

Census - STC - Download

Datasets

François Geerolf

Contents

Preamble	1
STC Presentation	1
Downloading	1
Computing Environment	2

Preamble

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

STC Presentation

Annual Survey of State Government Tax Collections (STC):

<https://www.census.gov/programs-surveys/stc/data/datasets.html>

Annual Survey of State Government Finances:

<https://www.census.gov/programs-surveys/state/data/datasets.All.html>

Downloading

```
"https://www2.census.gov/programs-surveys/stc/datasets/historical/state_tax_collections.zip" %>%
  curl_download(., destfile = "state_tax_collections.zip", quiet = FALSE)

unzip("state_tax_collections.zip")
unlink("state_tax_collections.zip")

census.stc <- read.xls("STC_Historical_DB (2017).xls") %>%
  filter(!is.na(Year)) %>%
  gather(variable, value, -Year, -State, -Name, -FY.Ending.Date) %>%
  mutate(value = gsub(",", "", value),
         value = as.numeric(value)) %>%
  mutate_at(vars(Year, State, Name, FY.Ending.Date), funs(paste(.))) %>%
  mutate_at(vars(Year, State, Name, FY.Ending.Date), funs(factor(.))) %>%
  full_join(read.csv("variables/variable.list.csv", header = FALSE) %>%
            select(variable = V1, description = V2), by = "variable") %>%
```

```
mutate(year = as.numeric(paste(Year)),
       state = substr(paste(Name), 1, 2),
       variable = factor(variable),
       state = factor(state)) %>%
select(-Year, -Name, -State, -FY.Ending.Date) %>%
select(state, year, variable, description, value) %>%
arrange(variable, state, year)
```

```
## Warning: attributes are not identical across measure variables;
## they will be dropped

## Warning in evalq(as.numeric(value), <environment>): NAs introduced by
## coercion

## Warning: Column `variable` joining character vector and factor, coercing
## into character vector

unlink("STC_Historical_DB (2017).xls")

save(census.stc, file = "census.stc.RData")
```

Computing Environment

```
Sys.time()
```

```
## [1] "2018-09-25 14:06:47 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  gdata_2.18.0   forcats_0.3.0  stringr_1.3.1
## [5] dplyr_0.7.6     purrr_0.2.5    readr_1.1.1    tidyr_0.8.1
## [9] tibble_1.4.2    ggplot2_3.0.0  tidyverse_1.2.1 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      xml2_1.2.0      httr_1.3.1      knitr_1.20
## [25] gtools_3.8.1     hms_0.4.2       rprojroot_1.3-2 grid_3.5.1
## [29] tidyselect_0.2.4 glue_1.3.0      R6_2.2.2        readxl_1.1.0
## [33] rmarkdown_1.10   modelr_0.1.2    magrittr_1.5     backports_1.1.2
## [37] scales_1.0.0     htmltools_0.3.6 rvest_0.3.2      assertthat_0.2.0
## [41] colorspace_1.3-2 stringi_1.2.4    lazyeval_0.2.1   munsell_0.5.0
## [45] broom_0.5.0      crayon_1.3.4
```