Firearm Confiscated Data 2008-2017
Louisville, KY
Exploratory Analysis

Christina Porter, cporter@bellarmine.edu

I. 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]

II. 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

Category	General Data Type	Data Type in CSV	Missing Data %
incident_number	nominal	object	0
year	ordinal	object	0.160786065
recovery_date	ordinal	datetime64[ns]	5.743635552
ucr_category	nominal	object	0.008932559
firearm_category	nominal	object	13.55069227
firearm_subcategory	nominal	object	13.82760161
firearm_manufacturer	nominal	object	18.52612774
firearm_model	nominal	object	46.5207682
firearm_caliber	nominal	object	16.57882983
address_geocode_type	nominal	object	27.79812416
block_address	nominal	object	5.743635552
street_address	nominal	object	27.79812416
city	nominal	object	0.21438142
state	nominal	object	0.142920947
zipcode	nominal	object	2.045556052
longitude	ordinal	float64	22.06342117
latitude	ordinal	float64	22.06342117
person_recovered_from_race	nominal	object	18.21348816
person_recovered_from_sex	nominal	object	17.8561858
person_recovered_from_age	ordinal	float64	18.93702546
confidence	ratio	float64	22.06342117

III. Data Set Summary Statistics

[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

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] 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

	count	mean	std	min	25%	50%	75%	max
longitude	8725	-79.44088906	22.43984542	-122.440321	-85.789472	-85.761323	-85.694804	0
latitude	8725	35.38630511	9.994028092	0	38.168676	38.229728	38.253961	41.266511
person_recovered_from_age	9075	30.49763085	12.23066772	0	21	27	36	87
confidence	8725	0.813987278	0.290922929	0	0.664	1	1	1

[2] Null Values are not represented or counted in each table.

Table 3: Proportions for ucr_category

ucr_category	count	proportion
Homicide	206	1.840271574
Suicide	380	3.394675719
Death_Investigation	38	0.339467572
Burglary	265	2.367339646
Wpns_Law_Violation	4511	40.29837413
Narcotics	2150	19.20671788
Aggravated_Assault	1542	13.77523673
Intimidation	134	1.197069859
Stolen_Property	683	6.101482937
Robbery	536	4.788279435
Vandalism	50	0.446667858
Auto_Theft	23	0.205467215
Simple_Assault	117	1.045202787
Shoplifting	22	0.196533857
Theft_Other	109	0.97373593
Kidnapping_Only	6	0.053600143
All_Other_Offenses	91	0.812935501
Theft_Fr_Veh	30	0.268000715
Fraud	18	0.160800429
Family_Offenses	35	0.3126675
Accidental_Shooting	62	0.553868144
Negligent_Homicide	1	0.008933357
Justifiable_Homicide	13	0.116133643
Liquor_Law_Viol	2	0.017866714
Narcotics_Equip	31	0.276934072
Fraud_Impersonation	9	0.080400214
Theft_Fr_Bldg	1	0.008933357
Promoting_Prostitution	1	0.008933357
Injured_Person_Report	19	0.169733786
Forcible_Rape	6	0.053600143
Counterfeiting	71	0.634268358
Disorderly_Conduct	6	0.053600143
Fraud_Credit_Card/Atm	4	0.035733429
Purse_Snatch	1	0.008933357
Bribery	14	0.125067
Theft_Mv_Parts	1	0.008933357
Theft_Fr_Vend_Machine	1	0.008933357
Sodomy_Force	3	0.026800071
Forcible_Fondling	1	0.008933357
Prostitution	1	0.008933357

Table 4: Proportions for zipcode

zipcode	count	proportion
40202	2194	20.00729528
40272	356	3.246397957
40211	1013	9.237643626
40203	871	7.942732081
40215	550	5.015502462
40206	218	1.987962794
40258	262	2.389202991
40241	116	1.057815065
40210	647	5.900054715
40214	500	4.559547693
40205	72	0.656574868
40216	515	4.696334124
40228	93	0.848075871
40212	647	5.900054715
40213	241	2.197701988
40229	162	1.477293452
40219	459	4.185664782
40207	37	0.337406529
40220	144	1.313149736
40118	125	1.139886923
40208	327	2.981944191
40218	503	4.586904979
40222	65	0.5927412
40204	216	1.969724603
40291	178	1.623198979
40299	94	0.857194966
40165	1	0.009119095
40217	113	1.030457779
40245	58	0.528907532
40223	57	0.519788437
40242	28	0.255334671
40209	34	0.310049243
40243	43	0.392121102
40023	10	0.091190954
40059	7	0.063833668
47130	1	0.009119095
40056	5	0.045595477
42259	1	0.009119095
40177	1	0.009119095
40058	1	0.009119095
40225	1	0.009119095

Table 5: Proportions for city

city	count	proportion
Louisville	10769	96.40139647
Shively	150	1.34276251
St_Matthews	18	0.161131501
Hurstbourne_Acres	3	0.02685525
Lyndon	56	0.501298004
Watterson_Park	16	0.143228001
Shepherdsville	1	0.00895175
St_Regis_Park	1	0.00895175
Audubon_Park	2	0.0179035
Worthington_Hills	6	0.0537105
Jeffersontown	20	0.179035001
Rolling_Hills	12	0.107421001
Plantation	2	0.0179035
Forest_Hills	1	0.00895175
Hurstbourne	9	0.080565751
Hollyvilla	8	0.071614001
West_Buechel	7	0.06266225
Moorland	2	0.0179035
Middletown	30	0.268552502
Fairdale	1	0.00895175
Druid_Hills	4	0.035807
Poplar_Hills	3	0.02685525
Jeffersonville	2	0.0179035
Creekside	2	0.0179035
Murray_Hill	1	0.00895175
Mammoth_Cave	1	0.00895175
Williamstown	2	0.0179035
New_Albany	2	0.0179035
Prospect	2	0.0179035
Lynnview	5	0.04475875
Green_Spring	4	0.035807
Houston_Acres	1	0.00895175
Strathmoor_Village	2	0.0179035
West_Point	1	0.00895175
Graymoor/Devondale	1	0.00895175
Norwood	3	0.02685525
Woodland_Hills	1	0.00895175
Blue_Ridge_Manor	1	0.00895175
Windy_Hills	2	0.0179035
Hickory_Hill	1	0.00895175
Douglass_Hills	8	0.071614001
Briarwood	1	0.00895175
Ft_Knox	1	0.00895175
Glenview	1	0.00895175
Brownsboro_Village	3	0.02685525
Woodlawn_Park	1	0.00895175
Fincastle	1	0.00895175

Table 6: Proportions for year

year	count	proportion
2017	383	3.426679789
2014	1494	13.36673526
2016	2245	20.08589067
2015	1756	15.71083475
2011	1365	12.2125794
2012	1363	12.19468551
2013	1378	12.32888968
2010	1189	10.63791715
2009	3	0.026840834
2008	1	0.008946945

Table 7: Proportions for firearm_subcategory

firearm_subcategory	count	proportion
Semi-Automatic	6668	69.11993366
Pump_Action	392	4.063439411
Revolver	1747	18.10925676
Derringer	95	0.984762102
Single-Shot	344	3.565875402
Bolt_Action	211	2.187208459
Flintlock	10	0.103659169
Percussion	24	0.248782005
Over_And_Under	6	0.062195501
Carbine	20	0.207318337
Lever_Action	67	0.69451643
Double_Barrel	35	0.36280709
Automatic	20	0.207318337
Jet_Propelled	8	0.082927335

Table 8: Proportions for state

state	count	proportion
KY	11176	99.97316397
IN	3	0.026836032

Table 9: Proportions for person_recovered_from_sex

sex	count	proportion
F	644	7.003044802
M	8552	92.9969552

Table 10: Proportions for firearm_category

firearm_category	count	proportion
Pistol	8112	83.81897086
Rifle	836	8.638148378
Shotgun	699	7.222566646
Submachine_Gun	18	0.185988841
Rifle-Shotgun_Combination	2	0.020665427
Airgun	6	0.06199628
Machine_Gun	3	0.03099814
Electronic_Control_Weapon	2	0.020665427

Table 11: Proportions for geocode_address_type

address_geocode_type	count	proportion
BLOCK	8083	100

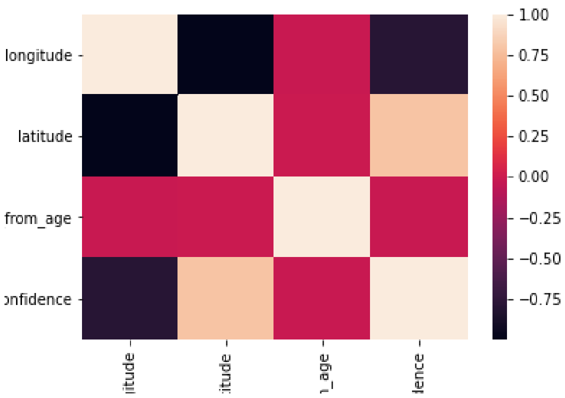
Table 12: Proportions for person_recovered_from_race

race	count	proportion
White	3199	34.93883792
Black	5759	62.8986457
Hispanic	153	1.671035387
Asian	32	0.349497597
Middle_Eastern	6	0.065530799
Indian/Burmese_Indian	2	0.0218436
Native_American	5	0.054609

Table 13: Correlation Table/Tables

	longitude	latitude	person_recovered_from_age	confidence
longitude	1	-0.996470496	-0.012459017	-0.787558106
latitude	-0.996470496	1	0.000834283	0.790198082
person_recovered_from_age	-0.012459017	0.000834283	1	-0.012421206
confidence	-0.787558106	0.790198082	-0.012421206	1

Figure 1: Correlation Heatmap



IV. 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.

A. 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.**

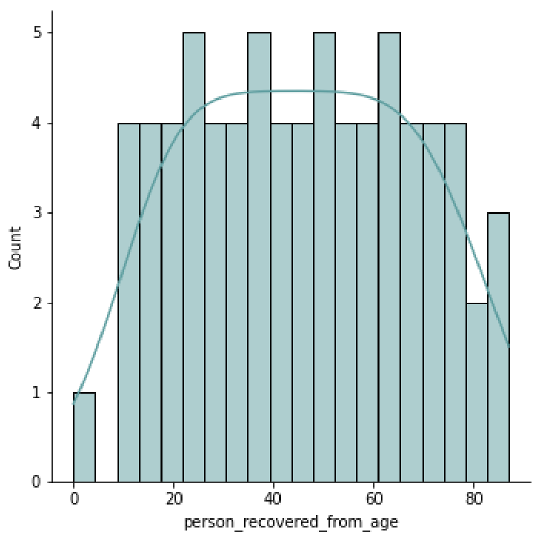


Figure 13: Count of Gun Confiscations per Age (Density)

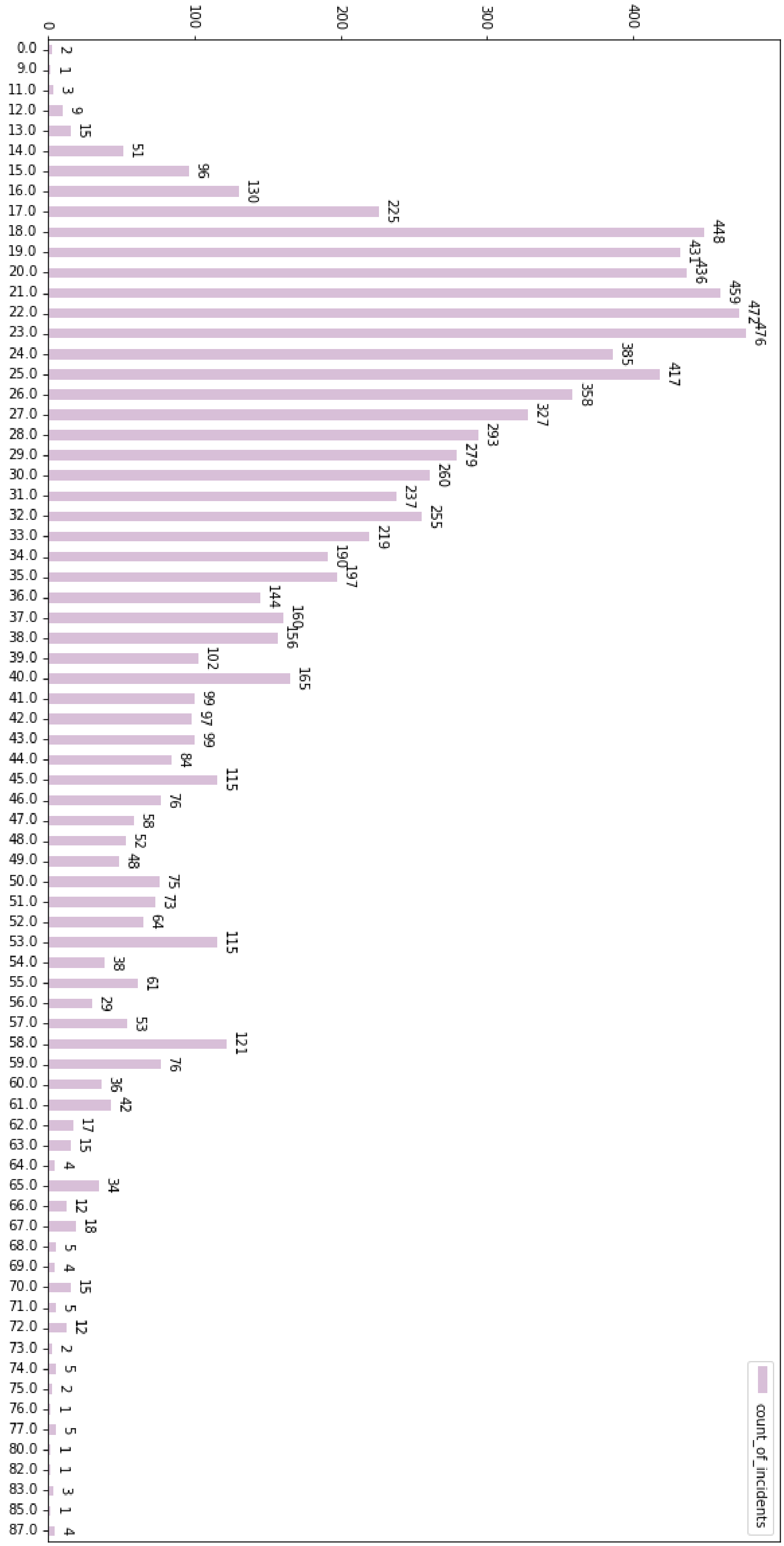


Figure 2: Count of Gun Confiscations per Age

B. 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**.

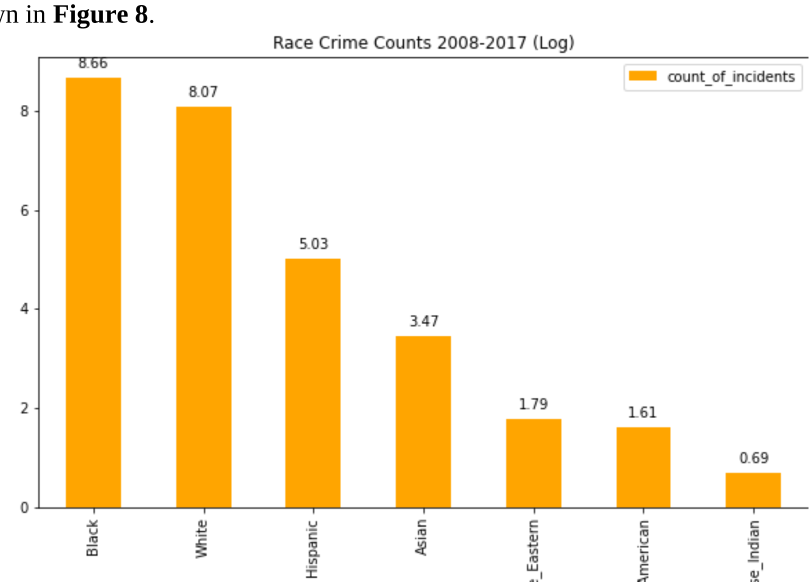


Figure 8: Guns Confiscated by Race

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.

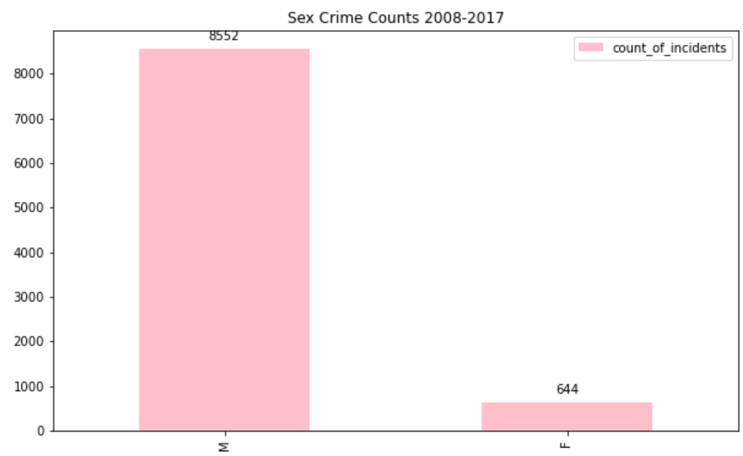


Figure 11: Firearms Confiscated by Sex

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.

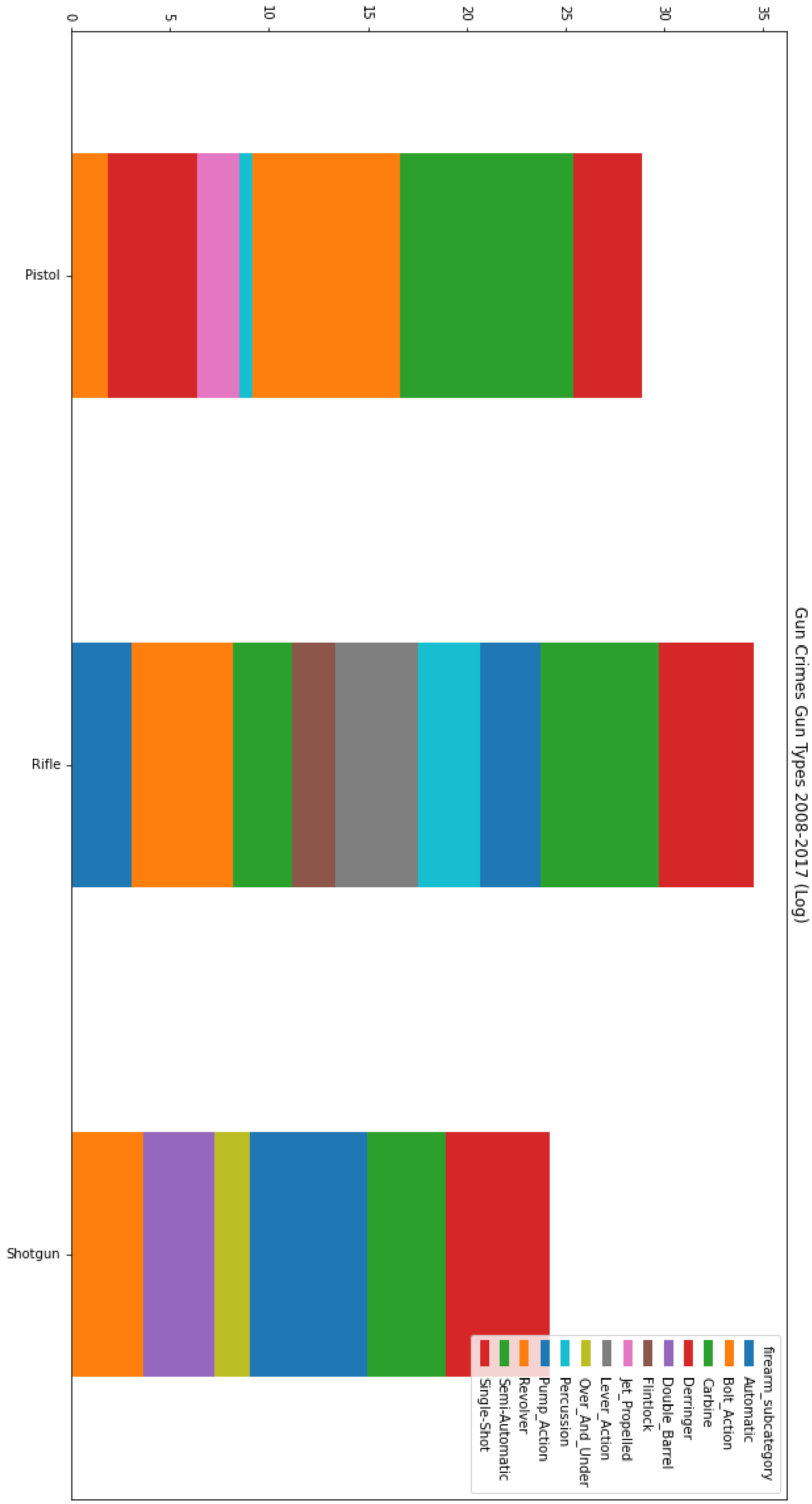


Figure 3: Stacked Gun Type Confiscations

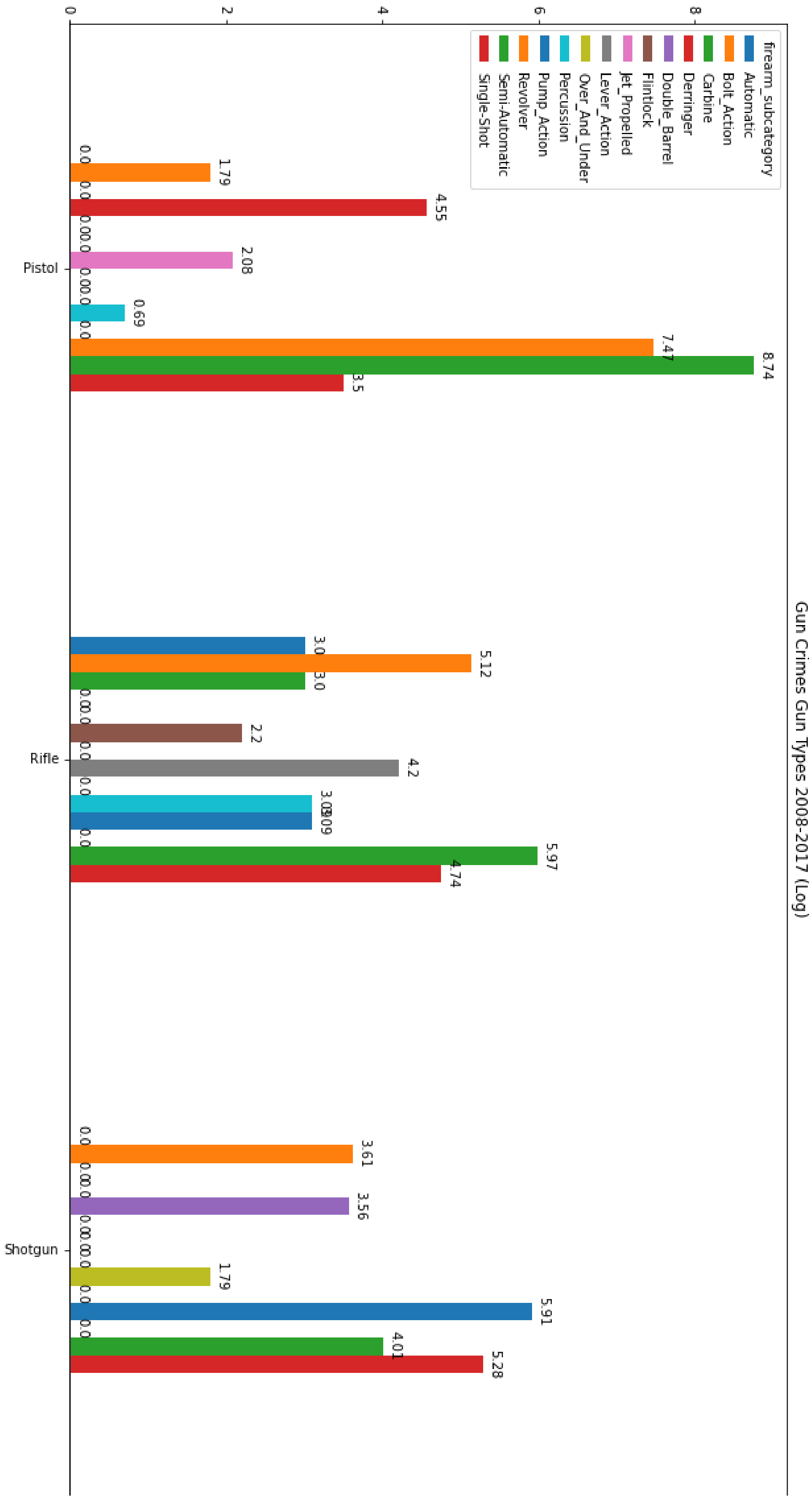


Figure 4: Unstacked Gun Type Confiscations

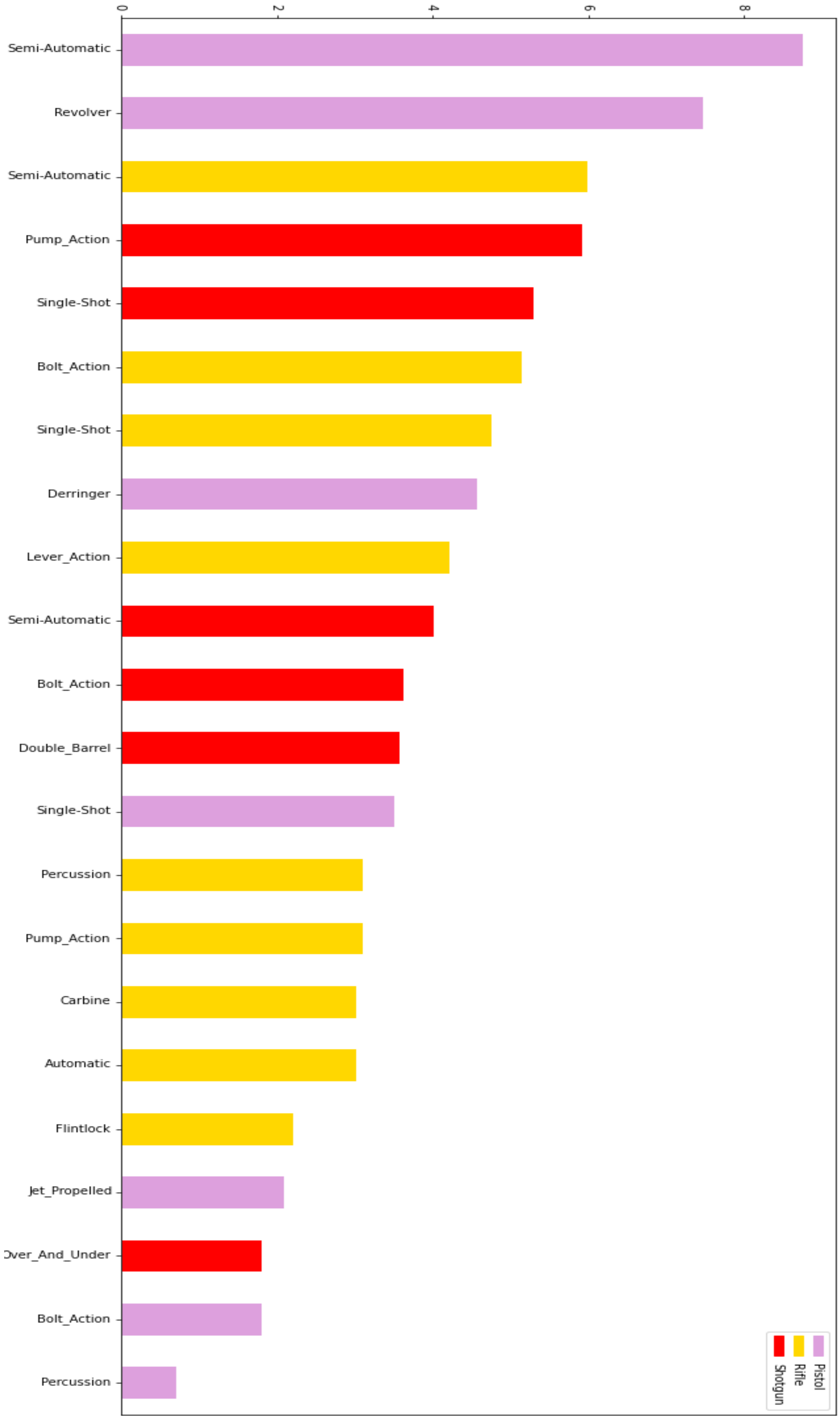


Figure 5: Top Guns Confiscations

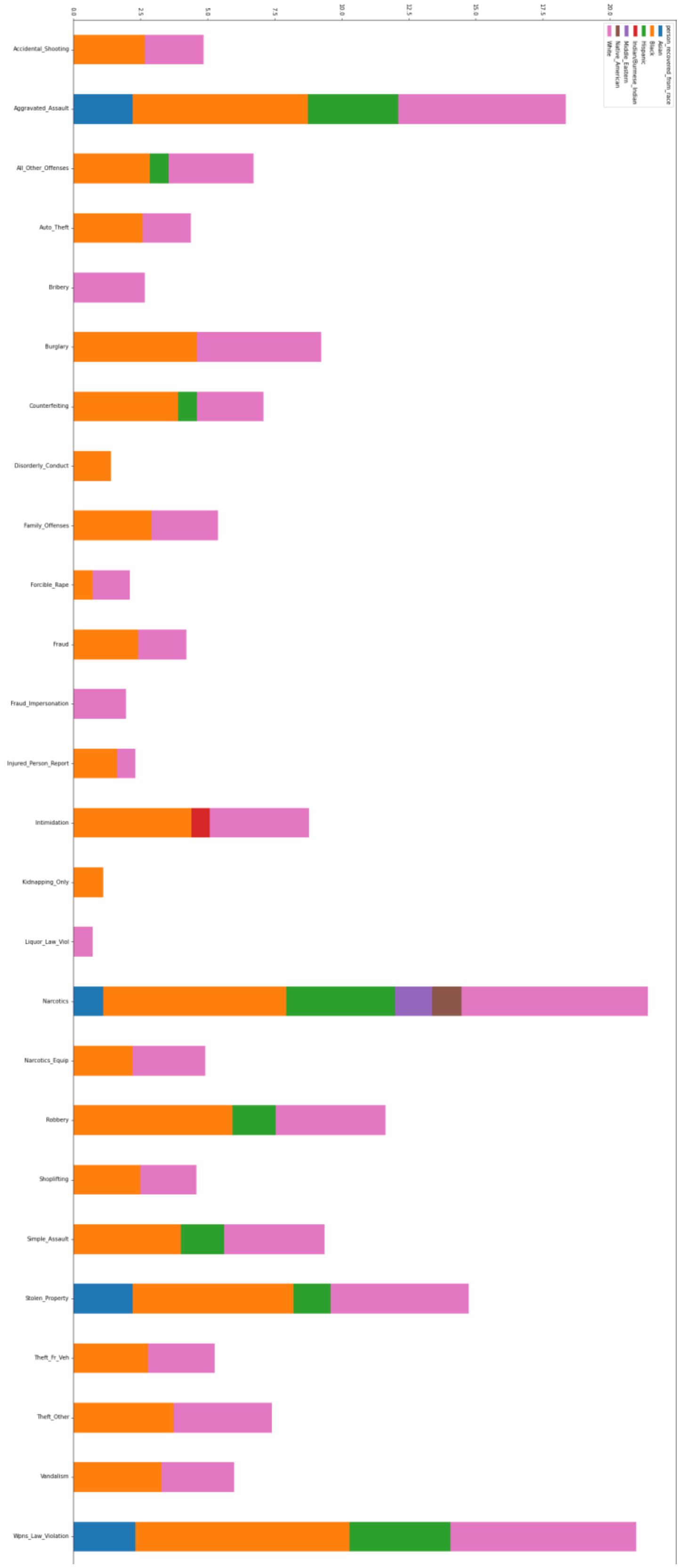


Figure 6: Crime Categories Divided by Race (Stacked)

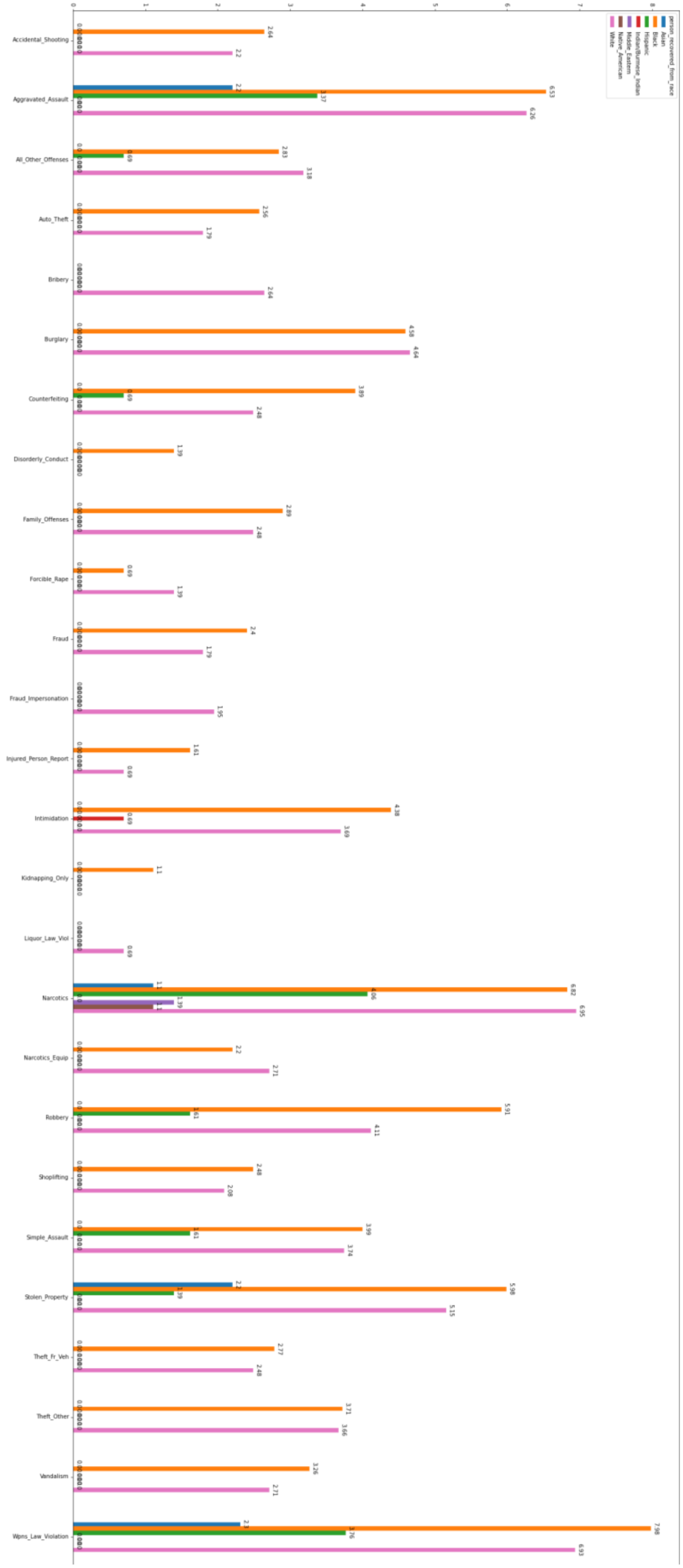


Figure 7: Crime Categories Divided by Race

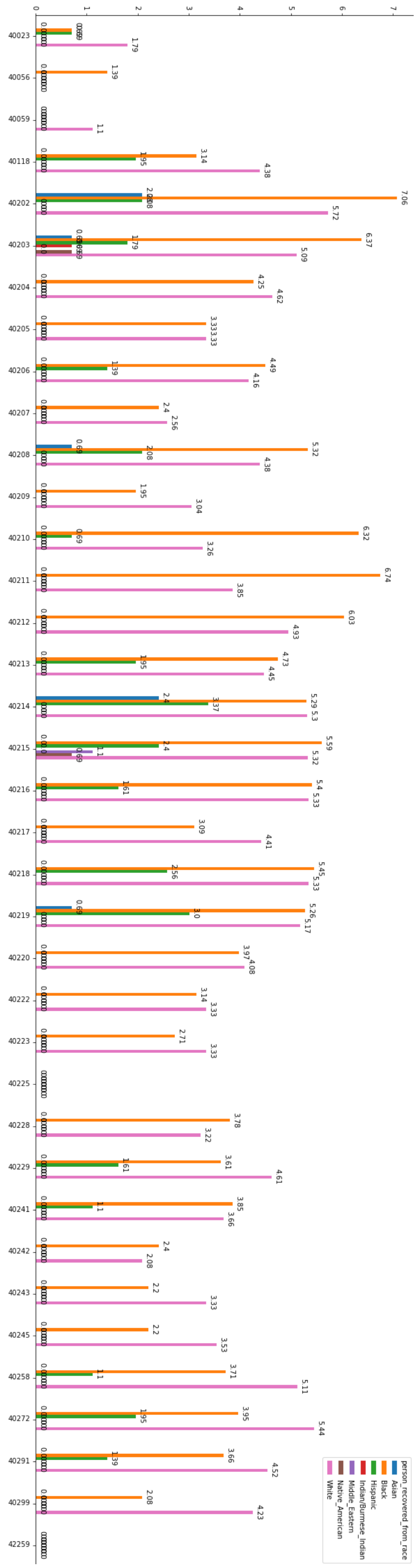


Figure 9: Firearms Confiscated by Zipcode/Race

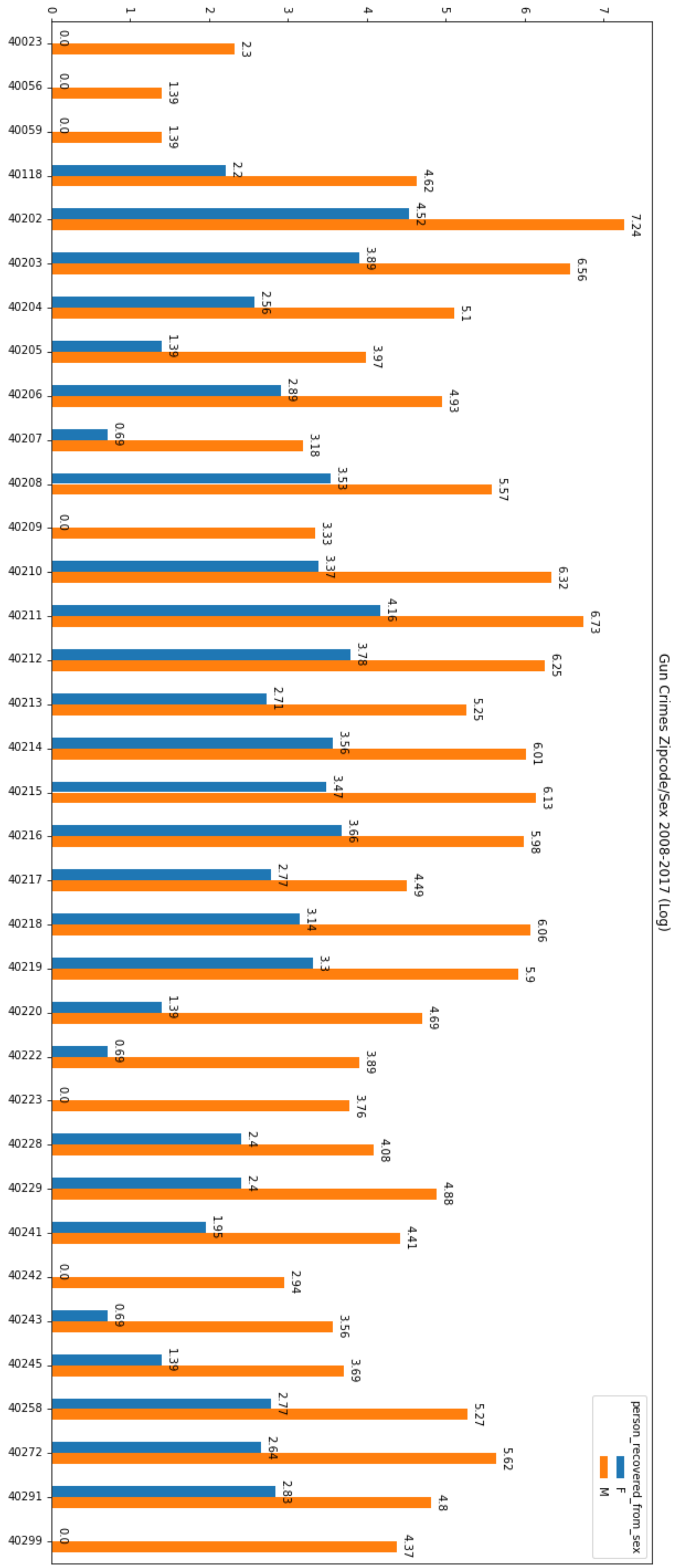


Figure 10: Confiscated Firearms Zipcode/Sex

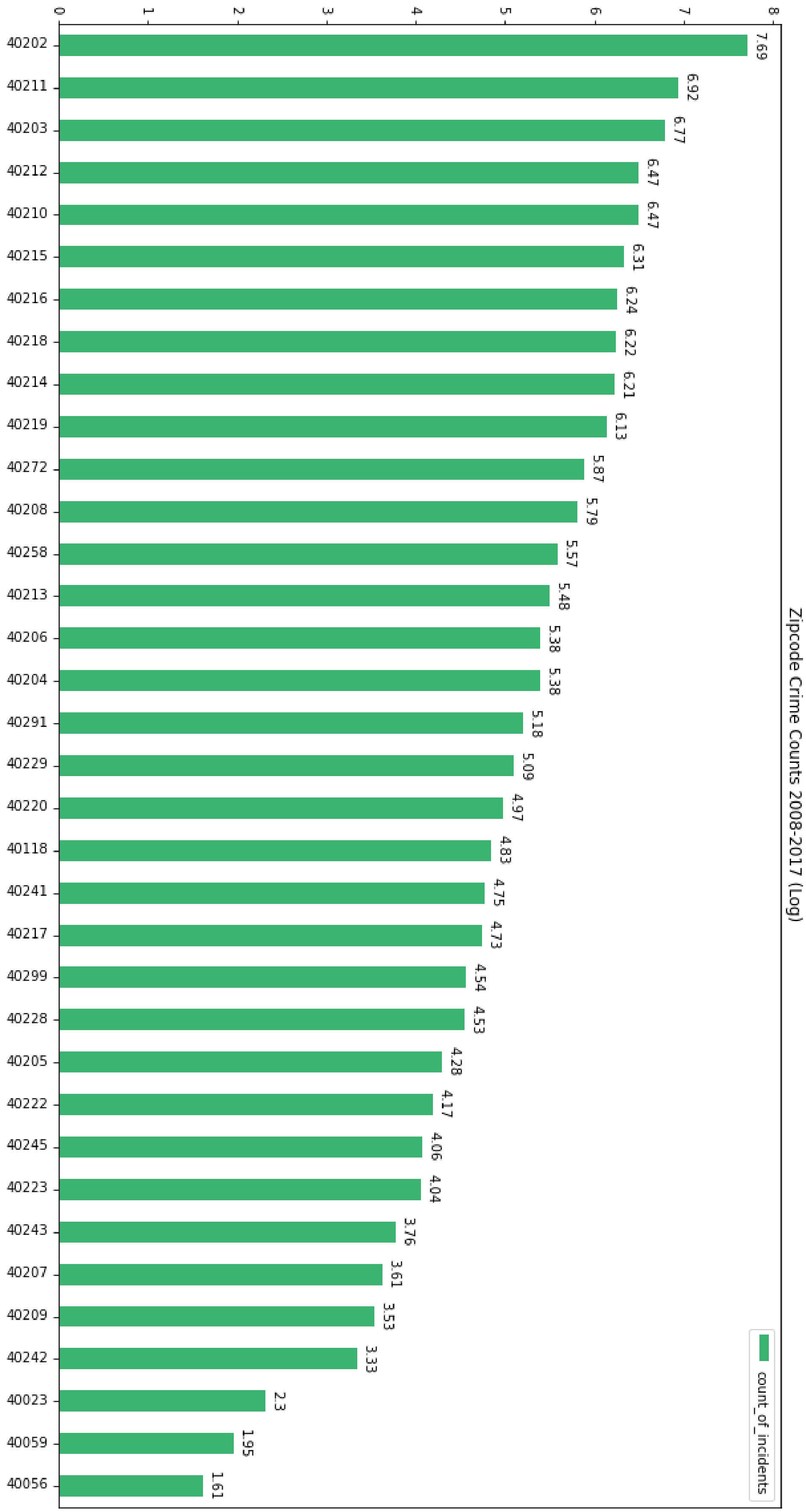


Figure 12: Firearms Confiscated by Zipcode

C. *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.**

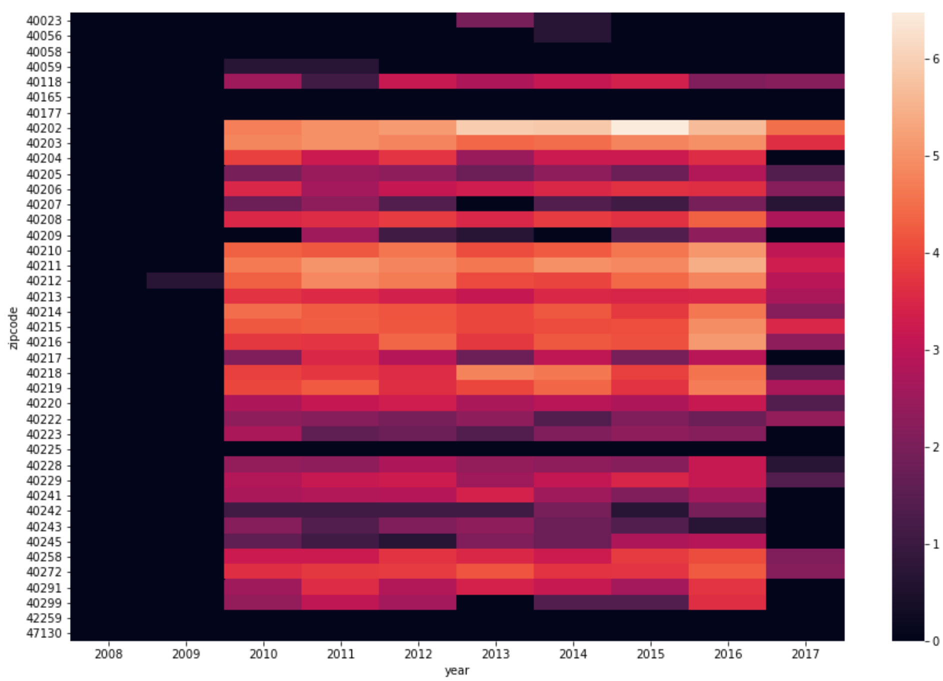


Figure 14: Heatmap of Gun Confiscations Year/Zipcode

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.

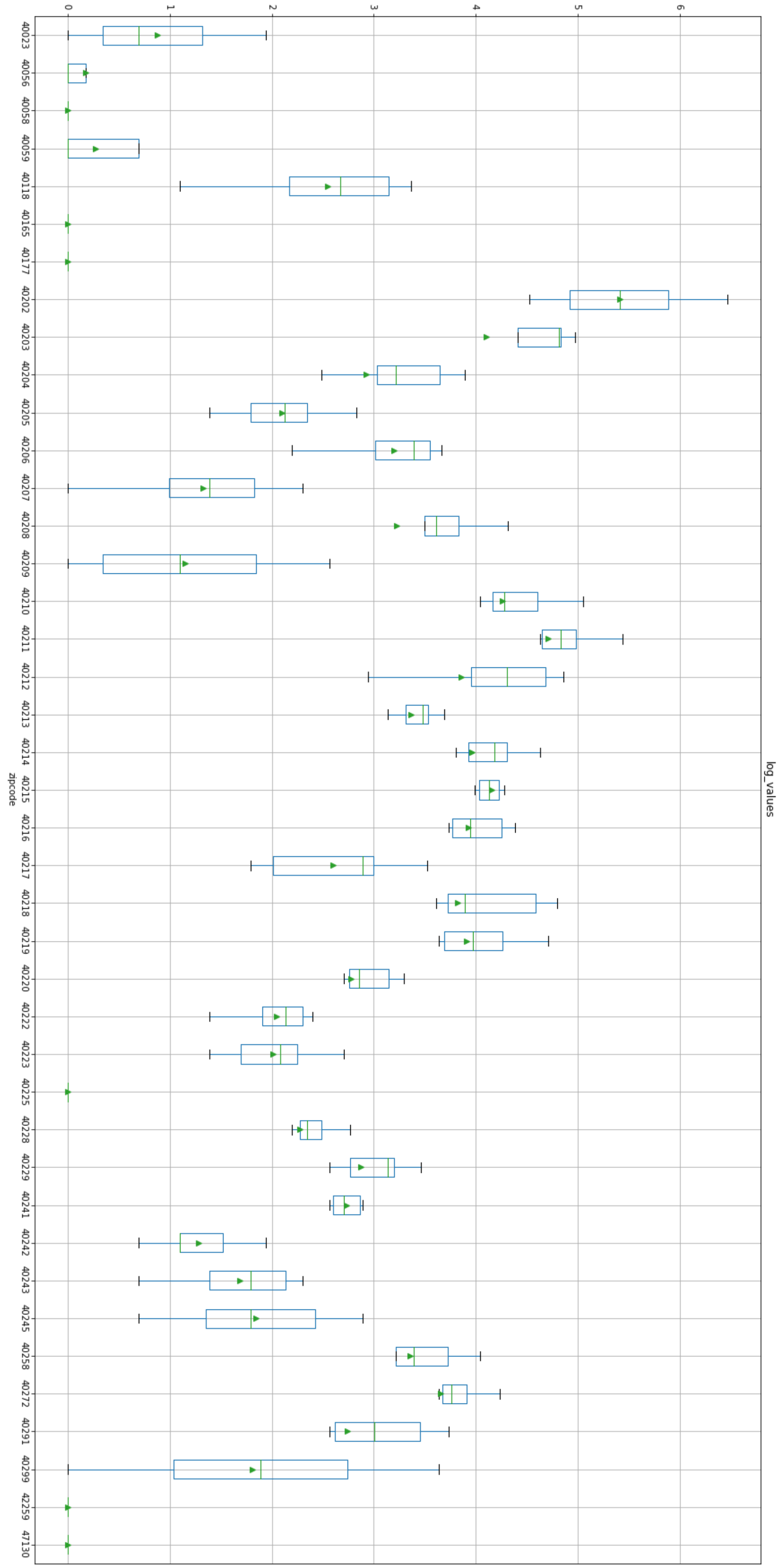


Figure 13: Boxplot of Zipcode/Confiscations Across 2008-2017

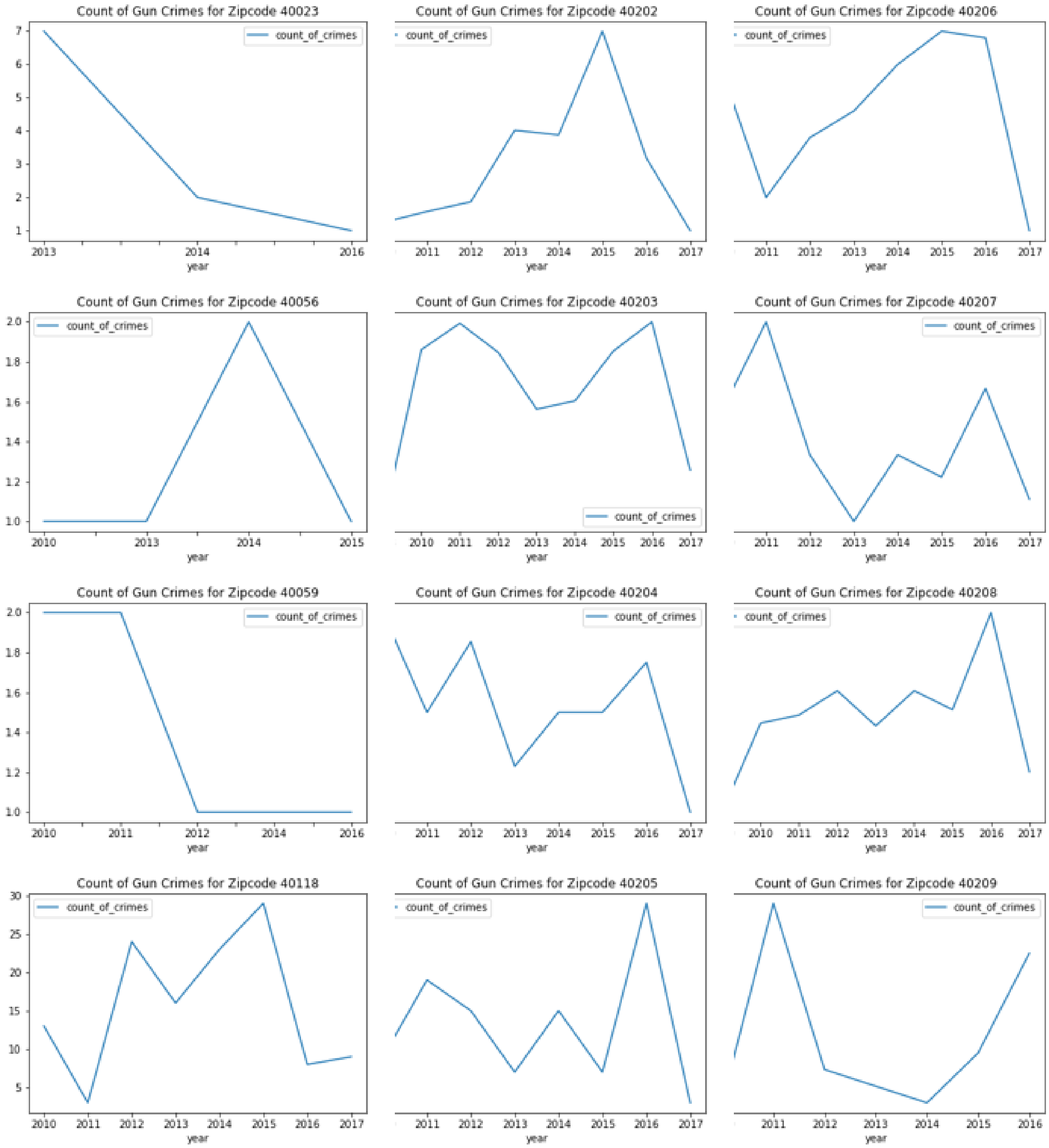


Figure 15: Zip Code Gun Confiscation Trends 2008-2017 (1)

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

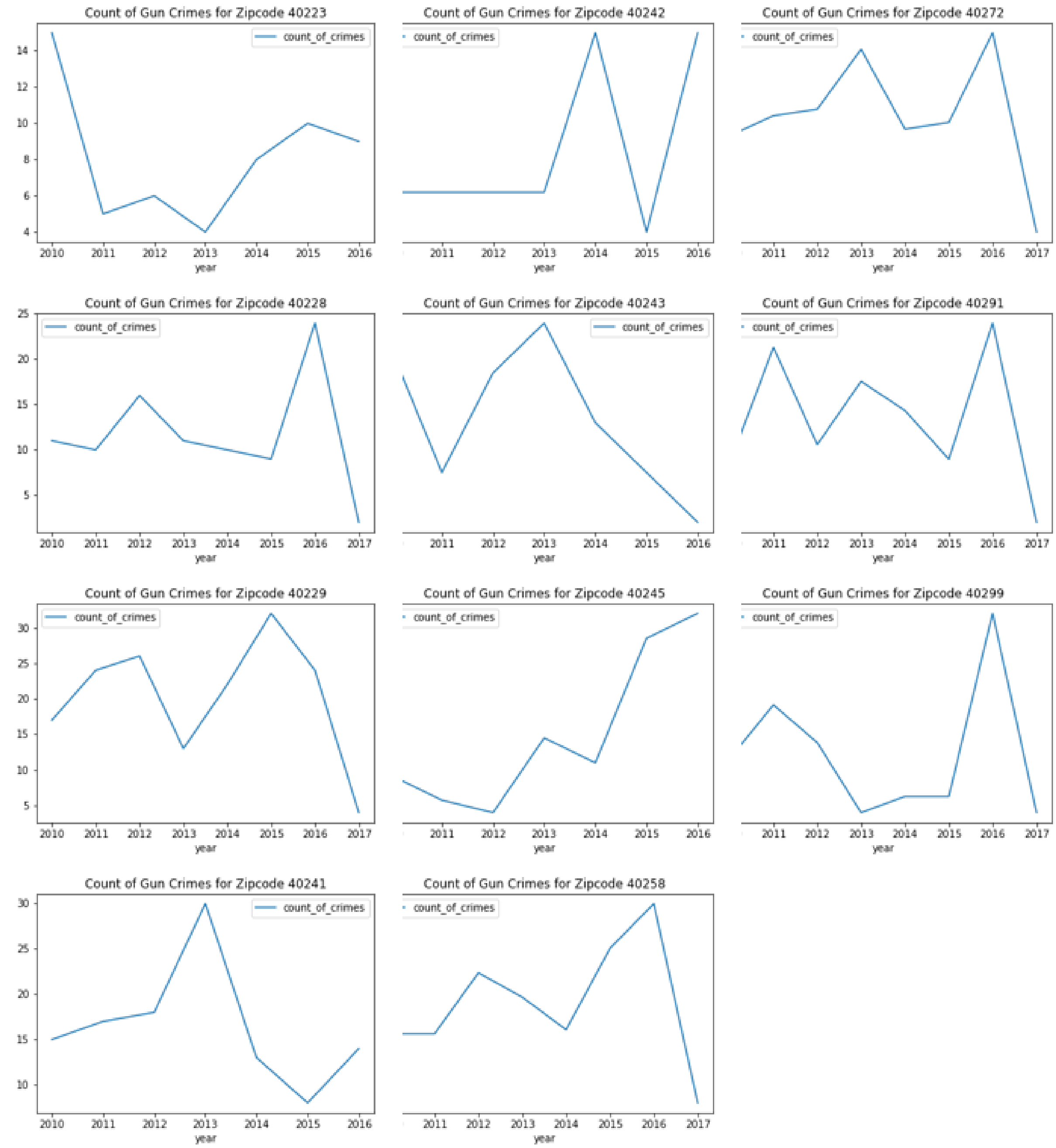


Figure 17: Zip Code Gun Confiscation Trends 2008-2017 (3)

V. 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.

VI. APPENDIX

A. 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.

i. *Incident Number*
https://github.com/cporterbellarmine/IP4_CNP/blob/main/unique_val_df_incident_number.csv

ii. *Firearm Manufacturer*
https://github.com/cporterbellarmine/IP4_CNP/blob/main/unique_val_df_firearm_manufacturer.csv

iii. *Firearm Model*
https://github.com/cporterbellarmine/IP4_CNP/blob/main/unique_val_df_firearm_model.csv

iv. *Firearm Caliber*
https://github.com/cporterbellarmine/IP4_CNP/blob/main/unique_val_df_firearm_caliber.csv

v. *Block Address*
https://github.com/cporterbellarmine/IP4_CNP/blob/main/unique_val_df_block_address.csv

vi. *Street Address*
https://github.com/cporterbellarmine/IP4_CNP/blob/main/unique_val_df_street_address.csv

B. This project with ACLU is still in progress. To see further progress, you can visit:
   i. https://github.com/cporterbellarmine/ButterflyProject

C. Other partners for this project include:
   i. Keturah Herron from ACLU
   ii. Dr. Rob Kelley from Bellarmine University
   iii. Dr. Monica Unseld from Microsoft Future of Work
   iv. Alisia McClain from Microsoft Future of Work
   v. Dr. Jakia Marie from Bellarmine University
   vi. Cassaria Edwards from Central High School
   vii. Eva Parke from Bellarmine University.