Data Science for Public Policy

Alex Engler & Aaron Williams - Georgetown University

PPOL 670 | Assignment 03
Data Visualization, Markdown, and Git

Due Date: Friday, February 14th at 11:59 PM

Deliverable: There are three deliverables to submit for this assignment. This includes (1) the resulting project .html file, (2) the .Rmd file with your R code, and (3) the URL of the Git repository. You may need to place the .Rmd and .html file into a folder, and compress (or zip) that folder, in order to submit it to Canvas.

Points: 12 points

Plagiarism on homework or projects will be dealt with to the full extent allowed by Georgetown policy (see http://honorcouncil.georgetown.edu).

Setup

Create a new folder and R project as in the prior assignments. Next, create a new R Markdown file (.Rmd) in that new folder.

Next, start a new git repository in that folder, using `git init`, `git add`, and `git commit`. Then, go to Github.com and log into your account. Click on the green button that says "New" to create a new GitHub repository.

Lastly, follow the second set of instructions, under "...or push an existing repository from the command line."

Please note: Setting up and using Git & GitHub for the first time is complex and unintuitive. We strongly recommend you start this assignment very early and come to office hours or otherwise reeach out with any issues.

Assignment Description

This assignment differs from those previous, in that you are tasked with finding and analyzing a dataset of your own choosing. This will result in you creating and submitting a single .html file, containing your code, a series of four visualizations made with ggplot2, and a brief discussion of what those visualizations mean.

First, find a public dataset available from the web, relevant to one of your policy interests. Add this data to a folder called 'data' in the same folder as your markdown file. You should also create a `.gitignore` file and add the file suffix of your data (likely .csv, .xlsx, or .sas7bdat) to that file, so git will ignore your data.

Then analyze the data, storing all your code and writing in your .Rmd file, following the requirements described below. There are three graded components of this analysis (Read all instructions before starting, as you must use git as you create the graphs).

1. Four ggplot2 graphs and markdown - 6 points

Your analysis should have four data visualizations of distinct graph types, made with ggplot2. Across all four graphs, use at least:

- Six different aesthetics;
- Six different non-aesthetic options;
- Four different geoms;
- Two different scales (meaning change the default scale used for at least two aesthetics).

Further, each graph must include:

- Correct usage of all visual encodings;
- Appropriate data sourcing;
- Proper labeling of ALL visual encodings;
- An appropriate title and subtitle;
- The code you used to generate each graphic, right above the graph (R Markdown should make this easy).

2. Written narrative & interpretation of visualizations - 3 Points

Write at least three to five sentences about each graph in your markdown document, describing what it says and how it informs relevant policy topics. You should (to the extent possible) write this as a narrative that ties together all of visualizations. You will be graded on the interpretation of the visualizations, and their relevance to the policy topic.

3. Git Repo and Commits - 3 Points

Use `git` thoroughly for this assignment. This means, at a minimum, you must commit when you first start the project and after the completion of each graph (a bare minimum of five commits). When you submit the assignment, all commits must have been pushed to GitHub. This means there will be a record of the commits on the public GitHub repository. You must submit the GitHub URL as part of the assignment, and will be graded on having committed your code as you worked on the assignment.

Submission

Upon completion of the assignment, knit the .Rmd file to .html, and submit both, along with the URL of the GitHub Repository.