CS614 – Prof. Berardi
Module 2
November 2-4, 2020

We've spent a lot of time becoming familiar with ggplot and RShiny, so this exercise will focus on visualizations. The data set for this module is called The Force Report and was created by NJ Advanced Media NJ, who over a 16-month period submitted Open Public Records Act (OPRA) requests for use-of-force forms from 2012 through 2016 for every single local police department. These forms are required to be completed by each officer involved in any incident in which force is used, including, but not limited to, compliance holds, takedowns, pepper spray, baton strikes, tasers, and firearms. The law that required these forms to be completed envisioned an early warning system to identify officers who might be using excessive force. However, this database was never built, so Advanced Media NJ built it. The incident specific version of the data is behind a paywall, so we will use the free version of the data which contains summary level statistics for each municipality. The data is provided on Canvas and a data dictionary can be found here.

It should be noted, that this is a database of incidents in which force, not necessarily excessive force, was used. Using force is a normal and necessary part of policing to protect both the public and the officer. Given that this is the single greatest power given to law enforcement, it is important, nonetheless, to monitor force usage to identify when boundaries of acceptable use of force are being crossed. This is not intended to be a database of "bad" policing and using more force is not necessarily indicative of something nefarious. However, this database has the ability to serve as a critical monitoring system to identify certain departments and/or individuals who might warrant closer investigation.

For this assignment, you will work with your partner to build an RShiny application that will allow the general public to interface with this data. This app should allow the user to filter the data into two groups. An easy stratification point to use is county, but you can also stratify by a variable in the data set (e.g., choose the top X% and bottom Y% of # of force incidents). Your app should automatically i) create a visualization and ii.) perform a statistical test that explores some aspect of the data of your choosing. There are a wealth of variables concerning the amount of force that is used, types of force used, the racial breakdown of the officers and force recipients, the number of officers meeting early warning criteria.

Ideally, your analysis will be a comparison of the two selected groups. The statistical test could be as simple as a correlation or as complex as a collection of linear regressions with interactions (or beyond if you have the skills). The results and interpretation of the statistical test(s) should be automatically displayed. Ideally, there should be some link between the visualization and the statistical model – but they do not have to summarize the exact same thing. If you'd like to incorporate other data about population, poverty, election results, etc., it is extremely easy to create an account on NJ Data Book and download municipality level statistics. This will require a data merge with the data I've provided. If you'd like to try something else, for example exploring other trends in the data or identifying outliers - see me to discuss.

Make sure that your app is clean with no errors that appear when to little or too much data is selected. Figures should be polished and legible with properly labeled axes, consistent capitalization, and legend titles with nice labels that are properly placed within or near the figure.

One you've finished, please host it on R Shiny's io service. You will submit a brief paragraph orienting me to your app, as well as the link. One submission per group is fine, but include all group members' names.