Will Gaston
COSC 5610: Data Mining
Professor Shion
11/5/2020

Cook County Regional Gang Intelligence Database

Part 2: Data Exploration and Visualization

1 Introduction

The Cook County Regional Gang Intelligence Database portrays information about 24,307 criminals mostly located near Chicago and within the states of IL and IN. The data includes specific physical characteristics of each gang member such as gender, height, weight, eye color, hair color, race, and age. The set also includes unique information such as what gang the criminal belonged to, what suburb he/she lived in, if the criminal had tattoos or not, and other questions related to gang affiliation. There are numerous characteristics of these criminals that describe what gang members in this region specifically look like. This allows for patterns among gang members to be found, so that other gang members and criminals who have not been captured can be more recognizable and easier to catch. I will be analyzing characteristics of this dataset in this exploration in order to help identify trends of gang members.

1.1 Cleaning and Brief Exploration

There is a lot of criminal data involved within this set, but many of the columns/characteristics of these gang members are unimportant for my research and exploration. I wanted to clean up and remove some of this useless data to make things easier for my actual analysis and visualization. I first went through the whole dataset, recognizing and identifying every column and member that seemed critical and also irrelevant for my study. I removed columns that were clearly useless and unfit for this project. Some of these columns included information about if the gang members were felons, on probation, or deceased. I erased time as well as geographic data such as dates and zip codes. These identifications of the criminals and their crimes will not inform much about what gang members look like, what they tend to look

like, and where they live, which are the main points of my exploration in this project. There were also some similar and repetitive columns I noticed throughout the data. I made sure to delete these repetitive columns of data to simplify what was needed and help me focus on the right things.

As I went through the data visually without any evaluating or statistical analysis, I wanted to point out some key points and traits in the data. It is important to establish an initial ground of exploration before digging deeper into the patterns of things within the data to give an opening about different aspects and patterns in the figures. It is important to see certain patterns within any dataset because some relationships may be obvious and without the need to explore further, while other connections can be hard to see and definitely in need of further evaluation. One thing I wanted to explore and use for more data exploration throughout this project was the comparison and contrast of male to female gang criminal members. The data shows male gang members as the clear, dominant sex for gang affiliation in this region. It is heavily favored away from the female sex, but this could be for different reasonings like geographic location or other information. It is important to know that there are hundreds of gangs within this set. This means a big piece of exploration should be associated with the location of the majority of gangs in the region as well as which gangs are the largest. The age of these gang members fit widely from eighteen to eighty, but it is important to acknowledge that most of the gang members fit mostly between thirty and fifty. There are less gang members associated within the age range of 18-25 as well as 65-80, however. One last key pattern with determining criminals and gang members is race. The majority of this dataset involves criminals of the black race. The next two in order seem to be Hispanic and then white.

These are all key features of the dataset that are important to know before digging deeper into the numbers. It gives us a brief introduction into the data as well as helps us establish some reasoning and beliefs about the data before we analyze further. Throughout this exploration, I aim to see the basic characteristics of this dataset. I want to see the see the different trends of these key features discussed above that are important to visualize and understand before digging into specific modeling.

2. Exploration

2.1 – Sex

As I mentioned before, the sex of the dataset is heavily favored toward the male sex. Due to this, there is no need to account for sex in future analyzing and predicting. I want to show the difference between male and female gang members. Figure 1 shows the count comparison of the two sexes while Figure shows the percent comparison. Male gang members surpass female gang members by 23,007 members and by 95.9%. I will account for both sexes throughout my further analysis, but it is important to know that the evaluation will basically be pointed towards trends affiliated with male gang members.

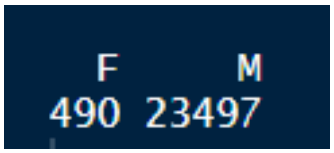


Figure 1 – Sex Count Comparison

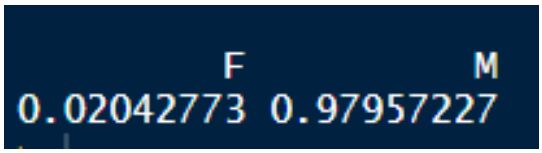


Figure 2 – Sex Percentage Comparison

The ‘F’ accounts for female gang members, while the ‘M’ accounts for male gang members. There are barely any female gang members in this dataset, which surprises me. I figured a heavier gang membership with the male sex, but not as large as this. I predicted a higher criminal account on the female side. It is interesting to know that the difference is so drastic in this region.

2.2 - Race

It is essential to acknowledge and see relationships among physical traits within gang members. Race is a physical trait that helps identify trends in gang violence and activity. I want to analyze the racial aspect of the data to see which races are involved in gang violence, which races are most popular for gang violence, and which races are least popular for gang violence. You can see this information in Figure 3 below.

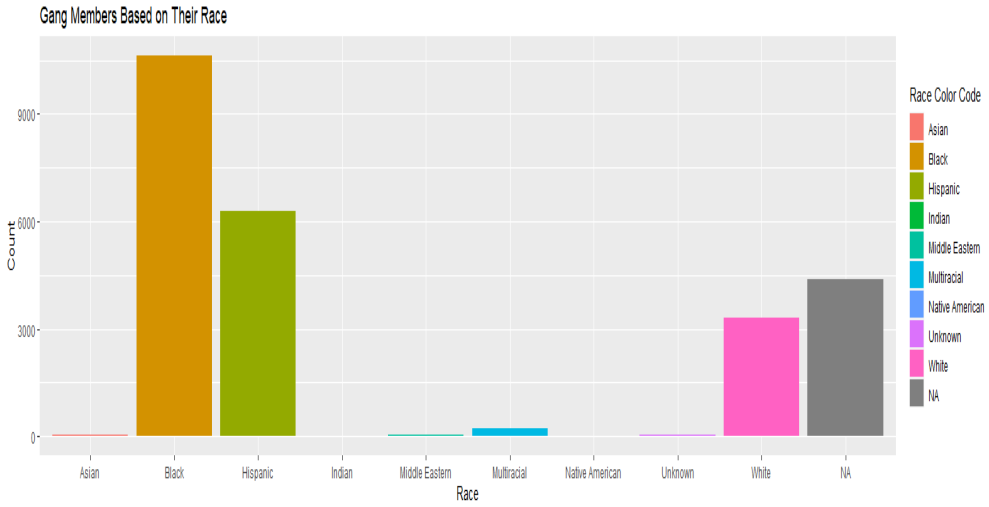


Figure 3 – Race Affiliation with Gang Members

There are eight races listed above with two other categories, unknown and NA. The unknown gang members were not identified racially when arrested, and the NA gang members' race were not known as well. It is obvious to see that the black, Hispanic, and white race all occupy the majority of the data. The other races do not have that much of an impact as their values are very small compared to the top three races discussed. It was sad to see that nearly 50% of the dataset includes black gang members and 25% of Hispanic gang members. That is a very high number for just one region. It was also interesting to see the white gang member percentage to be around 15%. I thought this number would be lower for the white race. This is a great indication of how race plays into gang violence.

2.3 – Age

Another important physical feature of gang members that I would like to oversee is age. Age is a great indicator of gang trends and criminal trends. Gangs and their members tend to stay close in age and behavior, so it is an important aspect to discuss with gang violence. You can see the age ranges of the gang members in Figure 4 below.

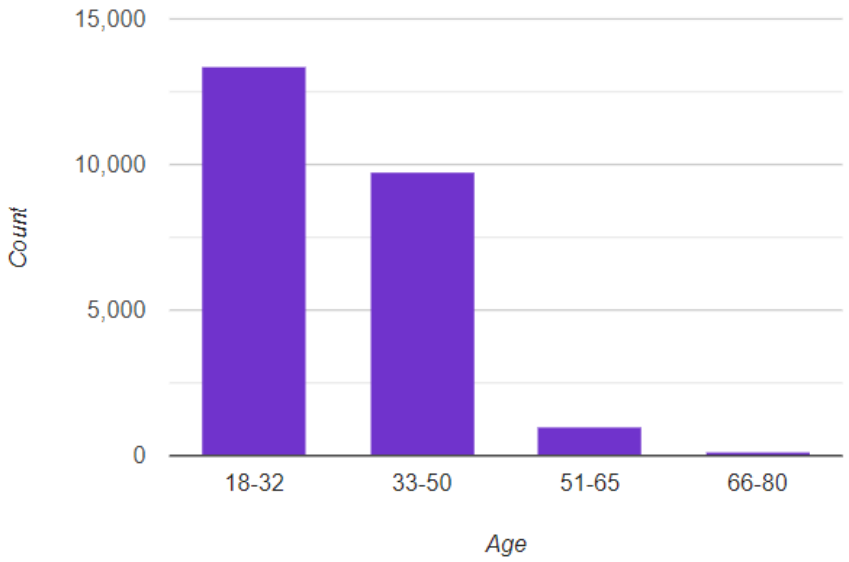


Figure 4 – Age Ranges of Gang Members

The ages of gang members throughout this dataset ranged from 18 to 80. I decided to split this range into four intervals as shown above. There were very heavy trends and drastic numbers from the age of 18 to 50. There were around 54% of gang members that were aged 18-32, and 41% of gang members aged 33-50 within the dataset. Gang members radically decreased between the ages of 50 and 80. This makes sense to me as gang members do not tend to be older. Gang violence involves many young and middle-aged adults. It was interesting to see that there were actual gang members within the age of 50 to 80. I even was surprised of the high number of gang members between 33 and 50. I thought gang members were usually within the range of 18 and 35. This is a great outlook of what specific ages gang members tend to be.

2.4 – Gangs

As I examined the data, there are hundreds of gangs included within these regions of the set. Many were located in small geographic locations, while others spread across many. It was interesting to see the different names and different numbers of gangs that actually exist. It was also scary to see how many gangs there are in just this region. It says a lot about the rest of the country and how gang violence really does exist in most of our world. You can see the range of gangs below in Figure 5 as well as how many suburbs they are located within.

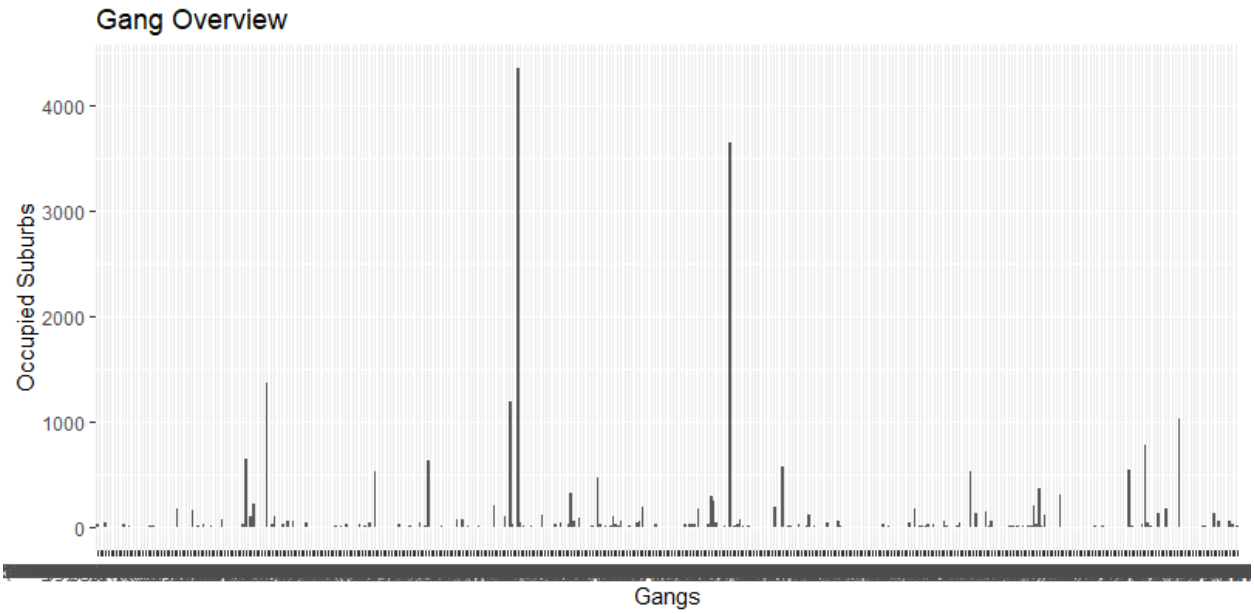
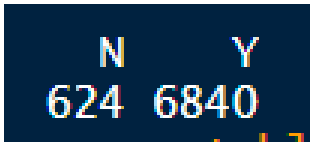


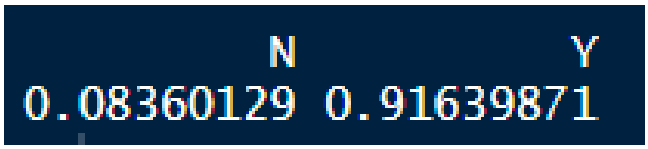
Figure 5 – Gang Members and Occupied Suburbs

There are hundreds and hundreds of gangs within this region. There are too many to analyze, but I needed to figure out how many exactly before digging further. Most of the gangs tend to occupy 0-500 suburbs, but there are also some popular gangs that surpass 500 suburbs and very few who reach the thousands. This data is also not just located within one state, however. These gangs occupy many cities located mostly within IL, IN, and few in other states. This gives me an overview of the vast number of gangs in the set. I want to further analyze later the more popular gangs within these regions instead of all of them.

2.5 – Armed

An important aspect of criminal activity and gang violence is weapons. When determining gang members and noticing trends throughout, protection is key. It is vital to know whether these gang members are armed out in public in front of innocent people. This will help protect innocent citizens and families. You can refer to Figure 6 and 7 below to examine the differences between armed gang members and unarmed gang members.





The ‘Y’ accounts for armed gang members, while the ‘N’ accounts for unarmed gang members. There is a big difference here. There are hardly any unarmed members as the count reaches 624, while the count for gang members reaches 6,840. This is a 92% chance a gang member is armed and a 8% chance they are not. Another big piece to notice is that the majority of this dataset, around 30%, do not have data on being armed or not. This tells us that the majority of gang members were unknown to have weapons or not. From the pattern, however, you can see the obvious trend of how many gang members bring their weapons out. These numbers do surprise me, however, because I figured there would be less armed members as usually only some members of a gang carry weapons instead of everyone. It is very fearful to realize how many have weapons and how many carry them around.

3. Conclusions Drawn

There were a lot of important trends and relationships within the dataset that will be useful in further modeling and analyzing. The majority of the gang members are male, so these trends and relationships throughout correspond to male gang members within this region. The different races of the gang members were identified. Three top races of black, Hispanic, and white were found to include higher gang activity. The ages of gang members tended to fit in between 18 and 50, mostly 18 and 32. There are numerous gangs in this set, so further modeling will be done for more popular gangs. Gang members also tend to be armed, another important factor for future modeling. This data exploration and visualization helped me really dig deep into this dataset and find important details among the data.

References

[1]. Ann Marie Sorenson, David Brownfield, and Kevin M. Thompson. 2010. Deviant Behavior, Vol. 22 Issue 1: Gang Membership, Race, and Social Class – a Test of the Group Hazard and Master Status Hypotheses. Taylor & Francis Online. https://www.tandfonline.com/doi/pdf/10.1080/01639620175006581O?needAccess=true.

[2] Anon. 2020. Criminal Justice Fact Sheet. (July 2020). Retrieved November 3, 2020 from https://www.naacp.org/criminal-justice-fact-sheet/.

[3] Anon. Bar Chart & Histogram in R (with Example). Retrieved November 2, 2020 from https://www.guru99.com/r-bar-chart-histogram.html.

[4] Anon. Categorical Data Descriptive Statistics. Retrieved November 2, 2020 from https://ucr.github.io/descriptives_categorical.

[5] Anon.Retrieved November 3, 2020 from https://towardsdatascience.com/a-guide-to-datavisualisation-in-r-for-beginners-ef6d41a34174.

[6]. J. A. Laskey. 1996. Gang Migration: The Familial Gang Transplant Phenomenon, Vol. 3 Issue 2: Journal of Gang Research. National Criminal Justice Reference Service. https://www.ncjrs.gov/App/abstractdb/AbstractDBDetails.aspx?id=160870.

[7]. M. Lane. 1989. Inmate Gangs, Vol. 51 Issue 4: Corrections Today. National Criminal Justice Reference Service. https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=118632.