

# BLS - CEX - Download

Datasets

*François Geerolf*

## Preamble

```
rm(list = ls())
pklist <- c("curl", "tidyverse", "rvest")
source("https://fgeerolf.github.io/datasets/load-packages.R")
options(tibble.print_max = 100)
```

## Introduction

The data for the CEX is available here: <https://www.bls.gov/cex/>

The flat data files of the CEX are: <https://download.bls.gov/pub/time.series/cx/>

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

## Scrapping Data

Elements of the scrapping data:

```
read_html(url) %>%
  html_nodes("a") %>%
  html_text(trim = TRUE) %>%
  as.data.frame %>%
  rename(X0 = ".") %>%
  as.tibble
```

```
# # A tibble: 12 x 1
#   X0
#   <fct>
# 1 [To Parent Directory]
# 2 cx.category
# 3 cx.characteristics
# 4 cx.contacts
# 5 cx.data.1.AllData
# 6 cx.demographics
# 7 cx.footnote
# 8 cx.item
# 9 cx.process
# 10 cx.series
# 11 cx.subcategory
# 12 cx.txt
```

```
read_html(url) %>%
  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)): argument is not an atomic vector; coercing

# # A tibble: 12 x 2
#   X1                                X2
#   <fct>                            <chr>
# 1 "<a href=\"/pub/time.series/\""  https://download.bls.gov/pub/time.se~
# 2 "<a href=\"/pub/time.series/cx/c~ https://download.bls.gov/pub/time.se~
# 3 "<a href=\"/pub/time.series/cx/c~ https://download.bls.gov/pub/time.se~
# 4 "<a href=\"/pub/time.series/cx/c~ https://download.bls.gov/pub/time.se~
# 5 "<a href=\"/pub/time.series/cx/c~ https://download.bls.gov/pub/time.se~
# 6 "<a href=\"/pub/time.series/cx/c~ https://download.bls.gov/pub/time.se~
# 7 "<a href=\"/pub/time.series/cx/c~ https://download.bls.gov/pub/time.se~
# 8 "<a href=\"/pub/time.series/cx/c~ https://download.bls.gov/pub/time.se~
# 9 "<a href=\"/pub/time.series/cx/c~ https://download.bls.gov/pub/time.se~
# 10 "<a href=\"/pub/time.series/cx/c~ https://download.bls.gov/pub/time.se~
# 11 "<a href=\"/pub/time.series/cx/c~ https://download.bls.gov/pub/time.se~
# 12 "<a href=\"/pub/time.series/cx/c~ https://download.bls.gov/pub/time.se~

datasets <- read_html(url) %>%
  html_nodes("a") %>%
  html_text(trim = TRUE) %>%
  as.data.frame %>%
  rename(X0 = ".") %>%
  cbind(read_html(url) %>%
    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)): argument is not an atomic vector; coercing

datasets %>%
  as.tibble

# # A tibble: 12 x 3
#   X0                                X1                                X2
#   <chr>                            <chr>                            <chr>
# 1 [To Parent D~ "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
# 2 cx.category   "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
# 3 cx.character~ "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
# 4 cx.contacts   "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
# 5 cx.data.1.Al~ "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
# 6 cx.demograph~ "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
# 7 cx.footnote   "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
# 8 cx.item       "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
# 9 cx.process    "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
# 10 cx.series     "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
# 11 cx.subcatego~ "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
# 12 cx.txt        "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~

```

## Downloading all data

```
for (i in 2:11){
  file <- datasets[i, "X0"]
  cat("\nDownloading from BLS Website CEX:", file)
  assign(file, read.csv(datasets[i, "X2"], sep = "\t"))
  do.call(save, list(file, file = paste0(file, ".RData")))
}
```

```
#
# Downloading from BLS Website CEX: cx.category
# Downloading from BLS Website CEX: cx.characteristics
# Downloading from BLS Website CEX: cx.contacts
# Downloading from BLS Website CEX: cx.data.1.AllData
# Downloading from BLS Website CEX: cx.demographics
# Downloading from BLS Website CEX: cx.footnote
# Downloading from BLS Website CEX: cx.item
# Downloading from BLS Website CEX: cx.process
# Downloading from BLS Website CEX: cx.series
# Downloading from BLS Website CEX: cx.subcategory

rm(datasets)
```

## Computing Environment

```
Sys.time()
```

```
## [1] "2018-09-23 22:12:06 PDT"
```

```
sessionInfo()
```

```
## R version 3.5.1 (2018-07-02)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS High Sierra 10.13.6
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] bindrcpp_0.2.2  rvest_0.3.2      xml2_1.2.0      forcats_0.3.0
## [5] stringr_1.3.1   dplyr_0.7.6      purrr_0.2.5     readr_1.1.1
## [9] tidyr_0.8.1     tibble_1.4.2     ggplot2_3.0.0   tidyverse_1.2.1
## [13] curl_3.2
##
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.18    cellranger_1.1.0 pillar_1.3.0    compiler_3.5.1
```

```
## [5] plyr_1.8.4      bindr_0.1.1      tools_3.5.1      digest_0.6.15
## [9] lubridate_1.7.4 jsonlite_1.5     evaluate_0.11    nlme_3.1-137
## [13] gtable_0.2.0    lattice_0.20-35  pkgconfig_2.0.2  rlang_0.2.2
## [17] cli_1.0.0       rstudioapi_0.7   yaml_2.2.0       haven_1.1.2
## [21] withr_2.1.2     httr_1.3.1       knitr_1.20       hms_0.4.2
## [25] rprojroot_1.3-2 grid_3.5.1       tidyselect_0.2.4 glue_1.3.0
## [29] R6_2.2.2        fansi_0.3.0      readxl_1.1.0     rmarkdown_1.10
## [33] selectr_0.4-1   modelr_0.1.2     magrittr_1.5     backports_1.1.2
## [37] scales_1.0.0    htmltools_0.3.6  assertthat_0.2.0 colorspace_1.3-2
## [41] utf8_1.1.4      stringi_1.2.4    lazyeval_0.2.1   munsell_0.5.0
## [45] broom_0.5.0     crayon_1.3.4
```