

Programs

true

2019-08-09

Program directory

This directory contains all programs necessary to run the cleaning, analysis, etc. They can be run separately, or in a single run.

Setup

Most parameters are set in the `config.R`:

```
source(file.path(rprojroot::find_rstudio_root_file(), "pathconfig.R"), echo=TRUE)
```

```
##
## > basepath <- rprojroot::find_rstudio_root_file()
##
## > acquired <- file.path(basepath, "data", "acquired")
##
## > interwrk <- file.path(basepath, "data", "interwrk")
##
## > generated <- file.path(basepath, "data", "generated")
##
## > outputs <- file.path(basepath, "analysis")
##
## > programs <- file.path(basepath, "programs")
##
## > for (dir in list(acquired, interwrk, generated, outputs)) {
## +   if (file.exists(dir)) {
## +     }
## +   else {
## +     dir.create(file.path(dir))
## +   }
## +   [TRUNCATED]

```

```
source(file.path(programs, "config.R"), echo=TRUE)
```

```
##
## > source(file.path(programs, "global-libraries.R"),
## +   echo = FALSE)
## Loading required package: dplyr
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
## Loading required package: devtools
```

```
## Loading required package: rprojroot
## Loading required package: tictoc
##
## > source(file.path(programs, "libraries.R"), echo = FALSE)
## Loading required package: tidyr
## Loading required package: readxl
## Loading required package: fuzzyjoin
## Loading required package: knitr
##
## > onet.src.base <- "https://www.onetcenter.org/dl_files/database/"
##
## > onet.src.version <- "23_2"
##
## > onet.src.file <- paste("db", onet.src.version, "excel.zip",
## +      sep = "_")
##
## > oes.src.base <- "https://www.bls.gov/oes/special.requests/"
##
## > oes.src.version <- "2018"
##
## > oes.src.file <- paste("oesm", oes.src.version, "nat.zip",
## +      sep = "")
```

Note that the path `interwrk` is transitory, and is only kept during processing. It will be empty in the replication archive.

Download and unpack the O*NET data

- Input data: O*NET
- Output data: path 'acquired'

```
source(file.path(programs, "01_download_onet.R"), echo=TRUE)
```

Mapping job titles to SOC

We also merge in BLS data on salaries.

```
##
## > source(file.path(rprojroot::find_rstudio_root_file(),
## +      "pathconfig.R"), echo = FALSE)
##
## > source(file.path(programs, "config.R"), echo = FALSE)
##
## > Alternate_Titles <- read_excel(file.path(acquired,
## +      "Alternate Titles.xlsx"))
##
## > Occupation_Data <- read_excel(file.path(acquired,
## +      "Occupation Data.xlsx")) %>% select(-Title)
##
## > BLS.data <- read_excel(file.path(acquired, paste0("national_M",
## +      oes.src.version, "_dl.xlsx")))
##
```

```
## > job_titles <- read_excel(file.path(generated, "job_titles.xlsx"))
##
## > names(job_titles) <- c("Job Title")
##
## > soc_job_titles <- Alternate_Titles %>% select("O*NET-SOC Code",
## +      Title) %>% distinct()
##
## > soc_job_alttitles <- Alternate_Titles %>% select("O*NET-SOC Code",
## +      "Alternate Title") %>% distinct()
##
## > primary <- stringdist_inner_join(y = soc_job_titles,
## +      x = job_titles, by = c(`Job Title` = "Title"), method = "jw",
## +      distance_col = "jw_ ..." ... [TRUNCATED]
##
## > secondary <- stringdist_inner_join(y = soc_job_alttitles,
## +      x = job_titles, by = c(`Job Title` = "Alternate Title"),
## +      method = "jw", dist .... [TRUNCATED]
##
## > nlm.titles <- bind_rows(primary, secondary) %>% left_join(Occupation_Data,
## +      by = "O*NET-SOC Code") %>% separate("O*NET-SOC Code", sep = 7,
## +      .... [TRUNCATED]
##
## > saverDS(nlm.titles, file = file.path(outputs, "nlm.titles.RDS"))
##
## > write.csv(nlm.titles, file = file.path(outputs, "nlm.titles.csv"))
```

Print out the table

```
kable(nlm.titles %>% select("Job Title", "Title", "Alternate Title", "A_PCT25", "A_MEDIAN", "A_PCT75"))
```

Job Title	Title	Alternate Title	A_PCT
Archivists	Archivists	NA	38090
Archivists	Archivists	Archivist	38090
Curators	Curators	NA	39580
Curators	Archeologists	Curator	48020
Curators	Archivists	Curator	38090
Curators	Curators	Curator	39580
Data Librarians	NA	NA	NA
Scientists	Biofuels/Biodiesel Technology and Product Development Managers	Scientist	112400
Scientists	Mathematicians	Scientist	73490
Scientists	Chemical Engineers	Scientist	81900
Scientists	Nuclear Engineers	Scientist	85840
Scientists	Nanosystems Engineers	Scientist	69890
Scientists	Manufacturing Engineering Technologists	Scientist	47500
Scientists	Nanotechnology Engineering Technologists	Scientist	47500
Scientists	Biologists	Scientist	56730
Scientists	Biochemists and Biophysicists	Scientist	64230
Scientists	Bioinformatics Scientists	Scientist	60250
Scientists	Geneticists	Scientist	60250
Scientists	Medical Scientists, Except Epidemiologists	Scientist	59580
Scientists	Astronomers	Scientist	74300
Scientists	Physicists	Scientist	85090
Scientists	Chemists	Scientist	56290

Job Title	Title	Alternate Title	A_PCT
Scientists	Climate Change Analysts	Scientist	53580
Scientists	Hydrologists	Scientist	61280
Scientists	Remote Sensing Scientists and Technologists	Scientist	75830
Scientists	Anthropologists	Scientist	48020
Scientists	Geographers	Scientist	63270
Policy Specialists	NA	NA	NA
Project Managers	Construction Managers	Project Manager	70670
Project Managers	Architectural and Engineering Managers	Project Manager	112400
Project Managers	Managers, All Other	Project Manager	75460
Project Managers	Wind Energy Project Managers	Project Manager	75460
Project Managers	Information Technology Project Managers	Project Manager	66410
Project Managers	Environmental Engineers	Project Manager	66590
Project Managers	Water/Wastewater Engineers	Project Manager	66590
Project Managers	Wind Energy Engineers	Project Manager	69890
Project Managers	Architectural Drafters	Project Manager	43660
Project Managers	Environmental Restoration Planners	Project Manager	53580
Project Managers	Social Science Research Assistants	Project Manager	35450
Project Managers	Remote Sensing Technicians	Project Manager	37940
Project Managers	Technical Directors/Managers	Project Manager	48520
Project Managers	Intelligence Analysts	Project Manager	57560
Project Managers	First-Line Supervisors of Construction Trades and Extraction Workers	Project Manager	52380
Researchers	Industrial Ecologists	Researcher	53580
Researchers	Anthropologists	Researcher	48020
Researchers	Historians	Researcher	40670
Senior Staffs	NA	NA	NA
Software Engineers	Computer and Information Research Scientists	Software Engineer	91650
Software Engineers	Software Developers, Applications	Software Engineer	79340
Software Engineers	Software Developers, Systems Software	Software Engineer	85610