5225 – Group Project Work Plan/Documentation Process

Contributors: Ruofan Chen, Liquan Zhong, Sara Rahim

*Overall objective*: Produce useful products using the FBI’s Crime Dataset (using R) for a client

It includes descriptive analysis and modeling.

***March 15th – March 28th***

Objective: Data cleaning and organizing/data-analysis ideas

First Group Meeting: March 13th

Tasks:

1. Organize all datasets in R (data cleaning)
2. Transform PDF files into excel then open into R
   1. Using Excel’s ‘Get Data’ -> ‘From PDF’ -> Manual fixing
3. Come up with ideas on how to analyze the data
   1. Implement Flexdashboard/Shiny (learn Flexdashboard/Shiny)
   2. Modeling/Clustering/Regression
   3. Using US Census Data
4. Read datasets from GitHub directly in R (learn GitHub)
5. Schedule meeting with “manager”
   1. First meeting scheduled: 3/16/21 at 4:45 pm

***April 3rd – April 9th***

Problem: Bugs found in code

Objective: Finish data cleaning/remove bugs + new ideas

Second Group Meeting: April 3rd

Tasks:

1. For years 1999-2004, check if all ‘states’ are showing correctly
2. For years 2005-2019, modify; add ‘Total’ using grepl
3. For years 1999 – 2019 remove numbers in the ‘state’ column using gsub
4. Choose only one rape definition (for 2012, use rape definition 1, and for 2013, use rape definition 2)
5. Data manipulation: make a column named “crime\_type”
6. “Rate” is calculated by crime\_type/(population/100,000)
7. Implementing Flexdashboard/Shiny ideas:
   1. **Map**
      1. State total – num/100,000 inhabitants
      2. Slide bar – choose year
      3. Also choose crime type
      4. Reference: <https://github.com/mamajumder/usa-crime>
   2. **Animated Bar Chart** 
      1. Crime type by area for years 2010-2019
   3. **Clustering** 
      1. Total cluster/violent crime and property crime
      2. Bubble plot
   4. **State Trends Analysis** 
      1. Barplots for property crime vs. violent crime
      2. Crime rates over all years
8. Schedule second meeting with “manager”
   1. Second meeting scheduled: 4/8/21 at 4:45 pm

***April 10th – April 18th***

Final stages of project

Objective: Minor data cleaning for specific analyses, work on Shiny results, do modeling, put everything together

Third Group Meeting: April 17th

Tasks:

1. Finish preparing 1995 and 1998 data
2. Finish removing new bugs found in 1999-2019 datasets
3. Make several pages in Dashboard
4. Add color scheme on the map (add interactive plot) + add Alaska and Hawaii using ‘usmap’ package
5. Finish writing cluster functions for 1995-2019
6. Find new variables that possibly affect crime rates (according to FBI’s website and report), get prepared for regression (time series regression and panel data regression)
7. Finish regression function and Granger Causality test.
8. Explain regression by state results and do visualization
9. Finish code for the barplot animation – some data cleaning (remove states that do not have all data available)
10. Add all results/functions to Dashboard – Edit titles/subtitles – Add notes where needed
11. Add documentation to codes/files – modify folder names in GitHub
12. Prepare presentation for April 22nd:
    1. Sara: map plot and animation plot
    2. Liquan: states trend and clustering
    3. Ruofan: regression, granger test, explanations, and conclusions