Mingqian Wu, Felix Lyu

Project Proposal

**Formulate a problem of interest:**

The root cause of crimes has been a hotly debated topic among sociologists. While this project will not directly investigate the root cause of crimes, it seeks to map out the correlations between a selected number of factors with the crime rate over some selected regions. For instance, the considered factors may be a region’s average income, average years of education, population density, average number of firearms per household, etc. We plan to consider only violent crimes in our project. We want to restrict the geographic scope into the following regions: New York, NY San Francisco, CA, Seattle, WA, Nashville, TN and Austin, TX.

**Gather a dataset that is relevant to the problem:**

Data sources are available on government agencies' open databases, such as data.gov.

For our violent crime’s data, we plan to look at:

<https://ucr.fbi.gov/crime-in-the-u.s/2010/crime-in-the-u.s.-2010/download-printable-files>

For our graduation rate’s data, we plan to look at:

<https://catalog.data.gov/dataset/edfacts-graduates-and-dropouts-201112>

For our education spending per student’s data, we plan to look at:

<https://www2.census.gov/govs/school/10f33pub.pdf>

For gun ownership, we plan to look at:

<http://worldpopulationreview.com/states/guns-per-capita/>

For socioeconomic data, such as poverty rate, unemployment rate, and education levels, we plan to look at:

<https://catalog.data.gov/dataset/county-level-data-sets>

Overall, we plan to only look at the year of 2010 and the factors listed above for now. We may expand the time range and investigated factors in the future. We plan to only extract the data for the above factors from the comprehensive dataset and group them by the investigated cities. We expect most of them to be quantitative data except for the investigated cities themselves.

**Derive a set of questions you want to answer from the data**

* What are some factors that lead to high or low violent crime rate for city?
* What is the relationship among the investigated factors? For instance, is there a correlation between average firearm per household and education spending?
* For cities that have high or low crime rate, what similarities do they have? Vice versa, what makes a city with a high crime rate different from a city with a low crime rate?
* If a city were to lower its crime rate, which aspect should it focus on that might be most effective in reducing its crime rate?

**Design a visualization that will help answer these questions.**

A series of scatterplots that show correlations between all the factors and the violent crime rate. For each scatterplot, the x-axis represents the specific factor, and the y-axis is the violent crime rate.

To compare the relationship between factors, we plan to use a parallel coordinates graph, each mapped to the line using a linear scale, and the cities will be encoded categorically using colors.

A better design is still to be determined.

**Features: what will your visualization achieve?**

We want interactions to select specific cities, factors or years.

**Visualization Design Sketch**

**Project Schedule**

Oct 21 - Oct 27:

Team formation and project description due on Friday.

Oct 28 - Nov 3:

Project proposal due on Friday.

Collection of raw data and basic preprocessing of raw data.

Nov 4 - Nov 10:

Start visualization implementation: basic website layout and titles.

Check point presentation on Friday.

Nov 11 - Nov 17:

Data joins and graph making.

Nov 18 - Nov 24:

Continue working on visualizations.

Prototype due on Monday.

Nov25 - Dec 1:

Continue working on visualizations and testing.

Dec 2 - Dec 8:

Presentation on Monday.

Write up results and thoughts.

Dec 9: Final Submission.