**Firearm Confiscated Data 2008-2017**

**Louisville, KY**

**Exploratory Analysis**

Christina Porter, [cporter@bellarmine.edu](mailto:cporter@bellarmine.edu)

1. **INTRODUCTION**

In Spring 2021, I have been participating in the Justice sector in a collaborative effort from Bellarine University, Central High School, and the Microsoft Future of Work Initiative for what is known as the Butterfly Project. In this project, my sector is working with ACLU to collect and analyze data on gun crimes in Louisville, KY after a devastating record of homicides in the year 2020. After this data is analyzed, ACLU should be able to take our findings and make a meaningful change within our community in response.

I will be analyzing a combination of three different datasets provided from the Open Data Archive of Louisville. The three datasets I will be using describe crimes recorded by the Louisville Metro Police Department where some type of firearm was confiscated. These datasets consist of:

* firearm\_data\_intersections13to17, which describes incidents at traffic intersections
* FirearmAddress, which describes incidents at a residential or commercial address
* FirearmsData, which describes all incidents.

I have merged these datasets into one called firearms\_data to obtain a more comprehensive view of the gun crimes committed in Louisville between the years of 2008 and 2017 and my findings are presented below.[[1]](#footnote-1)

1. **DATA SET DESCRIPTION**

This dataset contains 11195 distinct incidents described by 21 columns represented by strings, dates, and floats. A complete listing is shown in **Table 1**. For this data, I decided to keep my null values because they could prove important in determining whether a certain type of data is more likely to contain missing values.

**Table 1: Data Types and Missing Data**

|  |
| --- |
|  |

1. **Data Set Summary Statistics**

To further analyze the information within our data, I first generated a list of the summary statistics of the columns listed as float64 in Table 1. These statics contain the total count of values from each category, the mean, standard deviation, min, 25th percentile, 50th percentile, 75th percentile, and max. These findings are presented in **Table 2.** Further, I also aggregated tables from the remaining object and datetime64 values that present the unique values listed for that column, the number of times they appear, and the total proportion of that value within the column.[[2]](#footnote-2) These findings are presented in **Tables 3 – 12**. Please see the appendix for the proportions for the incident number, manufacturer, model, caliber, block address, and street address. Then, I created a correlation matrix that shows the correlation coefficient value between each of the four values represented within my summary statistics in **Table 13**. The graphical representation of this correlation matrix is shown in a heatmap in **Figure 1.**

**Table 2: Summary Statistics for firearms\_data**

**

Table 3: Proportions for ucr\_category



Table 4: Proportions for zipcode



Table 5: Proportions for city



Table 6: Proportions for year



**Table 7: Proportions for firearm\_subcategory**

**Table 9: Proportions for person\_recovered\_from\_sex**

Table 10: Proportions for firearm\_category



Table 11: Proportions for geocode\_address\_type



Table 12: Proportions for person\_recovered\_from\_race



Table 8: Proportions for state



Table 13: Correlation Table/Tables



Figure 1: Correlation Heatmap

Shape, square

Description automatically generated

1. **DATA SET GRAPHICAL EXPLORATION**

After taking a look at the values within my data, I was able to take a further look at the graphical representation of my data. There were not any numerical values where creating a scatterplot would make sense for analysis purposes, so there are no scatterplots included; however, we will be taking a more in-depth look at the distributions included within the data, bar charts of my categorical variables, and other analysis figures.

* 1. *Distributions*

In this project, I did not take the time to plot the proportional tables to see the distribution between each categorical variable; however, this tool would be useful for a more in-depth look to see how my data in each area is spread. While plotting my data, I happened to take a look at the total number of crimes committed for each age and I happened to notice that this data seems to take an approximately normal distribution that is skewed to the right. Additionally, the mode, or the data point that appears the most often, is 23, which means that more crimes were committed by those who were 23 years old. This can be shown in **Figure 2.** Noticing this, I then attempted to create a density plot with 20 bins. This density plot confirmed that the data is approximately normal that is skewed to the right as shown in **Figure 3.**

*Chart, histogram

Description automatically generated*

Figure 3: Count of Gun Confiscations per Age (Density)

*Chart, histogram

Description automatically generated*

Figure 2: Count of Gun Confiscations per Age

* 1. *Barcharts (categorical variables)*

The majority of my analysis has consisted of creating bar charts for various combinations of my categorical variables. These charts provided a very interesting analysis that could create more questions about the data that I am analyzing. Each chart provided is plotting using a logarithmic scale because the difference between some values in my data is so vast that it is hard to see some of the smaller values.

**Gun Types**

The first values that I plotted within my data was the breakdown between the gun types used for crimes. In my Python notebook, I grouped these values by first the category (Pistol, Shotgun, Rifle) and the subcategory or type of gun. The first plot I used was a stacked bar chart that presents each category in one column but breaks the column down into proportions of each type of gun that was used. This chart is useful in determining which gun category was used in crimes the most, which was the rifle as shown in **Figure 3.**

I also plotted these values side by side grouped by category type to allow a better view of the use of each gun subcategory, and interestingly enough, even though the rifle was the most used gun category in gun crimes, the semi-automatic pistol was actually used in the most crimes above any other type of gun as shown in **Figure 4.** I also plotted this same information in **Figure 5** which represents the amount of crimes committed with each gun subcategory with the gun category color coded. This plot shows us that the most used gun type was the semi-automatic pistol and the least used gun type was the percussion pistol.

**Types of Crimes Committed and Race**

Next, I was interested to see the breakdown of the various crimes committed. This analysis was done in conjunction with the analysis for the race, so I created a bar plot very similar to the plot created for the gun type analysis and used both a stacked and unstacked bar chart to analyze the types of crimes committed by each race. Shown in **Figure 6**, the type of crime where firearms were confiscated the most were weapons law violations followed closely by narcotics. This data is also portrayed in **Table 3** where the proportions of the ucr categories are discussed.

Using the unstacked **Figure 7,** I noticed some interesting information when we showed the incidents by race side-by-side. More firearms seem to be confiscated from black individuals for weapons law violations, theft, assault (including only reports of injured persons), fraud, and disorderly conduct categories. In comparison, more firearms are confiscated from white individuals for narcotics, liquor law violations, impersonation, rape, burglary, and bribery. Overall, more guns are confiscated from black individuals than any other race as shown in **Figure 8**.

Chart, bar chart

Description automatically generated

Figure 8: Guns Confiscated by Race

**Types of Crimes Committed by Zipcode**

To further analyze the data breakdown between the types of crimes committed for each race, I was curious to see whether this depending on location. In **Figure 9**, I grouped each zipcode in my data by race. As shown by Figure 9, significantly more guns were confiscated from black individuals with zipcodes 40056, 40202, 40203, 40206, 40208, 40210, 40211, 40212, 40213, and 40215 while more guns were confiscated from nonwhite individuals with zipcodes 40214, 40216, 40218, 40219, and 40229. In further analysis, I would be interested to see if this data is proportional to the amount of black or nonwhite individuals living in each zipcode. I did not graph the type of crime versus the zipcode because that would be a huge graph, but that would be a useful tool for further analysis with the combination of this information divided up by race.

Additionally, I performed the same analysis with the sex of individuals. We already know the crime where the most firearms are confiscated from is weapons violations, so I did not replicate this information with the sex. Rather, I went straight to using my unstacked plot to view where guns were confiscated from each sex for every zipcode. In every zipcode, more firearms were confiscated by males than females as shown in **Figure 10**. Based off this information, it is not surprising that more total firearms were confiscated from males as shown in **Figure 11.** In this analysis, I did not separate the ucr categories by sex but that information will be used in further analysis, and I also hope to break up this analysis by both sex and race.

Overall, the zipcode where more firearms were confiscated was 40202. Based off information in Figure 9 and Figure 10, the most firearms confiscated in this region were from black and male individuals, but not necessarily black male individuals. This possibility will be explored in further analysis.

*Chart

Description automatically generated*

Figure 11: Firearms Confiscated by Sex

Based off this information, I would be interested to see if there is a correlation between the number of times each zipcode is patrolled versus the number of confiscated firearms. This can possibly be explored in a separate analysis as I do not have access to this data.

*Chart, bar chart

Description automatically generated*

Figure 3: Stacked Gun Type Confiscations

*Chart, bar chart

Description automatically generated*

Figure 4: Unstacked Gun Type Confiscations

*A picture containing rectangle

Description automatically generated*

Figure 5: Top Guns Confiscations

*A picture containing text, writing implement, stationary, pencil

Description automatically generated*

Figure 6: Crime Categories Divided by Race (Stacked)

*Chart, histogram

Description automatically generated*

Figure 7: Crime Categories Divided by Race

*Chart, bar chart

Description automatically generated*

Figure 9: Firearms Confiscated by Zipcode/Race

*Chart, bar chart

Description automatically generated*

Figure 10: Confiscated Firearms Zipcode/Sex

*Chart, bar chart

Description automatically generated*

Figure 12: Firearms Confiscated by Zipcode

* 1. *Other Plots*

For the remainder of my analysis, I attempted to use other plots to convey my information. The first of these includes a boxplot for every zipcode. To create this plot, I found the number of incidents where guns were confiscated from each year for every zipcode. Each boxplot represents the summary statistics of the median, min, 25th percentile, 50th percentile, 75th percentile, and max. The green arrow on each plot is the mean.

Some things to note here:

* The tighter the plot, the more consistent the values were across every year. Likewise, the more spread out a plot is, the more variance there was in the number of guns confiscated from every year.
* Higher boxplots indicate more overall guns confiscated for that zipcode across every year.
* The tails on the boxplot indicate the max and min for each zipcode.
* If the mean on the boxplot is lower than the bar that represents the median, you likely have some outliers in the data.
* This boxplot is on the logarithmic scale and not the normal scale.

In this data, the logarithmic scale values are vastly different from the normal scale values, which on a graph are significantly more difficult to see and gives the normal scale boxplot an entirely different shape than the logarithmic scale boxplot. This boxplot is shown in **Figure 13.**

I was also able to create a pivot chart of the number of confiscated firearms per zipcode per year. From this pivot table I was able to create a heatmap that easily shows the contrast between the number of firearms confiscated on the logarithmic scale for each zipcode and each year. This heatmap is shown in **Figure 14.**

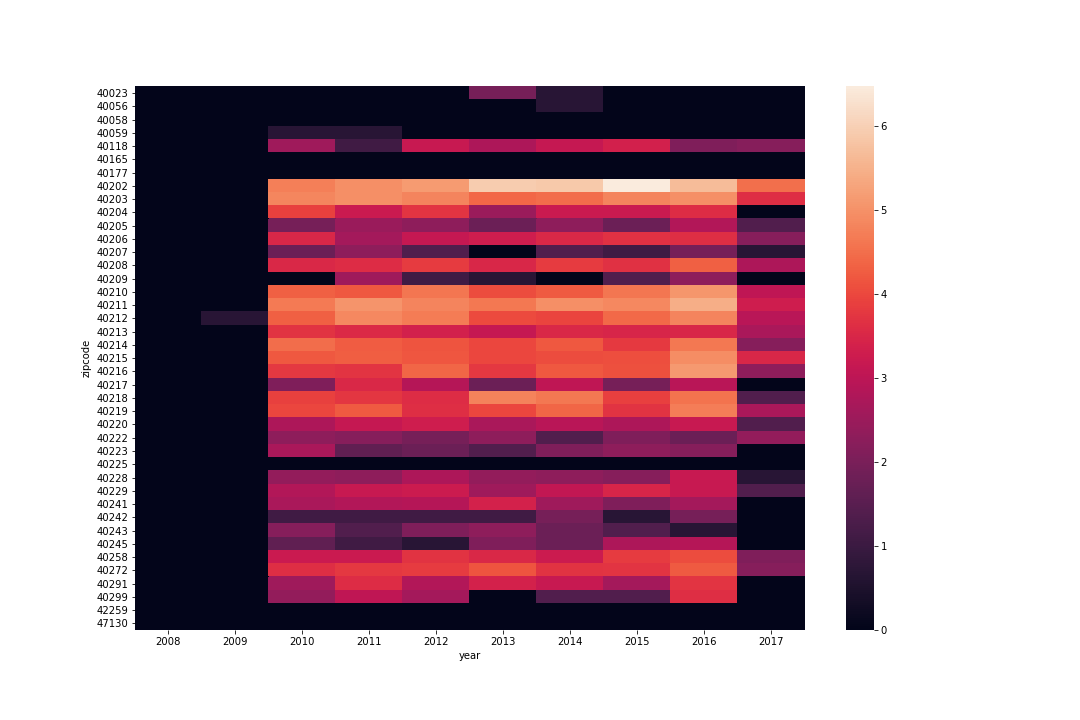
**

Figure 14: Heatmap of Gun Confiscations Year/Zipcode

Finally, I was able to aggregate several images of the trends in the number of firearms confiscated per zipcode every year. While the heatmap has the capability to show this information, each zipcode has its own line graph representing the trends in firearm confiscations for each year. These are shown in **Figures 15, 16, and 17.**  These images were plotted separately and thus are not on the same axes. On further analysis, I will combine these plots on the same axis to view comparable trends. For these plots, it is important to note that some data does not exist for all zipcodes for each year. Additionally, the data aggregated for the year 2017 is incomplete.

*Chart, box and whisker chart

Description automatically generated*

Figure 13: Boxplot of Zipcode/Confiscations Across 2008-2017

**

Figure 15: Zip Code Gun Confiscation Trends 2008-2017 (1)

**

Figure 16: Zip Code Gun Confiscation Trends 2008-2017 (2)

**

Figure 17: Zip Code Gun Confiscation Trends 2008-2017 (3)

1. **SUMMARY OF FINDINGS**

With the analysis on this data set thus far, I was not able to find a correlation between any variables that are included in this set; however, this data set would be better used to try to correlate between other data sets. For a more in-depth analysis, this dataset could be analyzed with numbers aggregated from additional data sets. From the data that I have aggregated, I was able to determine that while rifles are the gun that is confiscated by LMPD, the semi-automatic pistol is confiscated more than any individual type of rifle. Of gun confiscations, more guns seemed to be confiscated for weapons violations with narcotics following closely. Additionally, more guns are confiscated from people who are in their 20’s, and the most guns have been confiscated from individuals 23 years of age, black individuals, and within the 40203 zipcode.

Of the guns confiscated, more weapons were confiscated from black individuals for weapons law violations, theft, assault (including only reports of injured persons), fraud, and disorderly conduct categories. In comparison, more firearms are confiscated from white individuals for narcotics, liquor law violations, impersonation, rape, burglary, and bribery. While this analysis is still very much incomplete, and not all values may be entirely accurate because there a lot of missing values. For this analysis, I dropped categories that were included but didn’t have any corresponding information to them. I believe it provides a solid base and opens the way to develop more questions as the analysis deepens. Overall, this dataset is better suited to be paired with another dataset analyzing information including each category present in this dataset.

1. **APPENDIX**
   1. In my analysis, I was not able to include proportions for incident number, manufacturer, model, caliber, block address, and street address due to the size of the tables. Instead, each proportion table can be found by following the URLs provided below.
2. *Incident Number*

https://github.com/cporterbellarmine/IP4\_CNP/blob/main/unique\_val\_df\_incident\_number.csv

1. *Firearm Manufacturer*

https://github.com/cporterbellarmine/IP4\_CNP/blob/main/unique\_val\_df\_firearm\_manufacturer.csv

1. *Firearm Model*

https://github.com/cporterbellarmine/IP4\_CNP/blob/main/unique\_val\_df\_firearm\_model.csv

1. *Firearm Caliber*

https://github.com/cporterbellarmine/IP4\_CNP/blob/main/unique\_val\_df\_firearm\_caliber.csv

1. *Block Address*

https://github.com/cporterbellarmine/IP4\_CNP/blob/main/unique\_val\_df\_block\_address.csv

1. *Street Address*

https://github.com/cporterbellarmine/IP4\_CNP/blob/main/unique\_val\_df\_street\_address.csv

* 1. This project with ACLU is still in progress. To see further progress, you can visit:
     1. https://github.com/cporterbellarmine/ButterflyProject
  2. Other partners for this project include:
     1. Keturah Herron from ACLU
     2. Dr. Rob Kelley from Bellarmine University
     3. Dr. Monica Unseld from Microsoft Future of Work
     4. Alisia McClain from Microsoft Future of Work
     5. Dr. Jakia Marie from Bellarmine University
     6. Cassaria Edwards from Central High School
     7. Eva Parke from Bellarmine University.

1. \*\*About:\*\* This project is a sample of a project for Bellarmine University and Microsoft FutureLou. For more information, please visit https://cporter741.wixsite.com/butterflyprojectblog [↑](#footnote-ref-1)
2. Null Values are not represented or counted in each table. [↑](#footnote-ref-2)