BLS - SA - Download

Datasets

François Geerolf

Contents

Preamble **1**

SA Presentation **1**

Scrapping html page **1**
Scrapping database names 1
Scrapping database urls using regular expressions 2
Merging both in a Dataset 2

Downloading all data **3**

Computing Environment **5**

Preamble

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

SA Presentation

SA = State and Area Employment, Hours and Earnings

The flat data files of the SA database are here: https://download.bls.gov/pub/time.series/sa/

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

Scrapping html page

Scraping: "https://download.bls.gov/pub/time.series/sa/"

Scrapping database names

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

```
  rename(X0 = ".") %>%
  as.tibble
```

```
# # A tibble: 120 x 1
#    X0
#    <fct>
#  1 [To Parent Directory]
#  2 .message
#  3 sa.area
#  4 sa.contacts
#  5 sa.data.0.Current
#  6 sa.data.10a.Florida
#  7 sa.data.10b.Florida
#  8 sa.data.10c.Florida
#  9 sa.data.11a.Georgia
# 10 sa.data.11b.Georgia
# # ... with 110 more rows
```

Scrapping database urls using regular expressions

```
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: 120 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/sa/.~ https://download.bls.gov/pub/time.se~
##  3 "<a href=\"/pub/time.series/sa/s~ https://download.bls.gov/pub/time.se~
##  4 "<a href=\"/pub/time.series/sa/s~ https://download.bls.gov/pub/time.se~
##  5 "<a href=\"/pub/time.series/sa/s~ https://download.bls.gov/pub/time.se~
##  6 "<a href=\"/pub/time.series/sa/s~ https://download.bls.gov/pub/time.se~
##  7 "<a href=\"/pub/time.series/sa/s~ https://download.bls.gov/pub/time.se~
##  8 "<a href=\"/pub/time.series/sa/s~ https://download.bls.gov/pub/time.se~
##  9 "<a href=\"/pub/time.series/sa/s~ https://download.bls.gov/pub/time.se~
## 10 "<a href=\"/pub/time.series/sa/s~ https://download.bls.gov/pub/time.se~
## # ... with 110 more rows
```

Merging both in a Dataset

```
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: 120 x 3
#    X0            X1                        X2
#    <chr>         <chr>                     <chr>
#  1 [To Parent D~ "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
#  2 .message      "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
#  3 sa.area       "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
#  4 sa.contacts   "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
#  5 sa.data.0.Cu~ "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
#  6 sa.data.10a.~ "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
#  7 sa.data.10b.~ "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
#  8 sa.data.10c.~ "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
#  9 sa.data.11a.~ "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
# 10 sa.data.11b.~ "<a href=\"/pub/time.seri~ https://download.bls.gov/pub/~
# # ... with 110 more rows
```

Downloading all data

```
for (i in 3:119){
  file <- datasets[i, "X0"]
  cat("\nDownloading from BLS Website LA:", file)
  assign(file, read.csv(datasets[i, "X2"], sep = "\t", row.names = NULL))
  do.call(save, list(file, file = paste0(file, ".RData")))
}
```

```
#
# Downloading from BLS Website LA: sa.area
# Downloading from BLS Website LA: sa.contacts
# Downloading from BLS Website LA: sa.data.0.Current
# Downloading from BLS Website LA: sa.data.10a.Florida
# Downloading from BLS Website LA: sa.data.10b.Florida
# Downloading from BLS Website LA: sa.data.10c.Florida
# Downloading from BLS Website LA: sa.data.11a.Georgia
# Downloading from BLS Website LA: sa.data.11b.Georgia
# Downloading from BLS Website LA: sa.data.12.Hawaii
# Downloading from BLS Website LA: sa.data.13.Idaho
# Downloading from BLS Website LA: sa.data.14a.Illinois
# Downloading from BLS Website LA: sa.data.14b.Illinois
# Downloading from BLS Website LA: sa.data.14c.Illinois
# Downloading from BLS Website LA: sa.data.15a.Indiana
# Downloading from BLS Website LA: sa.data.15b.Indiana
# Downloading from BLS Website LA: sa.data.16.Iowa
# Downloading from BLS Website LA: sa.data.17.Kansas
```

```
# Downloading from BLS Website LA: sa.data.18.Kentucky
# Downloading from BLS Website LA: sa.data.19a.Louisiana
# Downloading from BLS Website LA: sa.data.19b.Louisiana
# Downloading from BLS Website LA: sa.data.1a.Alabama
# Downloading from BLS Website LA: sa.data.1b.Alabama
# Downloading from BLS Website LA: sa.data.2.Alaska
# Downloading from BLS Website LA: sa.data.20.Maine
# Downloading from BLS Website LA: sa.data.21.Maryland
# Downloading from BLS Website LA: sa.data.22a.Massachusetts
# Downloading from BLS Website LA: sa.data.22b.Massachusetts
# Downloading from BLS Website LA: sa.data.23a.Michigan
# Downloading from BLS Website LA: sa.data.23b.Michigan
# Downloading from BLS Website LA: sa.data.23c.Michigan
# Downloading from BLS Website LA: sa.data.23d.Michigan
# Downloading from BLS Website LA: sa.data.24.Minnesota
# Downloading from BLS Website LA: sa.data.25.Mississippi
# Downloading from BLS Website LA: sa.data.26a.Missouri
# Downloading from BLS Website LA: sa.data.26b.Missouri
# Downloading from BLS Website LA: sa.data.27.Montana
# Downloading from BLS Website LA: sa.data.28.Nebraska
# Downloading from BLS Website LA: sa.data.29.Nevada
# Downloading from BLS Website LA: sa.data.3.Arizona
# Downloading from BLS Website LA: sa.data.30.NewHampshire
# Downloading from BLS Website LA: sa.data.31a.NewJersey
# Downloading from BLS Website LA: sa.data.31b.NewJersey
# Downloading from BLS Website LA: sa.data.31c.NewJersey
# Downloading from BLS Website LA: sa.data.32.NewMexico
# Downloading from BLS Website LA: sa.data.33a.NewYork
# Downloading from BLS Website LA: sa.data.33b.NewYork
# Downloading from BLS Website LA: sa.data.33c.NewYork
# Downloading from BLS Website LA: sa.data.33d.NewYork
# Downloading from BLS Website LA: sa.data.33e.NewYork
# Downloading from BLS Website LA: sa.data.33f.NewYork
# Downloading from BLS Website LA: sa.data.33g.NewYork
# Downloading from BLS Website LA: sa.data.33h.NewYork
# Downloading from BLS Website LA: sa.data.33i.NewYork
# Downloading from BLS Website LA: sa.data.34a.NorthCarolina
# Downloading from BLS Website LA: sa.data.34b.NorthCarolina
# Downloading from BLS Website LA: sa.data.35.NorthDakota
# Downloading from BLS Website LA: sa.data.36a.Ohio
# Downloading from BLS Website LA: sa.data.36b.Ohio
# Downloading from BLS Website LA: sa.data.36c.Ohio
# Downloading from BLS Website LA: sa.data.36d.Ohio
# Downloading from BLS Website LA: sa.data.36e.Ohio
# Downloading from BLS Website LA: sa.data.36f.Ohio
# Downloading from BLS Website LA: sa.data.37.Oklahoma
# Downloading from BLS Website LA: sa.data.38a.Oregon
# Downloading from BLS Website LA: sa.data.38b.Oregon
# Downloading from BLS Website LA: sa.data.39a.Pennsylvania
# Downloading from BLS Website LA: sa.data.39b.Pennsylvania
# Downloading from BLS Website LA: sa.data.39c.Pennsylvania
# Downloading from BLS Website LA: sa.data.39d.Pennsylvania
# Downloading from BLS Website LA: sa.data.39e.Pennsylvania
# Downloading from BLS Website LA: sa.data.39f.Pennsylvania
```

```
# Downloading from BLS Website LA: sa.data.39g.Pennsylvania
# Downloading from BLS Website LA: sa.data.40.PuertoRico
# Downloading from BLS Website LA: sa.data.41.RhodeIsland
# Downloading from BLS Website LA: sa.data.42a.SouthCarolina
# Downloading from BLS Website LA: sa.data.42b.SouthCarolina
# Downloading from BLS Website LA: sa.data.43.SouthDakota
# Downloading from BLS Website LA: sa.data.44a.Tennessee
# Downloading from BLS Website LA: sa.data.44b.Tennessee
# Downloading from BLS Website LA: sa.data.45a.Texas
# Downloading from BLS Website LA: sa.data.45b.Texas
# Downloading from BLS Website LA: sa.data.45c.Texas
# Downloading from BLS Website LA: sa.data.45d.Texas
# Downloading from BLS Website LA: sa.data.45e.Texas
# Downloading from BLS Website LA: sa.data.46.Utah
# Downloading from BLS Website LA: sa.data.47.Vermont
# Downloading from BLS Website LA: sa.data.48a.Virginia
# Downloading from BLS Website LA: sa.data.48b.Virginia
# Downloading from BLS Website LA: sa.data.48c.Virginia
# Downloading from BLS Website LA: sa.data.49.VirginIslands
# Downloading from BLS Website LA: sa.data.4a.Arkansas
# Downloading from BLS Website LA: sa.data.4b.Arkansas
# Downloading from BLS Website LA: sa.data.50.Washington
# Downloading from BLS Website LA: sa.data.51a.WestVirginia
# Downloading from BLS Website LA: sa.data.51b.WestVirginia
# Downloading from BLS Website LA: sa.data.52a.Wisconsin
# Downloading from BLS Website LA: sa.data.52b.Wisconsin
# Downloading from BLS Website LA: sa.data.53.Wyoming
# Downloading from BLS Website LA: sa.data.5a.California
# Downloading from BLS Website LA: sa.data.5b.California
# Downloading from BLS Website LA: sa.data.5c.California
# Downloading from BLS Website LA: sa.data.5d.California
# Downloading from BLS Website LA: sa.data.5e.California
# Downloading from BLS Website LA: sa.data.5f.California
# Downloading from BLS Website LA: sa.data.5g.California
# Downloading from BLS Website LA: sa.data.6.Colorado
# Downloading from BLS Website LA: sa.data.7a.Connecticut
# Downloading from BLS Website LA: sa.data.7b.Connecticut
# Downloading from BLS Website LA: sa.data.8.Delaware
# Downloading from BLS Website LA: sa.data.9.DC
# Downloading from BLS Website LA: sa.data_type
# Downloading from BLS Website LA: sa.detail
# Downloading from BLS Website LA: sa.footnote
# Downloading from BLS Website LA: sa.industry
# Downloading from BLS Website LA: sa.period
# Downloading from BLS Website LA: sa.series
# Downloading from BLS Website LA: sa.state
```

Computing Environment

```
Sys.time()
```

```
## [1] "2018-09-24 20:19:37 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
```