

## *Poster Project Report*

### Project Topic and Relevance (~ 300 words)

I have always been interested in crime that is difficult to explain and even taboo to discuss. Being Bulgarian and my parents being immigrants with heavy accents I have personally seen and felt what its like to be hated for simply who you are. I wanted to understand the variables behind these types of crimes, especially, within this political climate I feel like now is a vital time to talk about this uncomfortable topic and understand the structure of these hate crimes. This data can also be flexible in the way it is interpreted giving the ability to understand many aspects of US culture. For example, this data can look at the relationship between the majority and the minority and what type of tensions there are among communities. This data can also look at where these crimes are taking place and in what jurisdictions, possibly giving an insight into poverty or racial/identity segregation within cities and states. This topic is relevant to so many people since America is a country of immigrants, thus affecting a lot of people and looking at why this hatred persists even in these “modern” times. When thinking of solutions to a problem you always have to listen and understand both sides of an event in order to get at the core of the problem so this investigation allows for a look at the construction of hate crimes piece by piece. In addition, the way to approach this topic is to look at trends within the US to get a sense of why certain races or identities are being targeted, predominately in which states, and what the bias behind the hate crime is. All these components are necessary to begin a constructive analysis of why this happens and then from there we can look at remedies and preventions of these types of crime. In addition, looking at this data allows us a sense of what type of cultural climate our country is in and what type of progress is being made and even predict the future rates of hate crime. Lastly, the most important part to me is that this data could allow possible solutions that can save future lives and families from assault simply due to their identity as a person because we will be able to predict patterns of hate crimes before they occur.

### Finding Data (~ 350 words)

I found my data through the ICPSR website where it hosts the Uniform Crime Reporting Program data on hate crimes. After using the ICPSR website in the first few labs I decided to use the website again because this website has free open access data, where anyone can download the entire data set. What I wanted to do with the hate crime data is discover why do hate crimes persist within this “modern” era? Basing my research off of this general question I wanted to specifically look at who is being predominantly targeted in the past and in the current time. I also wanted to know where these crimes take place not only in what city/state but also in what location because that

can also be just as telling as to why a crime was committed and the dynamics of the relationship of the victim/offender. This data looks at a special set of crime that criminology literature does not look at too much so this data would also fill in a gap that most do not fill and pave a way to looking at why hate crime persists. When I first took on this data set I knew I wanted to look at the bias motivation from the beginning but I was unsure of what other variables would be relevant to understand this bias motivation. I first thought that the variable GOFFRAC, the offenders race, would be the most obvious and most understanding of the bias motive but as I discovered along the process I found that this variable was not as helpful as intended. The rate at which the offenders race is known is either white or unknown, which does not tell us a lot about the dynamics of the offenders. Through this experience I found that the most obvious assumptions are not always the best ones to understand data. This caused me to look at less obvious variables such as crime location which turned out to be statistically significant with the bias motive. The entire process changed my view on how theoretical questions are answered through understanding data differently.

#### Data Structure & Data Munging (~ 500 words)

My original data had around 170 variables with about a little over a thousand incident reports. What I wanted to investigate with this data was to look at all of the incident reports but only look at a certain group of variables. A lot of my original data dealt with categorical data besides the amount of victim/offenders and population size of the city. The way that I wanted to filter out my data was to focus on the most important aspects as to why people commit these crimes. So the variables that I focused on was the bias motive, location of the crime, the offenders race, and the amount of victims/offenders within an incident, the city/state the crime took place, and the victim type. In order to get the 9 variables that I found the most useful in answering my question was to first create a subset of both my 2002 and 2013 data's by using the pipe operator to filter out all the other variables and create a data table to only show my relevant variables. This allowed me to digest both datasets data better and create a better understanding of how these variables work within each incident report. Then I rearranged my data making the variables within my new data table better to understand as if you are reading a book from left to right so my data began with the city name, then state, then population size and so on. For my data I did not have to merge any data except the total number of victims plus offenders to generally know how many individuals were associated within the incident to better understand how many people are involved in crimes like this. After creating my meaningful data set I began to create histogram visualizations on the frequency of each variable that I found useful. For example, I wanted to create visualizations that were simple and easy to understand and able to compare between 2002 and 2013 so I first created a histogram that graphed the frequency of each bias motivation to look at which type of hate crimes are the most in both years.

Then I created another frequency histogram of the location type of the crimes to know where these crimes were taking place the most. Lastly, I graphed the frequency of which states had the most hate crime incidents reported. These graphs allowed me to see and compare these frequencies, however, a draw back is these frequencies can leave out the unreported hate crimes and for the state frequency it doesn't take into account a states population size. In terms of a statistical analysis I did want to compare two variables and see if they were statistical significant to understand if these two variables are correlated in understanding hate crimes. When I ran a bivariate linear regression on the bias motive and the location of the crime I was able to see if these two were statistically significant and able to answer my questions. When doing this I learned that even if you have two variables that seem relevant they actually any not be significant which can also tell you something about the data and the nature of the data its self. On the flip side I found that these two variables, bias motivation and location of crime, are statistically significant meaning that they do have some relationship but to what extent I am not sure which made me realize that research is never done and there can always be more questions that need to be answered. Even if you don't find any correlation does not mean that the research was pointless. Throughout this process I learned that creating and conducting models will always give you an interesting finding into what you are looking for even if it is no what you expected. This surprised me the most about this type of research is that even if you have assumptions they will not necessary be supported by the data, thus, forcing you to develop different assumptions about people and their relationship to society.

### Reproducible Research (~ 350 words)

Reproducibility in research is incredibly important and necessary because it allows for the use of generalizability and thus affectively making conclusions about the topic/affects of what you are researching. In turn this allows for proper understanding and proper solutions to be implemented within the real world and at a broad scale. When another social scientist can take the same raw data that you worked with and come out with the same results that you generated this shows that first, your conclusions are accurate and hold more validity and second, your data can be presented as a generalizable idea with real world implications. The way we allowed for open data to possibly be replicated is by using github which is a free online cite that lets you upload your code and data processing. This is not only helpful for others to see how you analyzed your data but it was also helpful for me to understand the way that my data evolved and transformed throughout the weeks. I was also able to look back and understand more about how to create a better-organized dataset that way it is easier to interpret and thus reproduce. The ways that I wanted my data to be reproducible is to keep my analysis simple and understandable. Also when I am running bivariate linear regressions I wanted to look at only certain variables with each other so I can understand the relationship

between different variables thoroughly (which two variables were statistically significant or not). However, what I want to incorporate more into research moving forward is to learn more about creating a multivariate regression model for a lot of different categorical variables to see if they affect one another. I would also like to create a little bit more complex visualization because this will allow me to see an intersection between variables and create a deeper understanding of my data. What I would also like to incorporate in my data is numeric data because I have worked with majority only categorical variables and I think that adding other numeric variables will really help understand the volume of data, specifically my hate crime data. Lastly, what I would like to investigate about my hate crime data in the future is to try and account and understand hate crimes in the US that was not reported to the police and not recorded by the government and thus not in my data. I believe this unrecorded data will also tell us a lot about the US and its culture behind varying states that have low recorded hate crime data.