

2017 PSS SUMMER SCHOOL

Assignment 0

Due: 12:30pm on June 21

1 Introduction

- Eduard Danalache
- Rice University
- Business Analyst Intern, OCDO - DMID
- Justin Rinker, Manager

2 Section Work

I am working with the DMID section in the OCDO to streamline the data pipeline that we use at the Board. My two main projects are:

- converting existing legacy data-update/revision processes to a modern method that sources all of the data from the same place (rather than going to individual vendors' websites)
- designing a new, more flexible, data flow process that has more GUI-based elements and can receive data of any type (not just time series) from any source so that any member of the Board could use it, rather than just those specifically trained in it

I'd say my work is more policy-based, although not monetary policy but instead internal Board data-management and governance policy. In the past I've worked on a research project that involved performing a K-Means clustering analysis of N=411 mild traumatic brain injury (mTBI) patients to help detect and improve identification of conditions such as PTSD. After running a K-Means clustering algorithm on the data I performed a thorough statistical analysis of the significance of the groups using IBM's SPSS software.

3 Goals

I am taking this course to get more exposure to research and to learn how to use R for data exploration and analysis. I am a member of my school's Kaggle Club, and a lot of the data exploration that Kaggle does is through R, and although I currently use Python for this purpose I think R would be useful to learn as well, and it seems to me that it would be more useful for data exploration as well.

4 R Functions

IN AN ITEMIZED LIST (HINT: GOOGLE HOW TO DO THIS) DESCRIBE WHAT THE FUNCTIONS `setwd()`, `ls()`, and `str()` DO. THEY ARE USED IN THE R PORTION OF THIS ASSIGNMENT.

- `setwd(dir)`: Returns the current directory and then sets the working directory to that specified by "dir." It will give an error if it does not succeed.
- `ls(...)`: Lists the data sets and functions that have been defined in the current environment by default. Has other optional arguments as well..
- `str(obj, ...)`: Displays the internal structure of an R object specified by "obj." Has other optional arguments as well.

5 R Output

```
[1] "cars_df"      "longley_df" "my_name"     "students"    "years"
```

```
'data.frame': 16 obs. of 7 variables:
```

```
$ GNP.deflator: num 83 88.5 88.2 89.5 96.2 ...
```

```
$ GNP : num 234 259 258 285 329 ...
```

```
$ Unemployed : num 236 232 368 335 210 ...
```

```
$ Armed.Forces: num 159 146 162 165 310 ...
```

```
$ Population : num 108 109 110 111 112 ...
```

```
$ Year : int 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 ...
```

```
$ Employed : num 60.3 61.1 60.2 61.2 63.2 ...
```