Crime in 2013

Kevin Ma (kym5), Shangyuan Niu (sn522), Ajay Gandhi (aag255)

March 8, 2016

Data Retrieval

The data used in this presentation was spread across a series of websites, namely:

- Disaster Center http://www.disastercenter.com/crime/uscrime.htm
- worldatlas http://www.worldatlas.com/articles/us-poverty-rate-by-state.html

Disaster Center

In order to use data from the Disaster Center, we wrote a scraper in Node.js that saved each state's data together in a large JSON file. Because data on this website was displayed in HTML tables, the scraper mapped these tables to JSON. Additionally, numbers were displayed with comma-separated thousands, so each datapoint was cleaned and converted to an integer. The data on this website described yearly crime rates, broken down into the following categories:

- Violent
- Property
- Murder
- Forcible Rape
- Robbery
- Aggravated Assault
- Burglary
- Larceny - Theft
- Vehicle Theft

In addition to the above categories, the website provided yearly population and a "crime index", equal to the total property and violent crimes per year. Using the population and crime totals, we were able to compute per capita and per-hundred-thousand crime rates.

worldatlas

The data from worldatlas comprised of a measure of poverty in the United States. The states were ranked in descending order of poverty, along with a percentage representing the fraction of residents below the poverty threshold (the poverty threshold is the minimum amount of income a household needs). This data allowed us to compare crime rates, which are relatively meaningless on their own, to a seemingly related subject.

The data from worldatlas was also stored in an HTMl table; however it was a simple matter to copy and paste the data before formatting into JSON.

Data Visualization

We wanted to represent two facets of our data. The first was the simple question of how crime was distributed across the United States. The second, more compelling comparison was how poverty rates were correlated with crime rates.

Crime By Category

The most intuitive format to represent crime rates would be to create a choropleth of the United States and vary the value of each state's color based on its poverty percentage.

We created a color scale, using a different hue for each category of crime (violent, property, etc.). This allowed us to represent each category distinctly over the same base layout. Additionally, we computed the minimum and maximum of each category to create a sufficiently varied choropleth each time.

Crime and Poverty

In order to compare crime and poverty, a simple scatterplot would suffice. In order to create a more interesting illusion we created a scatterplot of states represented by circles. We varied the size of each circle by the state's population. This implied linear scales for the axes and a square root scale for the size of the circles. This allowes us to see a very direct relationship between the position of these circles and the data they represent.

What Can We Learn?

The choropleths show us a direct comparison of crime rates between states. At first glance, it becomes evident that New England seems much safer than any other region of the US, except perhaps the northern midwest. On a related note, the South demonstrates some of the highest crime rates in the States.

The scatterplots give us more complex insight into the relationship between poverty and crime. Intuitively, we would expect some categories of crime to demonstrate a stronger correlation than others; burglary would increase with poverty. Indeed, the scatterplots do tell us this - for states with higher poverty rates, burglary has quite a strong correlation with poverty. On the other hand, rape seems to be unrelated to poverty as one would expect.

Despite these results, there are several anomalies. For example, murder and poverty demonstrate a direct relationship, while larceny and robbery really show no evident relationship to poverty rates.