time.series/\""                     https://download.bls.gov
#  2 in.contacts           "<a href=\"/pub/time.series/in/in.contacts\""       https://download.bls.gov
#  3 in.country            "<a href=\"/pub/time.series/in/in.country\""        https://download.bls.gov
#  4 in.data.0.Current     "<a href=\"/pub/time.series/in/in.data.0.Current\"" https://download.bls.gov
#  5 in.data.1.AllData     "<a href=\"/pub/time.series/in/in.data.1.AllData\"" https://download.bls.gov
#  6 in.economicgroup      "<a href=\"/pub/time.series/in/in.economicgroup\""  https://download.bls.gov
#  7 in.economicseries     "<a href=\"/pub/time.series/in/in.economicseries\"" https://download.bls.gov
#  8 in.footnote           "<a href=\"/pub/time.series/in/in.footnote\""       https://download.bls.gov
#  9 in.industry           "<a href=\"/pub/time.series/in/in.industry\""       https://download.bls.gov
# 10 in.period             "<a href=\"/pub/time.series/in/in.period\""         https://download.bls.gov
# 11 in.periodicity        "<a href=\"/pub/time.series/in/in.periodicity\""    https://download.bls.gov
# 12 in.seasonal           "<a href=\"/pub/time.series/in/in.seasonal\""       https://download.bls.gov
# 13 in.series             "<a href=\"/pub/time.series/in/in.series\""         https://download.bls.gov
# 14 in.txt                "<a href=\"/pub/time.series/in/in.txt\""            https://download.bls.gov
```

Downloading all data

```r
setwd("/Users/geerolf/Drive/work/datasets/bls-ilc/")

for (i in 4:13){
  file <- datasets[i, "X0"]
  cat("\nDownloading from BLS Website ILC:", file)
  assign(file, read.csv(datasets[i, "X2"], sep = "\t"))
  do.call(save, list(file, file = paste0(file, ".RData")))
}
```

```
#
# Downloading from BLS Website ILC: in.data.0.Current
# Downloading from BLS Website ILC: in.data.1.AllData
# Downloading from BLS Website ILC: in.economicgroup
# Downloading from BLS Website ILC: in.economicseries
# Downloading from BLS Website ILC: in.footnote
# Downloading from BLS Website ILC: in.industry
# Downloading from BLS Website ILC: in.period
# Downloading from BLS Website ILC: in.periodicity
# Downloading from BLS Website ILC: in.seasonal
# Downloading from BLS Website ILC: in.series
```

```
rm(datasets)
```