

Socioeconomic Factors and Crime

Taylor Riesselman, Mike Javon, Derek Hasley
Reginald Dawson, Jena McJunkins

Introduction to Data Science
University of Nebraska at Omaha

October 23, 2014

Abstract

There are many factors that affect crime levels in the United States. In this project, we explore how socioeconomic factors such as population, unemployment rate, and income levels are correlated with crime levels. We also explore how crime levels change over time and by location. This is done by creating two interactive data products, in which the user can view crime data and independent variables affecting it in easily understandable, useful ways. One of the two data products produced is an interactive map in which the user can compare crime levels by state and year. The other data product produced is an interactive scatter plot generator which allows the user to view the data in many ways and interpret what it is showing. Both data products aid in answering the questions; how does crime change over time and by location, and what factors have an effect on different crime levels.

1 Introduction

Crime is a broad term used to describe an act in which laws are broken and punishment is warranted. Crime can be broken out into many different forms, including the ones used in this analysis as shown in Table 1. These different types of crime are documented and kept track of in terms of number of crimes. It can also be viewed in terms of a rate of crime per 100,000 people. Both ways of tracking and analyzing crime are useful and necessary in any society.

	Type of Crime
1	Violent crime total
2	Murder and nonnegligent Manslaughter
3	Forcible rape
4	Robbery
5	Aggravated assault
6	Property crime total
7	Burglary
8	Larceny theft
9	Motor vehicle theft

Table 1: The types of crime analyzed in this project.

Crime, in its many forms, is affected by multiple independent variables. Such variables include time, location, population, unemployment, and income levels. All of these variables play a part in the changes in crime and crime rate.

2 Explanation of the Data

The data used in this project to produce the two data products can be found in one of two places. The first is on the St. Louis FRED website. This website provides the yearly socioeconomic data for the United States. This data includes unemployment rate, income levels, and population. The website providing the different types of crime by state and year is the FBI's Uniform Crime Reporting Statistics website. From these two sites, enough data was obtained to create two interactive and informative data products.

2.1 Obtaining and Preparing the Data

The data for this project was obtained via the websites above and compiled into two CSV files to be read into R. The first CSV file contains the levels of crime, population, income, and unemployment for the United States by year. This data was read in and used directly, so no data manipulation in R was needed. The second CSV file contains the levels of different crimes by state and year. This data was edited and merged with map data from the ggplot2 package. The code used in editing the data for the map data product is shown below.

```
colnames(crimeData) <- gsub("\\.", "_", colnames(crimeData))
mdat <- map_data("state")
crimeData <- sqldf("select lower(State) as region, Violent_crime_total as value, Year from crimeData")
yearm <- merge(mdat, crimeData, by.x = 'region', by.y = 'region')
yearo <- yearm[order(yearm$order),]
```

3 Data Products

Two data products were generated in order to effectively convey crime levels in different ways. These data products were created using the shiny package. Both data products are interactive and allow the user to easily view and interpret the data. **Both of these data products will be modified and improved for the final project.**

3.1 Scatter Plot Data product

The first data product created allows the user to choose a socioeconomic factor and a type of crime. It then produces a scatter plot using the chosen variables. This plot makes it easy for the user to see the correlation between any combination of socioeconomic factor and crime type. Figure 2 shows an example of a plot that can be generated using the interactive scatter plot data product.

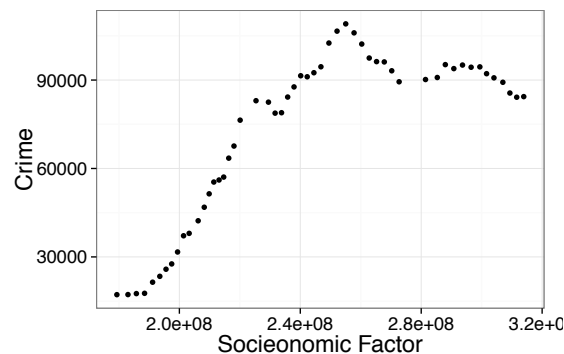


Figure 1: This is an example of a plot that can be generated using the interactive scatter plot data product. This plot represents the correlation between Population and Forcible rape

Note that this is the image of the scatter plot for one type of crime and one socioeconomic factor (in this case, Population and Forcible Rape). But in the interactive version there are two drop down lists allowing the user to choose any combination of socioeconomic factor and crime, which changes the data passed to the scatter plot. The next step we wish to take is to improve this plot with the use of a more sophisticated plotting package.

3.2 Map Data product

The map data product is an interactive map, in which the user can use the slide bar to view the violent crime totals by state for any year selected. An example of a map that can be created with this data product can be seen in Figure 1.

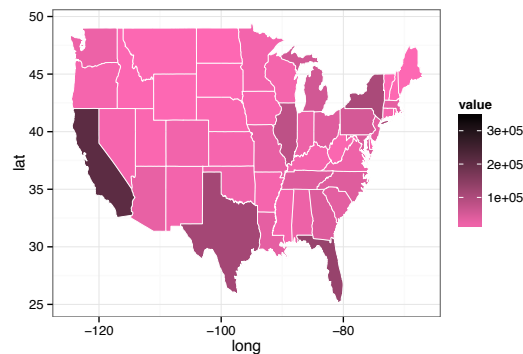


Figure 2: This is an example of a plot that can be generated using the interactive map data product. This map represents the total violent crimes in the year 2000.

Note that this is the image of the map at one instance of time (in this case, the year 2000). But in the interactive version there is a year slider that will alter the year, which changes the data passed to the mapping function, so you can quickly see the map change over time. The next step we wish to take is to add a drop-down where you can select different columns so you can visualize the change in a multitude of crime types.

4 Findings

Our findings are limited so far, as this is still a work in progress, but our preliminary findings are:

4.1 Scatter Plot Data product Findings

The scatter plot data product clearly illustrates the correlation between each type of crime and socioeconomic factor. Despite what we originally thought, unemployment does not seem to be strongly correlated with any of the types of crime. These are only preliminary findings; more analysis will be done.

4.2 Map Data Product Findings

With the map data product, it is easy to see that crime levels have continuously grown throughout time. This data product also shows that the states with larger cities, such as California and New York, generally have higher crime levels. These are only preliminary findings; more analysis will be done.

5 Conclusion

We have no real conclusion at this point as this is still very much a **work in progress**. When we submit our final version of the project our conclusion will go here.

References

- [1] Uniform Crime Reporting - Federal Bureau of Investigation,
<http://www.ucrdatatool.gov/Search/Crime/State/StatebyState.cfm>
- [2] Economic Research - Federal Reserve Bank of St Louis,
<http://research.stlouisfed.org/>

