BLS - LAUS - Download

Datasets

François Geerolf

Contents

Preamble

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

LAUS Presentation

The **Local Area Unemployment Statistics (LAUS)**:
- Monthly / annual employment, unemployment, labor force
- Census regions and divisions, States, counties, metropolitan areas, and many cities, by place of residence.

The documentation for the LAUS is available here: https://www.bls.gov/lau/
Databases for LAUS are here: https://www.bls.gov/lau/data.htm
The flat data files of the LAUS are here: https://download.bls.gov/pub/time.series/la/

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

Scrapping html page

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

Scrapping database names

```
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 la.area
# 3 la.area_type
# 4 la.areamaps
# 5 la.contacts
# 6 la.data.0.CurrentU00-04
```

Scrapping database urls using regular expressions

Regular expressions are a little bit hard getting used to. Here are useful cheap sheets:
- http://web.mit.edu/hackl/www/lab/turkshop/slides/regex-cheatsheet.pdf - https://www.dataquest.io/blog/large_files/python-regular-expressions-cheat-sheet.pdf

```
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/time.series/\""   https://download.bls.gov/pub/time.ser~
## 2 "<a href=\"/pub/time.series/la/l~ https://download.bls.gov/pub/time.ser~
## 3 "<a href=\"/pub/time.series/la/l~ https://download.bls.gov/pub/time.ser~
## 4 "<a href=\"/pub/time.series/la/l~ https://download.bls.gov/pub/time.ser~
## 5 "<a href=\"/pub/time.series/la/l~ https://download.bls.gov/pub/time.ser~
## 6 "<a href=\"/pub/time.series/la/l~ https://download.bls.gov/pub/time.ser~
```

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 %>%
  head
```

```
# # A tibble: 6 x 3
#   X0            X1                          X2
#   <chr>         <chr>                       <chr>
# 1 [To Parent D~ "<a href=\"/pub/time.serie~ https://download.bls.gov/pub/~
# 2 la.area       "<a href=\"/pub/time.serie~ https://download.bls.gov/pub/~
# 3 la.area_type  "<a href=\"/pub/time.serie~ https://download.bls.gov/pub/~
# 4 la.areamaps   "<a href=\"/pub/time.serie~ https://download.bls.gov/pub/~
# 5 la.contacts   "<a href=\"/pub/time.serie~ https://download.bls.gov/pub/~
# 6 la.data.0.Cu~ "<a href=\"/pub/time.serie~ https://download.bls.gov/pub/~
```

Downloading all data

```
for (i in 2:80){
  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: la.area
# Downloading from BLS Website LA: la.area_type
# Downloading from BLS Website LA: la.areamaps
# Downloading from BLS Website LA: la.contacts
# Downloading from BLS Website LA: la.data.0.CurrentU00-04
# Downloading from BLS Website LA: la.data.0.CurrentU05-09
# Downloading from BLS Website LA: la.data.0.CurrentU10-14
# Downloading from BLS Website LA: la.data.0.CurrentU15-19
# Downloading from BLS Website LA: la.data.0.CurrentU90-94
# Downloading from BLS Website LA: la.data.0.CurrentU95-99
# Downloading from BLS Website LA: la.data.1.CurrentS
# Downloading from BLS Website LA: la.data.10.Arkansas
# Downloading from BLS Website LA: la.data.11.California
# Downloading from BLS Website LA: la.data.12.Colorado
# Downloading from BLS Website LA: la.data.13.Connecticut
# Downloading from BLS Website LA: la.data.14.Delaware
# Downloading from BLS Website LA: la.data.15.DC
# Downloading from BLS Website LA: la.data.16.Florida
# Downloading from BLS Website LA: la.data.17.Georgia
# Downloading from BLS Website LA: la.data.18.Hawaii
# Downloading from BLS Website LA: la.data.19.Idaho
# Downloading from BLS Website LA: la.data.2.AllStatesU
```

```
# Downloading from BLS Website LA: la.data.20.Illinois
# Downloading from BLS Website LA: la.data.21.Indiana
# Downloading from BLS Website LA: la.data.22.Iowa
# Downloading from BLS Website LA: la.data.23.Kansas
# Downloading from BLS Website LA: la.data.24.Kentucky
# Downloading from BLS Website LA: la.data.25.Louisiana
# Downloading from BLS Website LA: la.data.26.Maine
# Downloading from BLS Website LA: la.data.27.Maryland
# Downloading from BLS Website LA: la.data.28.Massachusetts
# Downloading from BLS Website LA: la.data.29.Michigan
# Downloading from BLS Website LA: la.data.3.AllStatesS
# Downloading from BLS Website LA: la.data.30.Minnesota
# Downloading from BLS Website LA: la.data.31.Mississippi
# Downloading from BLS Website LA: la.data.32.Missouri
# Downloading from BLS Website LA: la.data.33.Montana
# Downloading from BLS Website LA: la.data.34.Nebraska
# Downloading from BLS Website LA: la.data.35.Nevada
# Downloading from BLS Website LA: la.data.36.NewHampshire
# Downloading from BLS Website LA: la.data.37.NewJersey
# Downloading from BLS Website LA: la.data.38.NewMexico
# Downloading from BLS Website LA: la.data.39.NewYork
# Downloading from BLS Website LA: la.data.4.RegionDivisionU
# Downloading from BLS Website LA: la.data.40.NorthCarolina
# Downloading from BLS Website LA: la.data.41.NorthDakota
# Downloading from BLS Website LA: la.data.42.Ohio
# Downloading from BLS Website LA: la.data.43.Oklahoma
# Downloading from BLS Website LA: la.data.44.Oregon
# Downloading from BLS Website LA: la.data.45.Pennsylvania
# Downloading from BLS Website LA: la.data.46.PuertoRico
# Downloading from BLS Website LA: la.data.47.RhodeIsland
# Downloading from BLS Website LA: la.data.48.SouthCarolina
# Downloading from BLS Website LA: la.data.49.SouthDakota
# Downloading from BLS Website LA: la.data.5.RegionDivisionS
# Downloading from BLS Website LA: la.data.50.Tennessee
# Downloading from BLS Website LA: la.data.51.Texas
# Downloading from BLS Website LA: la.data.52.Utah
# Downloading from BLS Website LA: la.data.53.Vermont
# Downloading from BLS Website LA: la.data.54.Virginia
# Downloading from BLS Website LA: la.data.56.Washington
# Downloading from BLS Website LA: la.data.57.WestVirginia
# Downloading from BLS Website LA: la.data.58.Wisconsin
# Downloading from BLS Website LA: la.data.59.Wyoming
# Downloading from BLS Website LA: la.data.60.Metro
# Downloading from BLS Website LA: la.data.61.Division
# Downloading from BLS Website LA: la.data.62.Micro
# Downloading from BLS Website LA: la.data.63.Combined
# Downloading from BLS Website LA: la.data.64.County
# Downloading from BLS Website LA: la.data.65.City
# Downloading from BLS Website LA: la.data.7.Alabama
# Downloading from BLS Website LA: la.data.8.Alaska
# Downloading from BLS Website LA: la.data.9.Arizona
# Downloading from BLS Website LA: la.footnote
# Downloading from BLS Website LA: la.map_info
# Downloading from BLS Website LA: la.measure
```

```
# Downloading from BLS Website LA: la.period
# Downloading from BLS Website LA: la.series
# Downloading from BLS Website LA: la.state_region_division
```

Computing Environment

```
Sys.time()
```

```
## [1] "2018-09-24 07:39:18 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
```