Robbery in Central San Francisco during the Summer of 2014

Dan Kirlin

December 29, 2015

Objective

In this project, we will be examining Crime in San Francisco from the Summer of 2014 using the `R` programming language and `ggplot2` package. In our investigation we will show several trends for San Francisco during this time period:

1. Theft is the most common crime type.
2. The Central District, although not a top 3 district for crime overall, is top 2 in theft and/or robberies.
3. There is no apparent trend towards Days for this type of crime in the Central District.

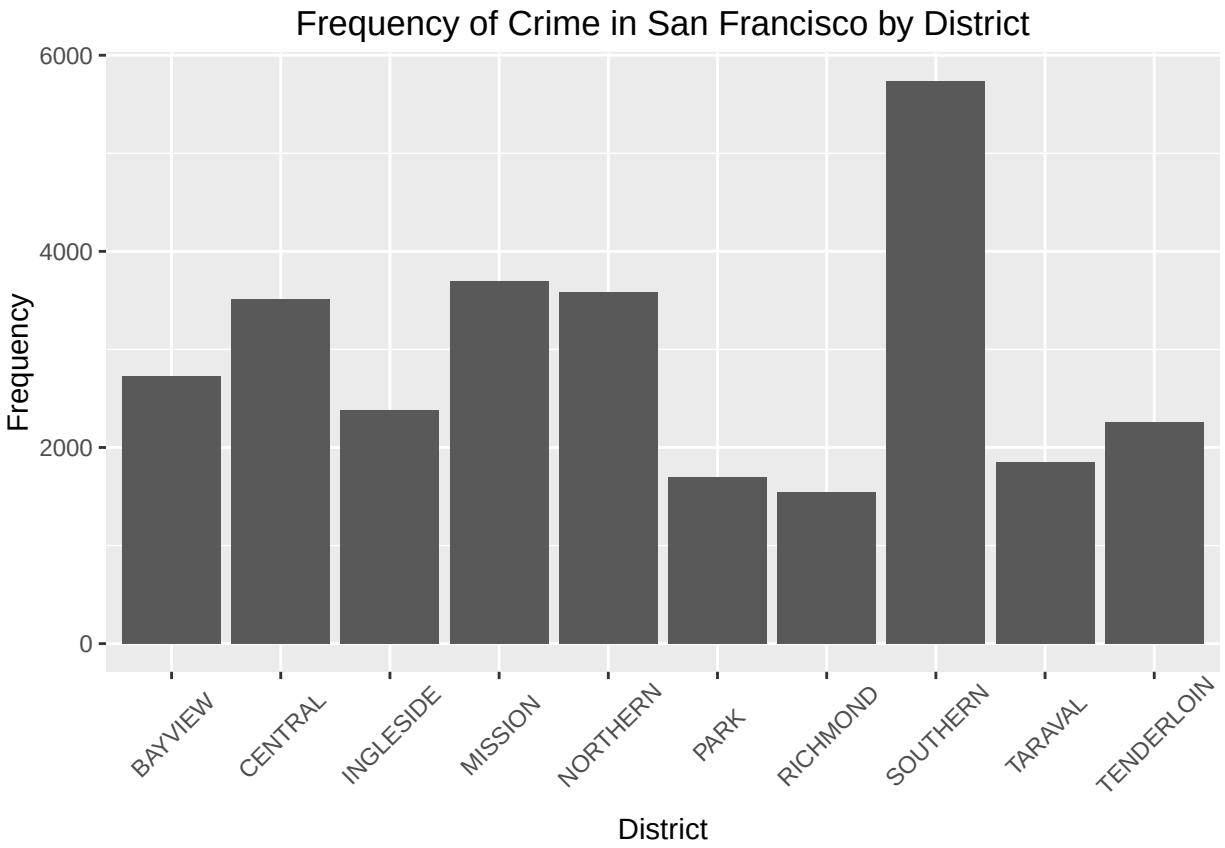
Paper

Let's begin by loading the ggplot2 package, dplyr (data manipulation), and the provided data.

```
require(ggplot2);require(dplyr);
sanfrancisco <- read.csv("sanfrancisco_incidents_summer_2014.csv")
```

For our first graph, let's examine how Crime is dispersed throughout the various districts.

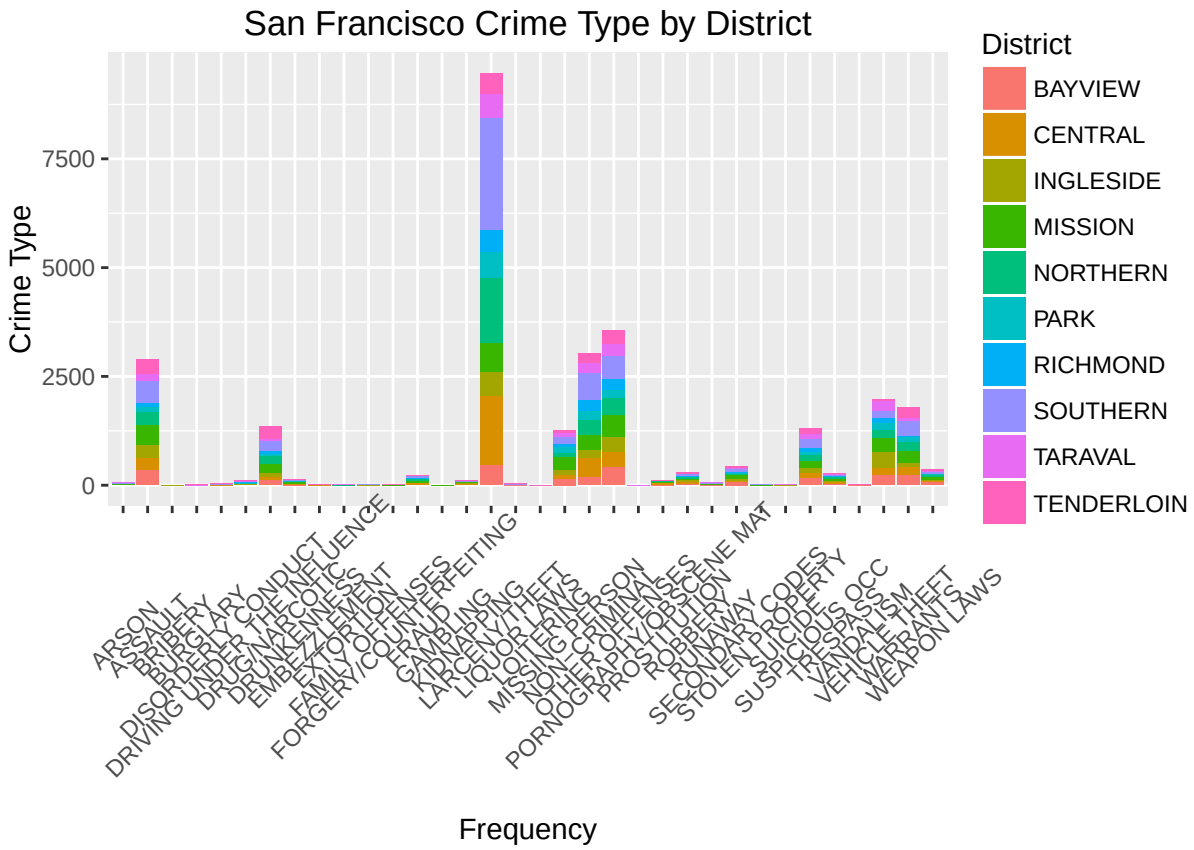
```
plot_neighborhood <- ggplot(sanfrancisco, aes(x = PdDistrict))
plot_neighborhood <- plot_neighborhood + geom_bar()
plot_neighborhood <- plot_neighborhood + labs(y = "Frequency", x = "District")
plot_neighborhood <- plot_neighborhood + theme(axis.text.x=element_text(angle=45,vjust=.6))
plot_neighborhood <- plot_neighborhood + ggtitle("Frequency of Crime in San Francisco by District")
plot_neighborhood
```



Here we see that the Southern District has nearly double that of the next highest (Central, Mission, and Northern).

Let's now examine this from a different angle - what type of crimes are taking place in each district?

```
plot <- ggplot(sanfrancisco, aes(x = Category, fill=PdDistrict))
plot <- plot + theme(axis.text.x=element_text(angle=45,vjust=.6))
plot <- plot + geom_bar()
plot <- plot + xlab("Frequency")
plot <- plot + ylab("Crime Type")
plot <- plot + ggtitle("San Francisco Crime Type by District")
plot <- plot + scale_fill_discrete(name = "District")
plot
```

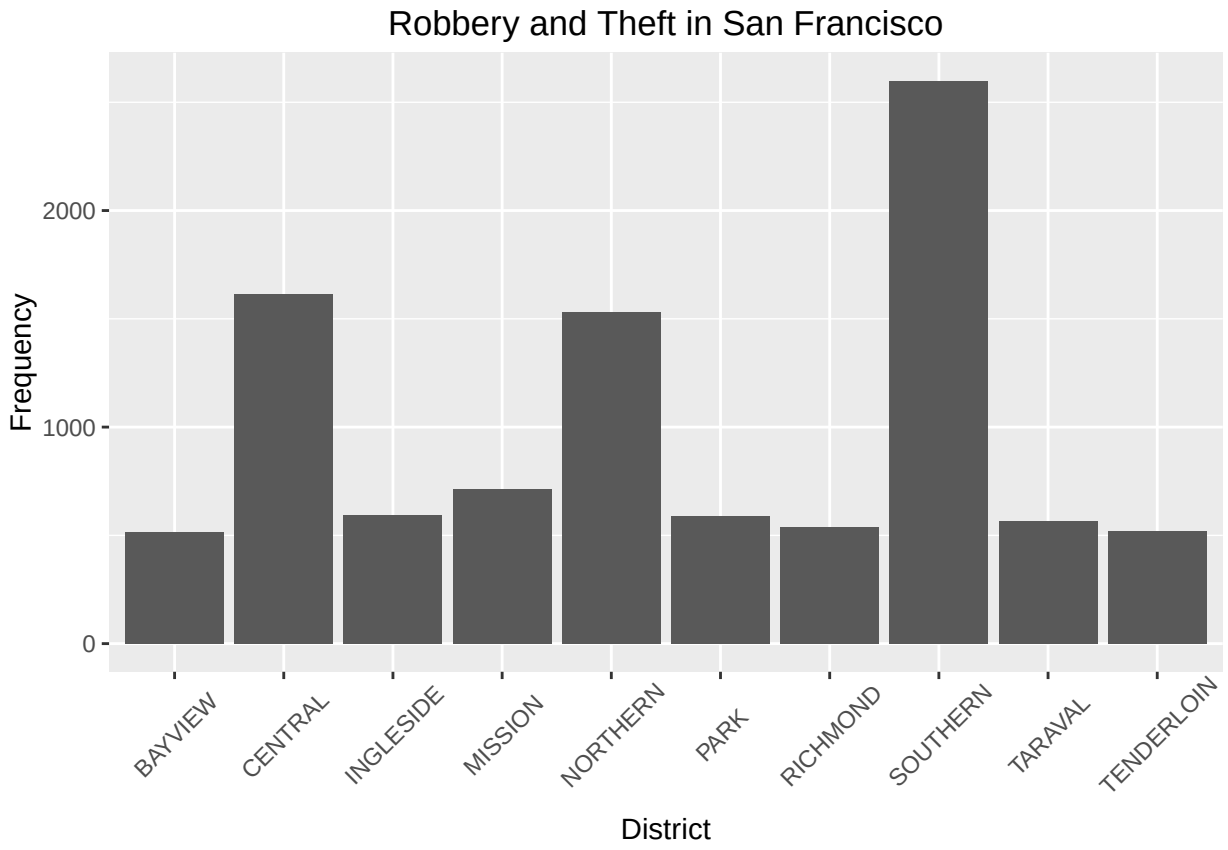


The most common crime by nearly double is Larceny/Theft. This appears to happen most frequently in the Central and Southern district.

Robbery Examination

From a purely Robbery/Theft perspective we see that central is only second to that of southern. Which when compares to only 4th overall for crime of the districts, is a significant jump.

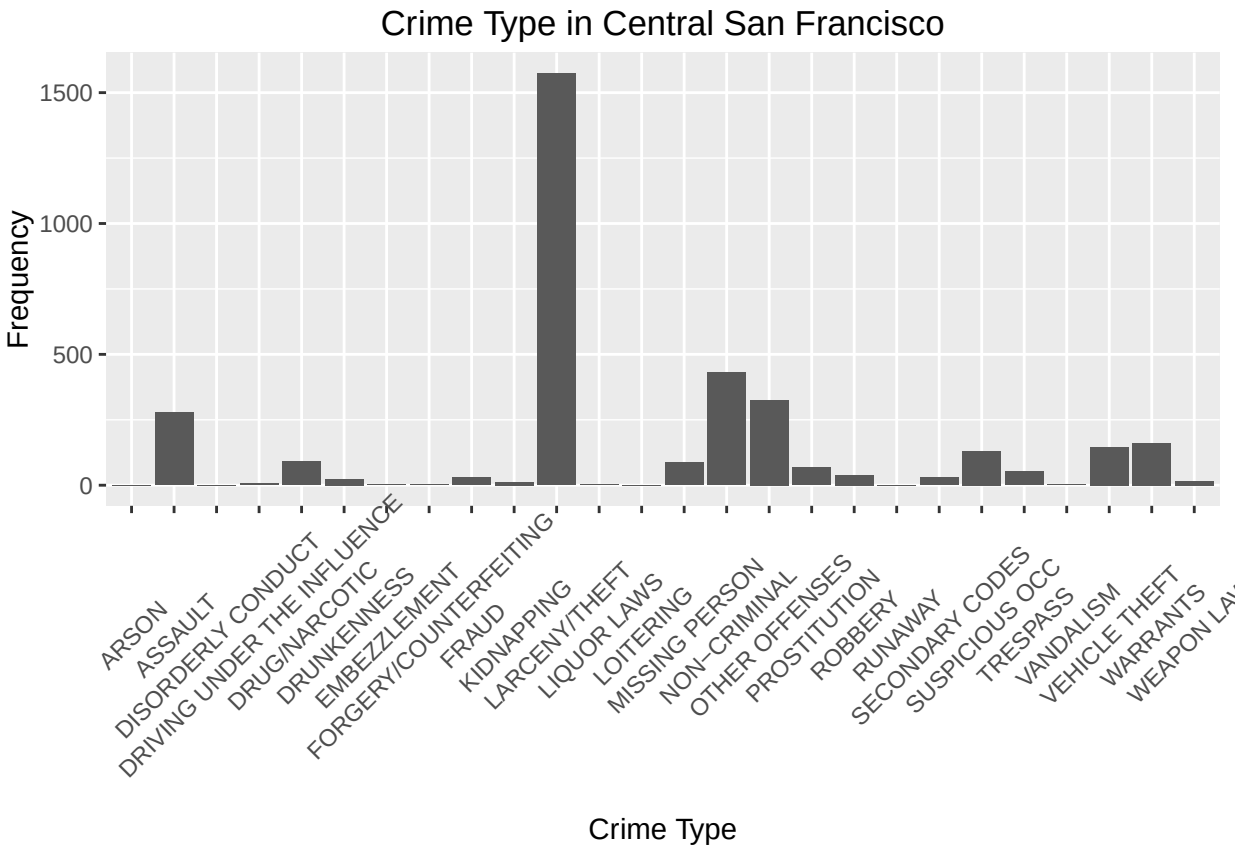
```
sanfrancisco_theft <- filter(sanfrancisco, Category == 'ROBBERY' | Category == 'LARCENY/THEFT')
plot_theft <- ggplot(sanfrancisco_theft, aes(x=PdDistrict))
plot_theft <- plot_theft + geom_bar()
plot_theft <- plot_theft + ggtitle("Robbery and Theft in San Francisco")
plot_theft <- plot_theft + labs(x = "District", y = "Frequency")
plot_theft <- plot_theft + theme(axis.text.x=element_text(angle=45,vjust=.6))
plot_theft
```



Central District Examination

If we narrow our scope to only the Central District exclusively, we can see several trends

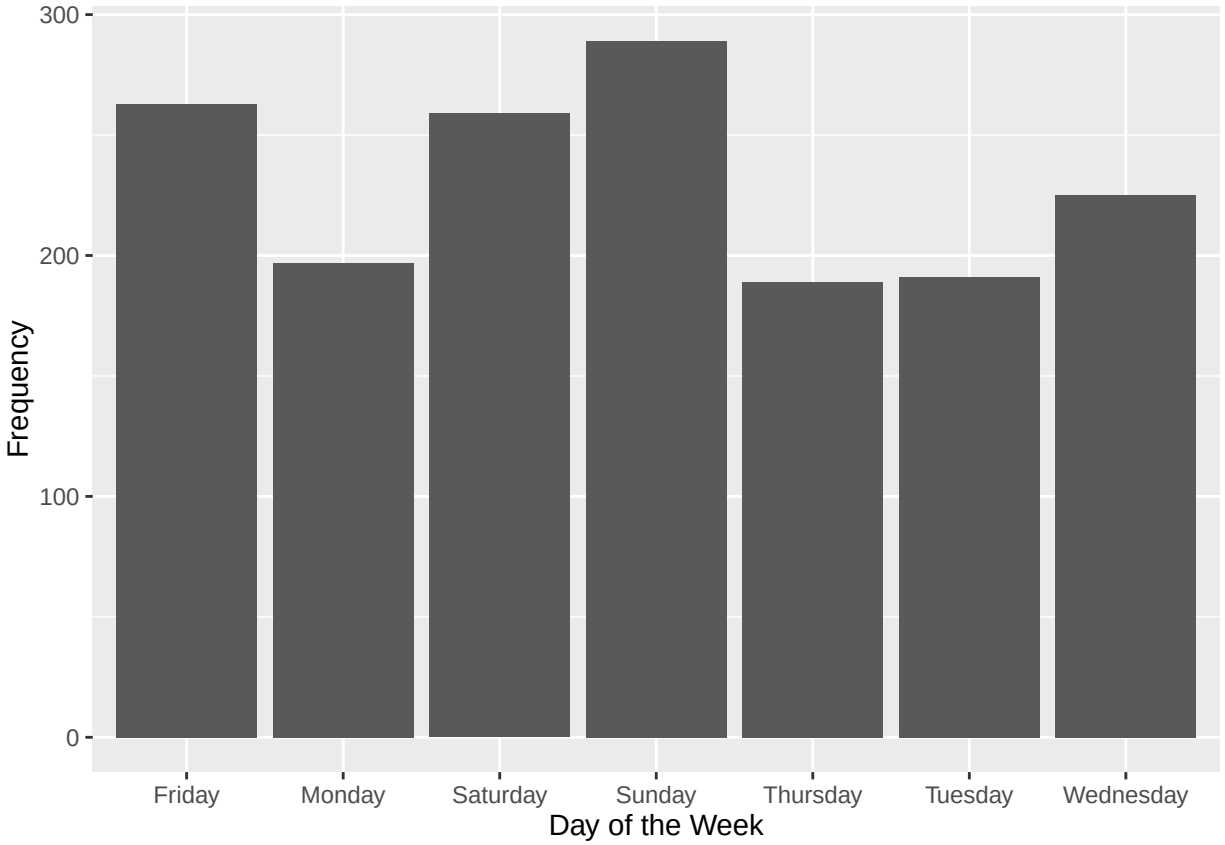
```
sanfrancisco_center <- filter(sanfrancisco, PdDistrict=="CENTRAL")
plot_incidents <- ggplot(sanfrancisco_center, aes(x=Category))
plot_incidents <- plot_incidents + geom_bar()
plot_incidents <- plot_incidents + theme(axis.text.x=element_text(angle=45,vjust=.6))
plot_incidents <- plot_incidents + labs(x = "Crime Type", y = "Frequency")
plot_incidents <- plot_incidents + ggtitle("Crime Type in Central San Francisco")
plot_incidents
```



For one - Larceny and Theft is even more strongly represented in this district (nearly quadruple the next crime type of "non-criminal").

Let's now attempt to see if there are any trends related to Day of the Week.

```
sanfrancisco_central_theft <- filter(sanfrancisco_center, Category == 'ROBBERY' | Category == 'LARCENY/'
plot_center_theft <- ggplot(sanfrancisco_central_theft, aes(x=DayOfWeek))
plot_center_theft <- plot_center_theft + geom_bar()
plot_center_theft <- plot_center_theft + labs(x="Day of the Week", y="Frequency")
plot_center_theft
```



Unfortunately - from the following graph we can see no trend in robbery to days of the week. One would asssume that Weekend nights would be a hot bed for this kind of activity however, data never lies. (A good point for a future project would be to visualize this data as a time series.)

Conclusion

In conclusion - we can see that Robbery is largely represented in the Central District as a percentage of it's crime, is the second overall for this type of crime among all districts and appears to have no trends with days of the week.