

CMDA-3654

EDA for Course Project

Bekalu Debelu

Due as a .pdf upload

Abstract

The purpose of this project is to examine the incident reports filed in San Francisco and analyze vehicle related crimes in the city. I focused on variables related to types of Incidents reported, the data of the reporting, and from which police districts the reports originated.

Analysis 1: Vehicle Related Incidents Reports Compared to Other Categories

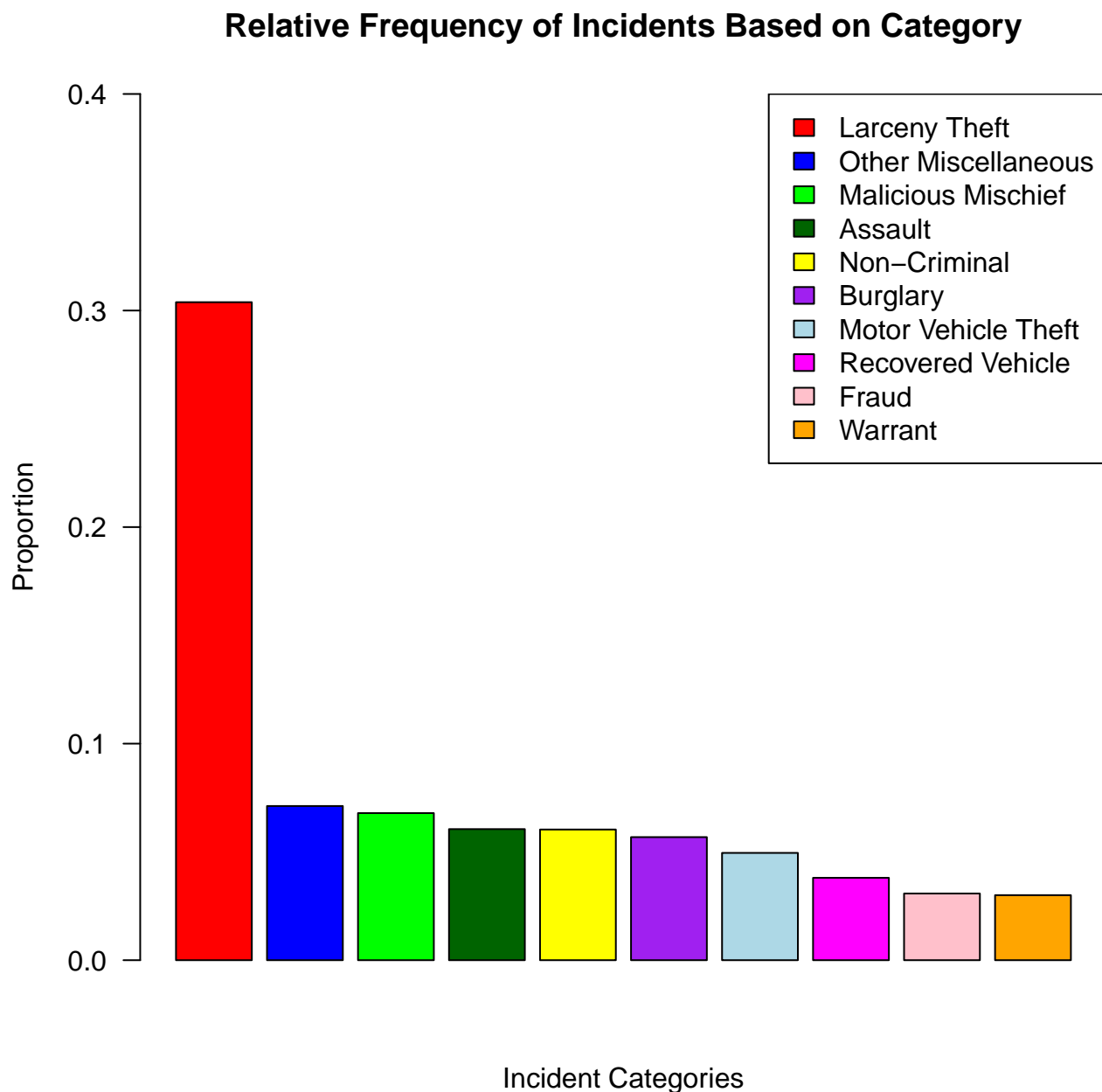


Figure 1: Plot of Frequency of Incident Reports based on Categories

From the graph we can see that the most frequent crime reported is Larceny Theft with Motor Vehicle Theft as the seventh most reported incident. By exploring the sub categories of Larceny related incident reports, we can see that larceny involves theft from vehicles which falls under our area of interest.

Analysis 2: Examining Larceny related reports

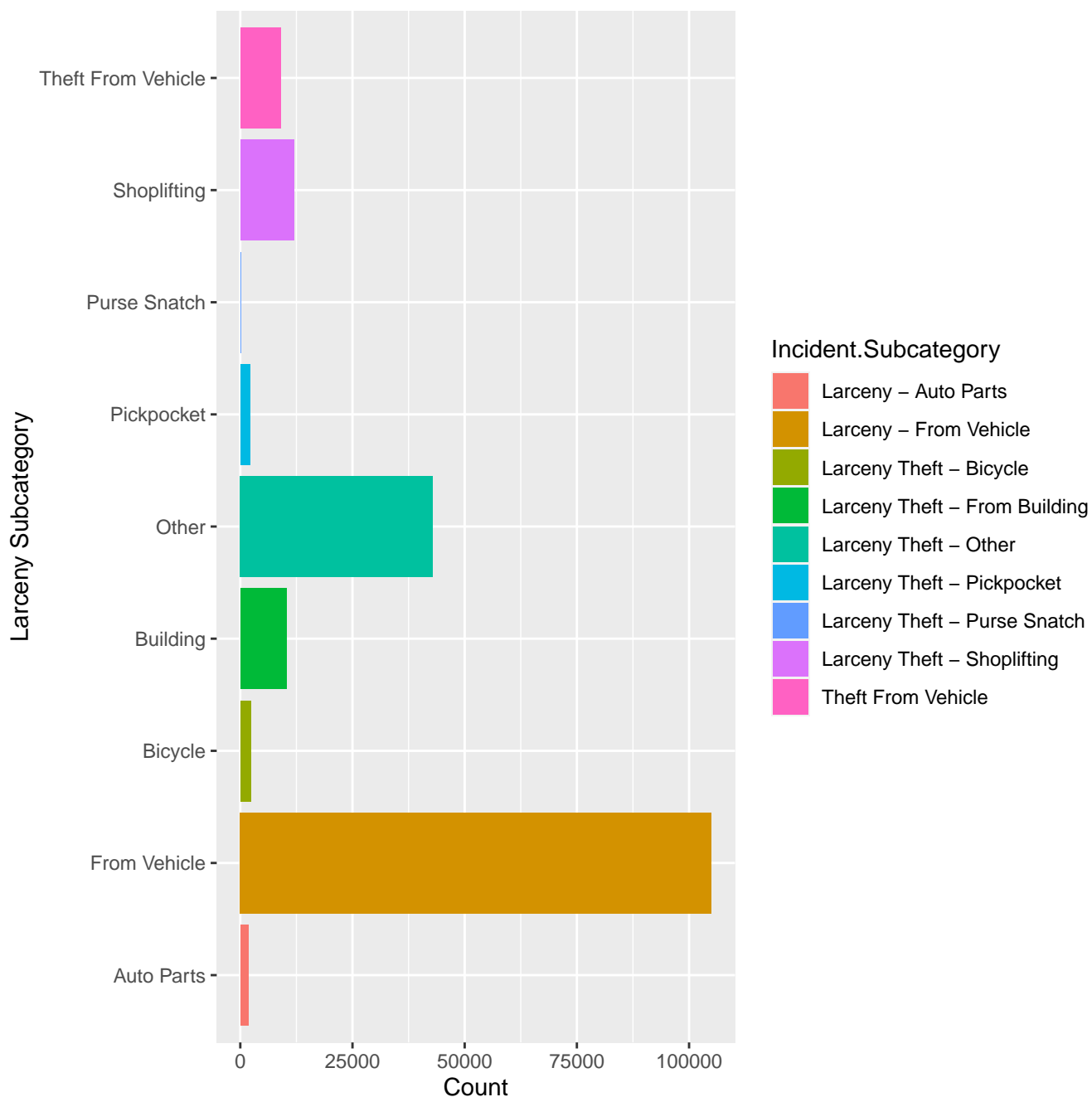


Figure 2: Barplot of Larceny Subcategories Based on Frequency

From the bar chart we can view that the majority of larceny related reports are vehicle break-ins including a lower portion of these crimes being incidents where parts from the vehicle itself were stolen.

Analysis 3: Count of vehicle related reports across the months

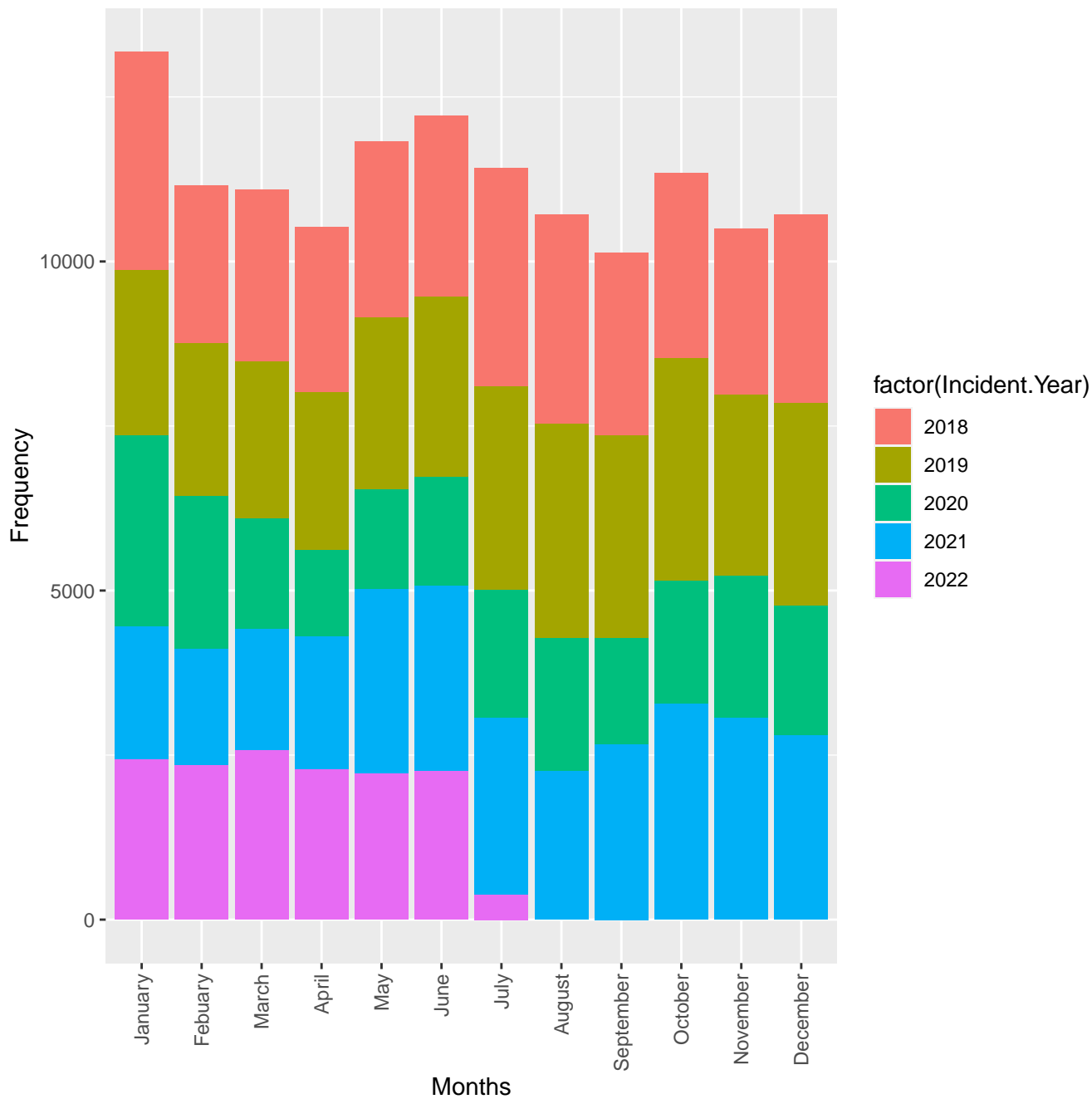


Figure 3: Plot of Months versus Frequency

Figure 3 shows us that vehicle reports are filed at nearly the same volume every year. This shows that we need to narrow our search and look by weeks instead. Find the average incidents that reported on each weekday and see if there are spikes when those days fall on holidays and city-wide events.

Analysis 4: Radar chart of Police Districts Based on Incidents Reported

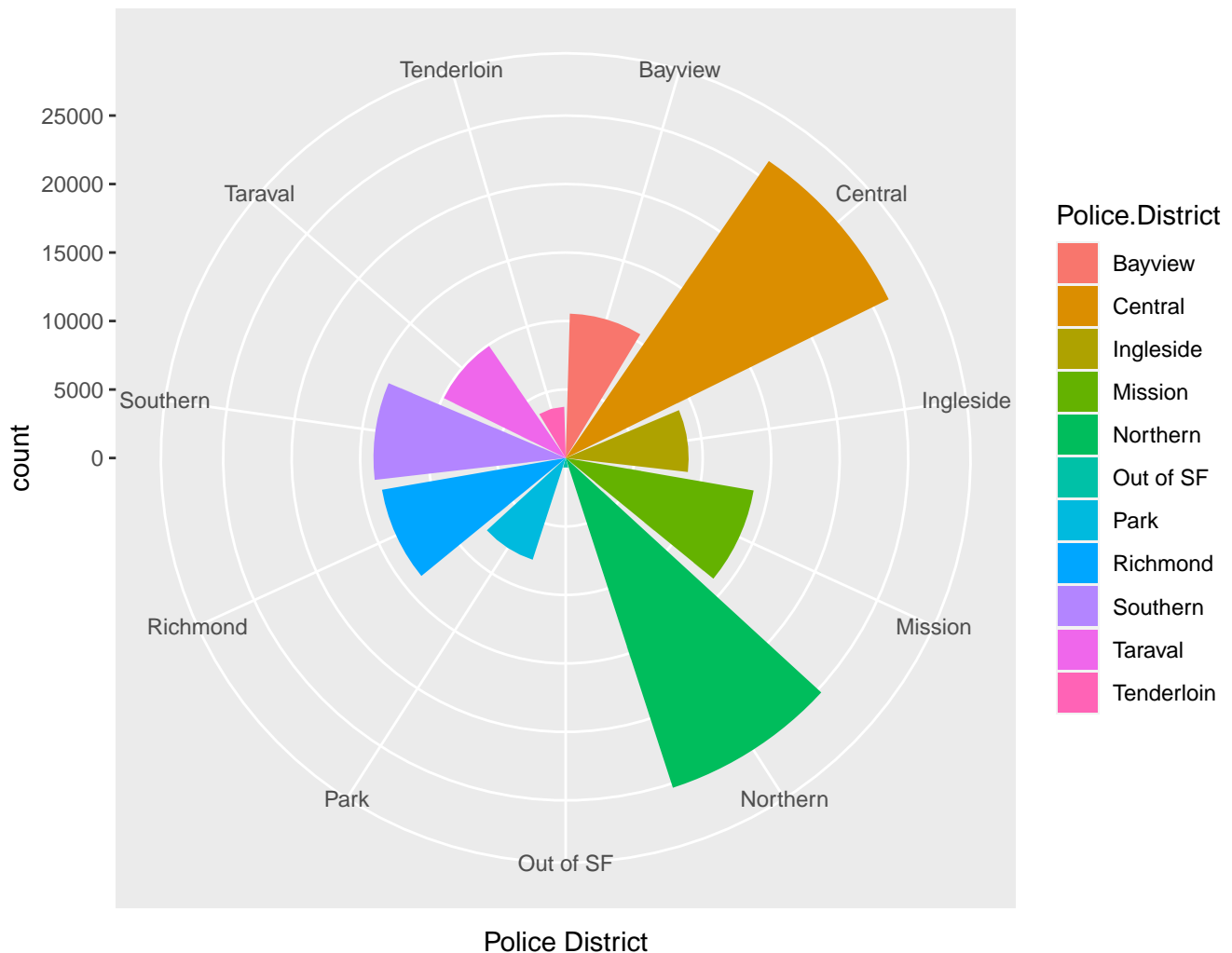


Figure 4: Plot of Incident Reports Based on Police Districts

Figure 4 gives us the distribution of reports across the various police districts in San Francisco. Our initial investigation shows that the districts in Central and Northern have larger number of reports compared to other districts. Our next steps could be to look into these districts and see if they are residential neighborhoods, industrial spaces, etc.

Appendix

4.1 Code for Analysis 1

```
#Count the instances of incident reports based on categories
reports_by_category = table(police_reports$Incident.Category)

proportional_table = reports_by_category / sum(reports_by_category)
```

```
#Pick the ten most frequently occurring incident reports
top_10_reports = order(proportional_table, decreasing = T)[1:10]

palette(c("red","blue","green","darkgreen","yellow","purple","lightblue","magenta","pink","orange"))

barplot(proportional_table[top_10_reports], col = c(1:10), ylim = c(0,0.40), las = 2, cex.names = 0.7,
        xaxt="n", ylab = "Proportion",
        xlab = "Incident Categories",
        main = "Relative Frequency of Incidents Based on Category")

legend("topright", legend = row.names(proportional_table)[top_10_reports], fill = c(1:10))
```

4.2 Code for Analysis 2

```
#Separate larceny related reports to investigate further
larceny = subset(police_reports,
                 police_reports$Incident.Category == "Larceny Theft")

ggplot(larceny) + geom_bar(aes(x = Incident.Subcategory, fill = Incident.Subcategory)) +
scale_x_discrete(labels=c("Auto Parts", "From Vehicle", "Bicycle","Building",
                          "Other","Pickpocket", "Purse Snatch","Shoplifting",
                          "Theft From Vehicle")) +

coord_flip() +
labs(x = "Larceny Subcategory", y = "Count",
     main = "Frequency of Larceny Subcategories")
```

4.3 Code for Analysis 3

```
vehicle = subset(police_reports, police_reports$Incident.Subcategory == "Larceny - From Vehicle" | police_reports$Incident.Subcategory == "Larceny - From Vehicle")

#First we have to extract the months from the Incident.Date variable and append it as its own variable to the data
Incident.Month <- as.Date(vehicle$Incident.Date, format = "%m/%d/%Y")
Incident.Month <- format(Incident.Month, "%m")
vehicle$Incident.Month = Incident.Month

ggplot(vehicle) +
  geom_bar(mapping = aes(x = Incident.Month,
                        fill = factor(Incident.Year))) +
  labs(x = "Months", y = "Frequency",
       main = "Frequency of Incident ") +
  scale_x_discrete(labels=c("January","February","March","April","May","June",
                            "July","August","September","October","November",
                            "December")) +
  guides(x = guide_axis(angle = 90))
```

4.4 Code for Analysis 4

```
ggplot(vehicle) +
  geom_bar(mapping = aes(x = Police.District, fill = Police.District)) +
  coord_polar() +
  labs(main = "Plot of Incident Reports Based on Police Districts",
       x = "Police District")
```