

BLS - All Datasets - Download

Datasets

François Geerolf

Contents

Preamble	1
Scrapping Data	1
Datasets from the BLS	1
Datasets from the BLS	2
Elements of the scrapping data	3
List all Datasets	4
Downloading	5
Computing Environment	6

Preamble

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

Link is the following url:

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

Scrapping Data

Datasets from the BLS

Datasets:

```
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 from the BLS

```
datasets %>%
  select(X0, X2) %>%
  as.tibble
```

```
## # A tibble: 69 x 2
##   X0                X2
##   <chr>            <chr>
## 1 [To Parent Director~ https://download.bls.gov/pub/
## 2 ap               https://download.bls.gov/pub/time.series/ap/
## 3 bd               https://download.bls.gov/pub/time.series/bd/
## 4 bg               https://download.bls.gov/pub/time.series/bg/
## 5 bp               https://download.bls.gov/pub/time.series/bp/
## 6 cc               https://download.bls.gov/pub/time.series/cc/
## 7 cd               https://download.bls.gov/pub/time.series/cd/
## 8 ce               https://download.bls.gov/pub/time.series/ce/
## 9 cf               https://download.bls.gov/pub/time.series/cf/
## 10 ch              https://download.bls.gov/pub/time.series/ch/
## 11 ci              https://download.bls.gov/pub/time.series/ci/
## 12 cm              https://download.bls.gov/pub/time.series/cm/
## 13 compressed      https://download.bls.gov/pub/time.series/compress~
## 14 cs              https://download.bls.gov/pub/time.series/cs/
## 15 cu              https://download.bls.gov/pub/time.series/cu/
## 16 cw              https://download.bls.gov/pub/time.series/cw/
## 17 cx              https://download.bls.gov/pub/time.series/cx/
## 18 eb              https://download.bls.gov/pub/time.series/eb/
## 19 ec              https://download.bls.gov/pub/time.series/ec/
## 20 ee              https://download.bls.gov/pub/time.series/ee/
## 21 ei              https://download.bls.gov/pub/time.series/ei/
## 22 ep              https://download.bls.gov/pub/time.series/ep/
## 23 esbr            https://download.bls.gov/pub/time.series/esbr/
## 24 fi              https://download.bls.gov/pub/time.series/fi/
## 25 fm              https://download.bls.gov/pub/time.series/fm/
## 26 fw              https://download.bls.gov/pub/time.series/fw/
## 27 gg              https://download.bls.gov/pub/time.series/gg/
## 28 gp              https://download.bls.gov/pub/time.series/gp/
## 29 hc              https://download.bls.gov/pub/time.series/hc/
## 30 hs              https://download.bls.gov/pub/time.series/hs/
## 31 ii              https://download.bls.gov/pub/time.series/ii/
## 32 in              https://download.bls.gov/pub/time.series/in/
## 33 ip              https://download.bls.gov/pub/time.series/ip/
## 34 is              https://download.bls.gov/pub/time.series/is/
## 35 jl              https://download.bls.gov/pub/time.series/jl/
## 36 jt              https://download.bls.gov/pub/time.series/jt/
## 37 la              https://download.bls.gov/pub/time.series/la/
## 38 le              https://download.bls.gov/pub/time.series/le/
## 39 li              https://download.bls.gov/pub/time.series/li/
## 40 ln              https://download.bls.gov/pub/time.series/ln/
## 41 lu              https://download.bls.gov/pub/time.series/lu/
## 42 ml              https://download.bls.gov/pub/time.series/ml/
## 43 mp              https://download.bls.gov/pub/time.series/mp/
## 44 mu              https://download.bls.gov/pub/time.series/mu/
## 45 mw              https://download.bls.gov/pub/time.series/mw/
```

```
## 46 nb      https://download.bls.gov/pub/time.series/nb/
## 47 nc      https://download.bls.gov/pub/time.series/nc/
## 48 nd      https://download.bls.gov/pub/time.series/nd/
## 49 nl      https://download.bls.gov/pub/time.series/nl/
## 50 nw      https://download.bls.gov/pub/time.series/nw/
## 51 oe      https://download.bls.gov/pub/time.series/oe/
## 52 or      https://download.bls.gov/pub/time.series/or/
## 53 overview.txt https://download.bls.gov/pub/time.series/overview~
## 54 pc      https://download.bls.gov/pub/time.series/pc/
## 55 pd      https://download.bls.gov/pub/time.series/pd/
## 56 pr      https://download.bls.gov/pub/time.series/pr/
## 57 sa      https://download.bls.gov/pub/time.series/sa/
## 58 sdmx    https://download.bls.gov/pub/time.series/sdmx/
## 59 sh      https://download.bls.gov/pub/time.series/sh/
## 60 si      https://download.bls.gov/pub/time.series/si/
## 61 sm      https://download.bls.gov/pub/time.series/sm/
## 62 su      https://download.bls.gov/pub/time.series/su/
## 63 tu      https://download.bls.gov/pub/time.series/tu/
## 64 wd      https://download.bls.gov/pub/time.series/wd/
## 65 wm      https://download.bls.gov/pub/time.series/wm/
## 66 wp      https://download.bls.gov/pub/time.series/wp/
## 67 ws      https://download.bls.gov/pub/time.series/ws/
## 68 xx      https://download.bls.gov/pub/time.series/xx/
## 69 yy      https://download.bls.gov/pub/time.series/yy/
```

Elements of the scrapping data

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

```
## # A tibble: 6 x 1
##   X0
##   <fct>
## 1 [To Parent Directory]
## 2 ap
## 3 bd
## 4 bg
## 5 bp
## 6 cc
```

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

```
## 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: 6 x 2
##   X1                                X2
##   <fct>                           <chr>
## 1 "<a href=\"/pub/\""             https://download.bls.gov/pub/
## 2 "<a href=\"/pub/time.series/ap/~ https://download.bls.gov/pub/time.seri~
## 3 "<a href=\"/pub/time.series/bd/~ https://download.bls.gov/pub/time.seri~
## 4 "<a href=\"/pub/time.series/bg/~ https://download.bls.gov/pub/time.seri~
## 5 "<a href=\"/pub/time.series/bp/~ https://download.bls.gov/pub/time.seri~
## 6 "<a href=\"/pub/time.series/cc/~ https://download.bls.gov/pub/time.seri~
```

List all Datasets

```
datasets2 <- datasets %>%
  # Remove Datasets absent
  filter(!X0 %in% c("nl", "xx", "ln", "yy")) %>%
  select(X0, X2) %>%
  as.tibble %>%
  filter(nchar(X0) == 2) %>%
  mutate(link = paste0(X2, X0, ".txt"),
         file = paste0(X0, ".txt")) %>%
  select(link, file)
```

```
datasets2 %>%
  head
```

```
# # A tibble: 6 x 2
#   link                                file
#   <chr>                             <chr>
# 1 https://download.bls.gov/pub/time.series/ap/ap.txt ap.txt
# 2 https://download.bls.gov/pub/time.series/bd/bd.txt bd.txt
# 3 https://download.bls.gov/pub/time.series/bg/bg.txt bg.txt
# 4 https://download.bls.gov/pub/time.series/bp/bp.txt bp.txt
# 5 https://download.bls.gov/pub/time.series/cc/cc.txt cc.txt
# 6 https://download.bls.gov/pub/time.series/cd/cd.txt cd.txt
```

```
datasets2[, "link"]
```

```
# # A tibble: 60 x 1
#   link
#   <chr>
# 1 https://download.bls.gov/pub/time.series/ap/ap.txt
# 2 https://download.bls.gov/pub/time.series/bd/bd.txt
# 3 https://download.bls.gov/pub/time.series/bg/bg.txt
# 4 https://download.bls.gov/pub/time.series/bp/bp.txt
# 5 https://download.bls.gov/pub/time.series/cc/cc.txt
# 6 https://download.bls.gov/pub/time.series/cd/cd.txt
# 7 https://download.bls.gov/pub/time.series/ce/ce.txt
# 8 https://download.bls.gov/pub/time.series/cf/cf.txt
# 9 https://download.bls.gov/pub/time.series/ch/ch.txt
# 10 https://download.bls.gov/pub/time.series/ci/ci.txt
# 11 https://download.bls.gov/pub/time.series/cm/cm.txt
# 12 https://download.bls.gov/pub/time.series/cs/cs.txt
# 13 https://download.bls.gov/pub/time.series/cu/cu.txt
```

```
# 14 https://download.bls.gov/pub/time.series/cw/cw.txt
# 15 https://download.bls.gov/pub/time.series/cx/cx.txt
# 16 https://download.bls.gov/pub/time.series/eb/eb.txt
# 17 https://download.bls.gov/pub/time.series/ec/ec.txt
# 18 https://download.bls.gov/pub/time.series/ee/ee.txt
# 19 https://download.bls.gov/pub/time.series/ei/ei.txt
# 20 https://download.bls.gov/pub/time.series/ep/ep.txt
# 21 https://download.bls.gov/pub/time.series/fi/fi.txt
# 22 https://download.bls.gov/pub/time.series/fm/fm.txt
# 23 https://download.bls.gov/pub/time.series/fw/fw.txt
# 24 https://download.bls.gov/pub/time.series/gg/gg.txt
# 25 https://download.bls.gov/pub/time.series/gp/gp.txt
# 26 https://download.bls.gov/pub/time.series/hc/hc.txt
# 27 https://download.bls.gov/pub/time.series/hs/hs.txt
# 28 https://download.bls.gov/pub/time.series/ii/ii.txt
# 29 https://download.bls.gov/pub/time.series/in/in.txt
# 30 https://download.bls.gov/pub/time.series/ip/ip.txt
# 31 https://download.bls.gov/pub/time.series/is/is.txt
# 32 https://download.bls.gov/pub/time.series/jl/jl.txt
# 33 https://download.bls.gov/pub/time.series/jt/jt.txt
# 34 https://download.bls.gov/pub/time.series/la/la.txt
# 35 https://download.bls.gov/pub/time.series/le/le.txt
# 36 https://download.bls.gov/pub/time.series/li/li.txt
# 37 https://download.bls.gov/pub/time.series/lu/lu.txt
# 38 https://download.bls.gov/pub/time.series/ml/ml.txt
# 39 https://download.bls.gov/pub/time.series/mp/mp.txt
# 40 https://download.bls.gov/pub/time.series/mu/mu.txt
# 41 https://download.bls.gov/pub/time.series/mw/mw.txt
# 42 https://download.bls.gov/pub/time.series/nb/nb.txt
# 43 https://download.bls.gov/pub/time.series/nc/nc.txt
# 44 https://download.bls.gov/pub/time.series/nd/nd.txt
# 45 https://download.bls.gov/pub/time.series/nw/nw.txt
# 46 https://download.bls.gov/pub/time.series/oe/oe.txt
# 47 https://download.bls.gov/pub/time.series/or/or.txt
# 48 https://download.bls.gov/pub/time.series/pc/pc.txt
# 49 https://download.bls.gov/pub/time.series/pd/pd.txt
# 50 https://download.bls.gov/pub/time.series/pr/pr.txt
# 51 https://download.bls.gov/pub/time.series/sa/sa.txt
# 52 https://download.bls.gov/pub/time.series/sh/sh.txt
# 53 https://download.bls.gov/pub/time.series/si/si.txt
# 54 https://download.bls.gov/pub/time.series/sm/sm.txt
# 55 https://download.bls.gov/pub/time.series/su/su.txt
# 56 https://download.bls.gov/pub/time.series/tu/tu.txt
# 57 https://download.bls.gov/pub/time.series/wd/wd.txt
# 58 https://download.bls.gov/pub/time.series/wm/wm.txt
# 59 https://download.bls.gov/pub/time.series/wp/wp.txt
# 60 https://download.bls.gov/pub/time.series/ws/ws.txt
```

Downloading

```
# Map(function(u, d) download.file(u, d, mode = "wb"), datasets2[, "link"],
#   datasets2[, "file"])
```

Computing Environment

```
Sys.time()
```

```
## [1] "2018-09-24 19:13:02 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
##
## 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  curl_3.2        yaml_2.2.0
## [21] haven_1.1.2     withr_2.1.2     httr_1.3.1      knitr_1.20
## [25] hms_0.4.2       rprojroot_1.3-2 grid_3.5.1      tidyselect_0.2.4
## [29] glue_1.3.0      R6_2.2.2        fansi_0.3.0     readxl_1.1.0
## [33] rmarkdown_1.10 selectr_0.4-1    modelr_0.1.2    magrittr_1.5
## [37] backports_1.1.2 scales_1.0.0    htmltools_0.3.6 assertthat_0.2.0
## [41] colorspace_1.3-2 utf8_1.1.4      stringi_1.2.4   lazyeval_0.2.1
## [45] munsell_0.5.0   broom_0.5.0     crayon_1.3.4
```