The Crimes that Divide Us

Keenan Venuti

Kvenuti@college.harvard.edu

Harvard University

The Politics of Personal Data

Gov 1430

ABSTRACT

This research began with a single purpose, but it has since branched to tackle a larger goal. Starting out by simply analyzing differences in crime amongst states, I quickly realized that data and metadata can tell an intricate story about individuals and populations. This analysis aims to show key differences in criminal activity across states over a 10 year period supplemented by etiological explanations. Using clustering, an unsupervised machine learning classification technique, I found significant clusters of states indicating similarities in crime rates within specific regions of the U.S. By classifying different states, I hope to provide a framework to further research legislative, cultural, societal, fiscal, geographical, and demographical similarities between these states that may cause their likeness in criminal activity.

INTRODUCTION

In our current America, we face constant reminders of how divided our country is. Political lines, societal differences, financial gaps, all serve as pertinent acknowledgments to how different two citizens can be. These divides can be self-created, self-enforced, or a product of larger environmental pressures. While we have always known of our inherent differences, accumulation and analysis of data has prevented us from fully expressing and investigating the subtleties that divide us. Using newer data science tools and sensors, developed to understand the influx of data-overload, we can see trends across time and location. To do this, I pursued an analysis of the changes in crime, per state, across custom made crime rate indexes. My goal is to group states based on similarities that one can't see unless a crime index is used. This analysis is ongoing as I look for and identify the causalities behind statewide criminal similarities.

For example, numerous psychology studies have been performed to understand location's effect on behavior and personality. One study found that men from Southern states exhibit more aggression when insulted (Cohen, 1996). Research like this suggests causation between location and crime rate; but, within the context of this study, excessive crime happens everywhere and in all types of places. Starting with research in cultural differences based on geography, I hope to explore these relationships further to understand why distinct state clusters exist.

MATERIALS & METHODS

Relevant Definitions

- Clustering: Classifying an observation through its similarities and differences to other observations.
- Crime Index: A multi-dimensional score/tool, based on either violent crime or theft related crime statistics, used to find distances between states for clustering.
- K-means clustering: A clustering algorithm that repeatedly calculates a cluster mean and creates a new cluster based on observations closest to that mean.
- Agglomerative clustering: A clustering algorithm that initially creates an individual cluster for each observation and iteratively combines clusters until there is only one cluster comprised of all observations.

To better understand how crime was collected on a city and state level, I harvested two data sources. I used urldefense.proofpoint.com to collect individual crime data and ucrdatatool.gov to collect crime metadata across years and states. Once I had my datasets, I performed a cleaning process to isolate crime occurrences in each state across 10 years. The data can then be divided into smaller subsets based on type of crime (theft, murder, assault, etc.), state, and year. With a cleaned dataset, I isolated the years 2004, 2009 and 2014. From there, I created two crime indexes based on violent crime and theft related crime. This allowed me to cluster states using the two major types of crime. I used both a k-means and agglomerative clustering algorithm to generate my clusters as they are two very different algorithms. This helped me get a better read on the layout of the data; there are pros and cons to any clustering algorithm. Each algorithm looks at individual states/observations and calculates distances to other observations based on the crime index previously created. From there, it links similar states together into a cluster which is comprised of similar observations. Within this analysis, there are usually 2 or 3 clusters.

RESULTS

I checked how well my clustering algorithm worked by using a silhouette score. It determines how many observations were clustered appropriately. The algorithms implemented all achieved positive silhouette scores, implying that the clustering was a success and similar states were grouped. Here are 2 time-lapse maps of the U.S. and the states that fall under each cluster. The darker the color of the state, the more crime occurs there. As you can see, states can undergo many changes across 10 years.

2004 Clustering on Violent Crimes

2009 Clustering on Violent Crimes

2014 Clustering on Violent Crimes

2004 Clustering on Theft Related Crimes

2009 Clustering on Theft Related Crimes

2014 Clustering on Theft Related Crimes

CONCLUSIONS

From these visualizations, we can clearly see a progression of crime in different states through the 10 year period. However, one should focus on the surprising segregation of North and South consistent across these maps. Even West coast states like California and East coast states like Florida seem to cluster with the South. These clear divisions, in tandem with well-separated clusters, shows a divided America. However, seeing the division is much easier than knowing why it exists.

NEXT STEPS

My next step will be to research and document the legislative, cultural, societal, and geographical reasons this divide exists. I believe there are precise reasons for these distinct clusters, just as there are well-understood reasons for political and fiscal divides. Using meta-crime data is helpful in understanding macro-social trends, but this is just the surface of the full story. I plan on utilizing crime datasets composed of individual crimes to better understand motive on a micro level.

REFERENCES

Cohen, D., Nisbett, R. E., Bowdle, B. F., & Schwarz, N. (1996). Insult, aggression, and the southern culture of honor: An" experimental ethnography.". Journal of personality and social psychology, 70(5), 945.

Crime State by State. UCR. Accessesd 4/26/17. https://www.ucrdatatool.gov/Search/Crime/State/StatebyState.cfm

Haidt, J. (2012). The righteous mind: Why good people are divided by politics and religion. Vintage.

Police Records Data. URL Defense. Accessesd 4/26/17. https://urldefense.proofpoint.com/v2/url?u=http-3A__publicsafetydataportal.org_participating-2Dagencies_&d=CwMGaQ&c=WO-RGvefibhHBZq3fL85hQ&r=i4dmVVIljwEegFeQ_qgf1Er6StxS0coKka6i2YgNHUE&m=Fjzb6gg2mMj2URFDPLw8y2bg8wfkV_uLZhRmZauoS1Y&s=i_ofthr8yvYhMNMS8tOEcabYb0O-Bcl-NZaMXgHP4oM&e=