Project Proposal Group 3

Cornelius C. Johnson

March 19, 2019

Group Members

Group 3 consists of Cornelius Johnson, Zhengqi Dong, and Zhaohong Wang. We are a diverse group, as Cornelius is an Economics major, Zhengqi is a Computer Science major, and Zhaohong is an Actuarial Science major.

Data Set

We propose the development of a Shiny app using the Stanford Open Policing Project data, which is a recently posted (3/19/19) Tiny Tuesday data set. The data is hosted online at `https://openpolicing.stanford.edu/data/` and is available in both CSV and RDS formats. Additionally, the available data sets are separated by state and/or city. The Summary-level data set can be accessed by executing the following:

```
combined_data <- readr::read_csv("https://raw.githubusercontent.com/5harad/openpolicing/master/results/c
```

Researchers on the Stanford project use RStudio for their analyses, and have generously posted their code along with numerous tutorials on the website. This is very encouraging as we will have an abundance of resources to rely upon when carrying out our own project.

Another encouraging aspect of this choice of data set is its richness. We will likely use the Summary-level data set in order to simplify the workload, but even so there are some very interesting variables at our displosal. Consider the following table:

variable	class	description
location	character	County/District location for each incidence
state	character	State for each incidence
driver_race	character	Driver's race
stops_per_year	double	Number of stops per year
stop_rate	double	Stop rate (stop = police stop of a vehicle) (%)
search_rate	double	Search rate (%)
consent_search_rate	double	Consent to search rate (%)
arrest_rate	double	Arrest rate (%)
citation_rate_speeding_stops	double	Citation rate for speeding stops (%)
hit_rate	double	Hit rate (%): proportion of searches that find contraband
inferred_threshold	double	Inferred threshold - based off the threshold test described in the paper

In total, our `combined_data` data set has 2688 observations, and thankfully it is already tidy.

Questions

Our Shiny App will primarily be concerned with issues of racial discrimination in police traffic stops. Specifically, we would like to answer the following questions:

- Which states have the highest stop rates and highest search rates?
- Do different states have different racial distributions for each type of stop?
  - If there are significant differences, can we understand the spatial distribution of these differences?
- Specifically, do black drivers have higher search and citation rates?
- Does racial bias exist?
  - If so, what is the spatial distribution of racial biases?
- What are possible solutions for racial bias that are supported by the data?

A few of our racial bias questions have been touched upon in some way by the Stanford researchers. However, we still believe that it would be beneficial to address these questions independently and compare our findings. Additionally, we were surprised to find that the Stanford researchers have not significantly looked into the spatial distributions of their data set. We are excited to do so!

Design Ideas

Regarding design ideas, the following itemized list is an early ordering of potential interests.

- Scatter plots to identify and characterize potential relationships between variables.
- Stacked bar charts to display differences in driver race for various stop types.
- Map depicting the spatial distribution of each variable, along with racial interaction terms.
  - This would be at varying levels of aggregation.

We are particularly interested in including a map of the US with the color for each state being dependent on different factors. We would certainly like to do this dependent on the number of stops per year, but it may be more inciteful to include some of our new interaction terms between race and stop, citation, and/or hit rates in order to understand the spatial distributions mentioned previously.

Further accuracy could be achieved by allowing readers to zoom in on the map and look at local/county levels of aggregation. Specific inspiration for the map design stems from an example found in the Shiny gallery, at `https://datasociety.com/kitamba-the-opportunity-project/`. This is not something that has been covered in the course, but we believe (encouraged by the Project Proposal instructions) that it is possible to learn and complete within the allotted timeframe.