General Design Document

1. Summary

All folders included here consists of codes useful for downloading and/or cleaning data for professor Simon Mongey’s research project on dental association in the United States.

**There are no datasets, except papers and notes, in the folder. They are downloaded with R scripts.**

Data would be available in Stata DTA format and CSV format.

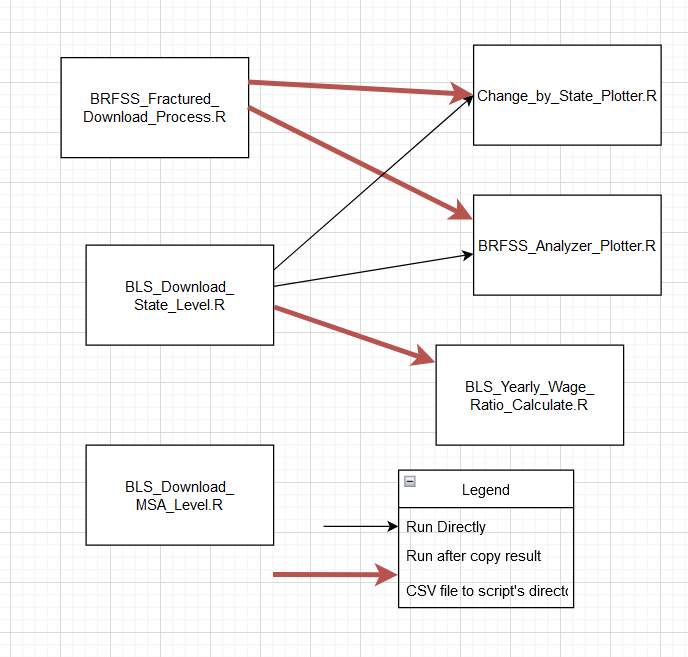
The documentation is organized by dataset. While you can run scripts in each folder, I advise first looking into flowchart in part (3)

1. Included Datasets

|  |  |  |
| --- | --- | --- |
| Data Source | Found under folder: | Name |
| Bureau of Labor Statistics | BLS\_Data | Wages\_by\_State.csv  Employment\_by\_State.csv  State\_Wage\_Ratio.csv |
| American Dental Education Association | ADEA\_Data | ADEA\_Admission\_09\_21.csv  ADEA\_GPA\_DAT\_00\_21.csv |
| Behavioral Risk Factor Surveillance System | BRFSS\_Data/BRFSS\_Data\_Processed\_Folder | Processed\_BRFSS\_1999\_to\_2020.csv |
| National Health and Nutrition Examination Survey | NHANES\_Data | NHANES\_Result.csv |
| Graphic Reproduction Instructions | Graphic\_Reproduction\_Folder | All RMD, PDF, and TEX file under the folder. |
| Other Research Documents | Xianbin Paper and Notes | Notes.docx  Oral Data Notes.docx  Dental Hygienist to Dentist Ratio By State.docx |
| United States Population estimation data | NOT present in any folder. Downloaded within script upon use. | N/A |

1. Order of Script Execution:

**ADEA Data and NHANES Data are independent from other dataset.**



1. Bureau of Labor Statistics Data

The Bureau of Labor Statistics Data (BLS) Data contains employment information.

Since the raw data is numerous, the downloading of raw data is done within R script.

Available Variables: Wages, By year/hour, Mean and Median

Available By: State-Year-Occupation (Selected)

Availability Range: 50 US States, D.C., and Puerto Rico. Year 1999-2017, Occupation as listed bwloe.

In our script, we included only the following occupation:

|  |  |  |
| --- | --- | --- |
| BLS Code | Occupation Title | Calculated |
| 29-1021 | Dentists, General |  |
| 29-1020 | Dentists, All | X |
| 29-2021 | Dental Hygienists |  |
| 29-1062 | Family and General Practitioners |  |
| 29-1069 | Physicians and Surgeons, All Other |  |
| 29-1060 | Physicians, All | X |
| 29-1111 | Register Nurse (Code used before 2012) |  |
| 29-1141 | Register Nurse (Code used after 2012) |  |

By having “X” in “Calculated” column, it means that the Mean Annual Wages and Total Employment was calculated from its sub-categories.

Order of Script Execution:

* + 1. Run BLS\_Download\_State\_Level.R
    2. Run BLS\_Result\_Plot.R or BLS\_Yearly\_Wage\_Calculate.R

Usage of Sctipts:

* BLS\_Download\_State\_Level.R: Download BLS Data and write in CSV and DTA
* BLS\_Result\_Plot.R: Create Scatterplots of variables calculated from BLS Data
* BLS\_Yearly\_Wage\_Calculator.R: Calculate ratio of wages between different occupations.
* BLS\_Download\_MSA\_Level: This script was **No Longer Used.** It downloads BLS data but in Metropolitan Statistical Area level. However, changing MSA code and lack of yearly population data made it less useful. Still, it can be executed and used if you wish so.

For More Specific Technical Details, See **readme.txt under BLS\_Data folder**, and **comments in R script.**

1. ADEA Data

The American Dental Education Association Data provides information on enrollment and applications to U.S. dental schools. It helps assess the amount of new dental students by year and their academic abilities.

There are actually two dataset:

a) ADEA\_Admission\_09\_21.csv

Available Variables: Admission and Acceptance, in-state and all over the United States.

Available By: Year-School

Availability Range: All US dental school listed in ADEA (About 60), year 2009 to 2021

b) ADEA\_GPA\_DAT\_00\_21.csv

Available Variables: GPA and DAT(Dental Admission Test), for enrolled and applicants.

Available By: Year. National Level

Availability Range: National Level Average, Year 2000 to 2021

Order of Script Execution:

* + - * 1. Make sure you have file “ADEA Dental School Applicants and Enrollees x Entering Class.xlsx” in the directory, where x ranges from 2010 to 2021. These file was downloaded and stored in the directory since they are not large file and does not have to be downloaded in-script. They can be found at [https://www.adea.org/data/students/#collapse0](https://www.adea.org/data/students/" \l "collapse0).
        2. Make sure file “ADEA\_2009\_Applicant\_By\_State.xls” and “ADEA\_2009\_Enroll\_By\_State.xls” are present in the directory. The tables are found at <https://www.adea.org/publications/pages/2009-applicants-and-enrollees.aspx>, table 7 and 8 respectively.
        3. Run ADEA\_Admission\_Data\_Compiler.R and/or ADEA\_GPA\_Data\_Compiler.R

Both scripts reads the raw data and compile them into usable tables.

See technical notes in **Comments in each R Script**.

1. BRFSS Data

The Behavioral Risk Factor Surveillance System (BRFSS) Data was provided by CDC and records the health conditions, behavior, and other health-related issues of surveyed Americans. It is useful to see changes in oral health.

Since the raw data is huge, the downloading of raw data is done within R script.

The output, if you run BRFSS\_Fractured\_Download\_Process.R without changing its configurations, shall be **Processed\_BRFSS\_1999\_to\_2020.csv.** It is available in Individual Response level, with year of response and state of residence recorded. For its Availability, Variables Included, and their usage: See **BRFSS\_Data\_Downloaded\_Codebook.xlsx**. Sheet 1 for variables included with their meanings, sheet two for availability.

After the data mentioned above is available, run BRFSS\_Analyzer\_Plotter.R to make plots and visualize the data in state-year level.

Order of Script Execution:

* + 1. Make sure you copied result from BLS Script, namely **Employment\_By\_State.csv** and **Wages\_By\_State.csv**, under the directory of BRFSS\_Data
    2. Run BRFSS\_Fractured\_Download\_Process.R
    3. Run BRFSS\_Analyzer\_Plotter.R

Usage of Sctipts:

* BRFSS\_Fractured\_Download\_Process.R: Download BRFSS Data and write them in CSV and/or DTA format. Configurations Available upon changing and running the original script.
* BRFSS\_Analyzer\_Plotter.R: Calculate some variables and make plots, by state/year level.
* Change\_by\_State\_Plotter.R: Make plots indicating the change of several variables over years, for each state. Useful for both BLS and BRFSS data.

For More Specific Technical Details, See **Comments in each R script.**

1. NHANES Data

The National Health and Nutrition Examination Survey is a CDC dataset on various health-related questions, including behavioral questions and examination data. It is useful to assess yearly change in oral health condition of Americans.

Since the raw data is huge, the downloading of raw data is done within R script.

Available By: Individual Entries, Averaging only possible in National Level

Availability and Codebook: See **NHANES\_Data\_Downloaded\_Codebook.xlsx**. Sheet 1 for variables included with their meanings, sheet two for availability.

Order of Script Execution: Run NHANES\_Download\_Process.R

There is NO separate analyzing/plotting script for NHANES data. However, near the end of the R script, there are a few lines on how to plot with the processed NHANES data.

For More Specific Technical Details, See **Comments in each R script.**

1. Graphic Reproduction

This part are several RMD, PDF, and TEX file on how to replicate the graphics I produced over the summer. Most of the descriptions are easy to read in PDF files, though R Markdown files are preferred to replicate codes.

1. Other Research Documents

a) Notes.docx

These included notes on a series of papers and reports regarding dentistry, dental association, and dental health problems in the United States. For each entry, there is a short summary and link to the original file. It should be useful as beginning point as literature review.

b)Oral Data Notes.docx

This file recorded several sources from which we can learn about oral health condition of Americans, as suggested in past literature and from my own research.We used BRFSS mainly and NHANES.

c) Dental Hygienist to Dentist Ratio By State.docx

This file consists a record of policies of 24 US States and their regulations on dental hygienist each dentist may supervise and their historical changes. Earlier pages included also scope of practice and their changes, while later one had only dental hygienist to dentist ratio and code numbers where scope of practice was found.

Many states have not and have never had such limitations.

Due to limited time, I did only 24 States.

1. United States Population Data

BLS Data was combined with estimation of US populations by state-year to calculate dentist-per-1,000-people and other variables. They were not downloaded in a separate script, but usually done in a few lines within other scripts. See comments in BLS\_Download\_State\_Level.R for additional information.