Robbery in Central San Francisco

Dan Kirlin
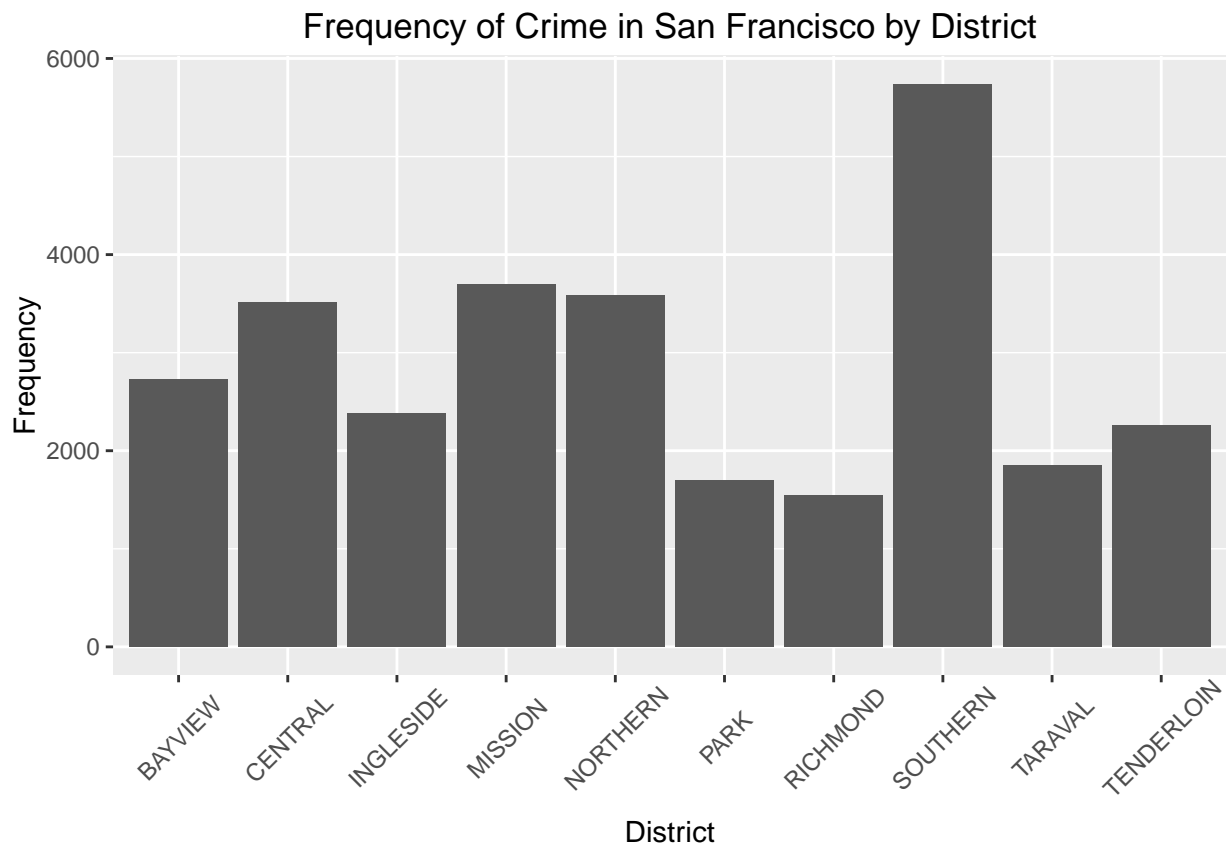
December 29, 2015

In this project, we will be examining Crime in San Francisco using the `R` programming language and `ggplot2` package. Let's begin by loading the ggplot2 package, dplyr (data manipulation), and the provided data.

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

```r
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
```
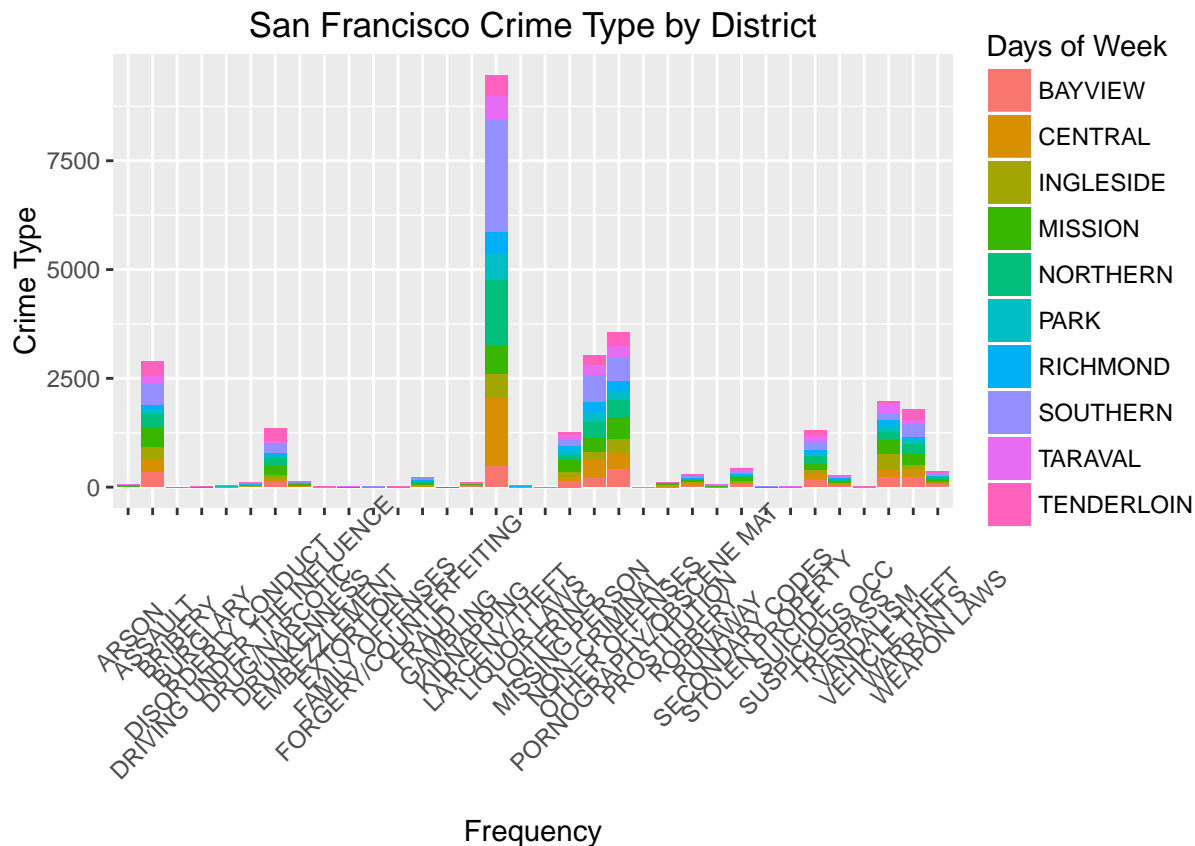

The most common crime by nearly double is Larceny/Theft. This appears to happen most frequently in the Central and Southern district.

```
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 = "Days of Week")
plot
```
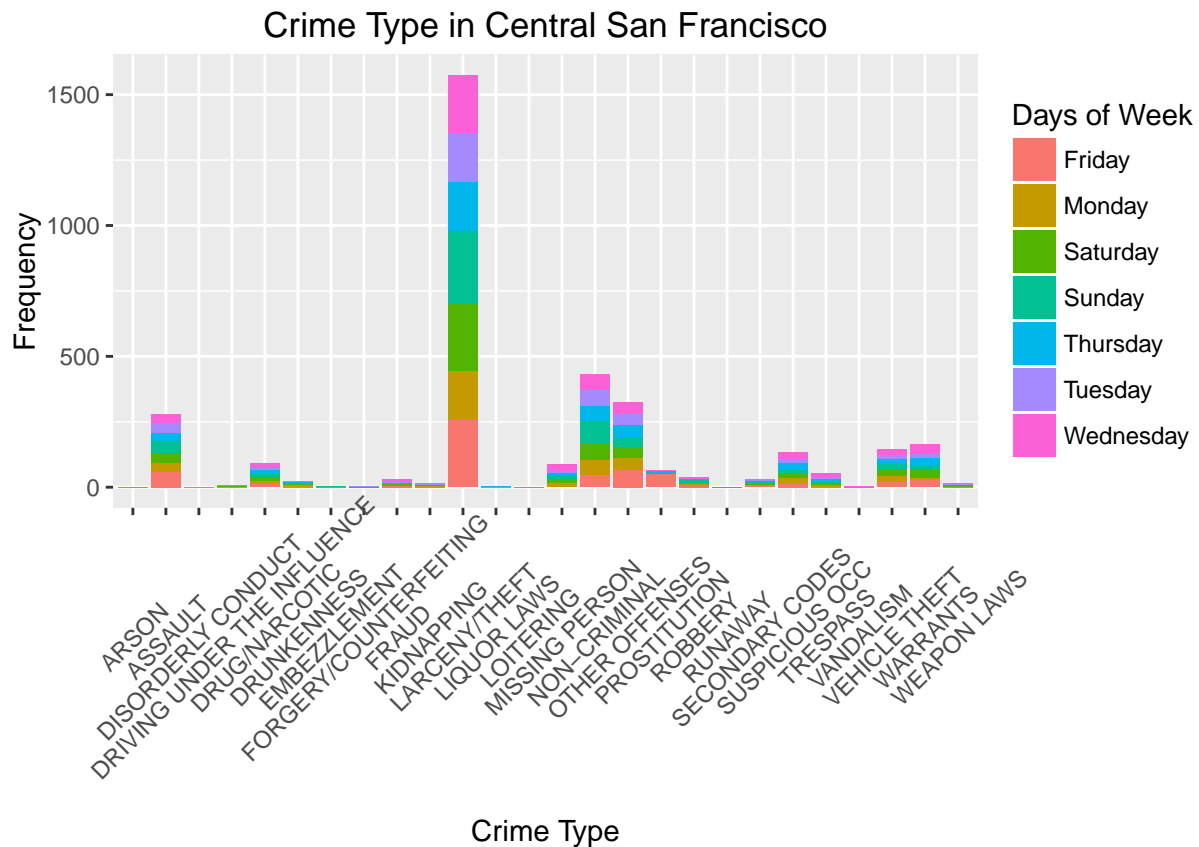

In Central San Francisco we can see that Larceny and Theft is even more strongly represented in this district (nearly quadruple the next crime type of "non-criminal").
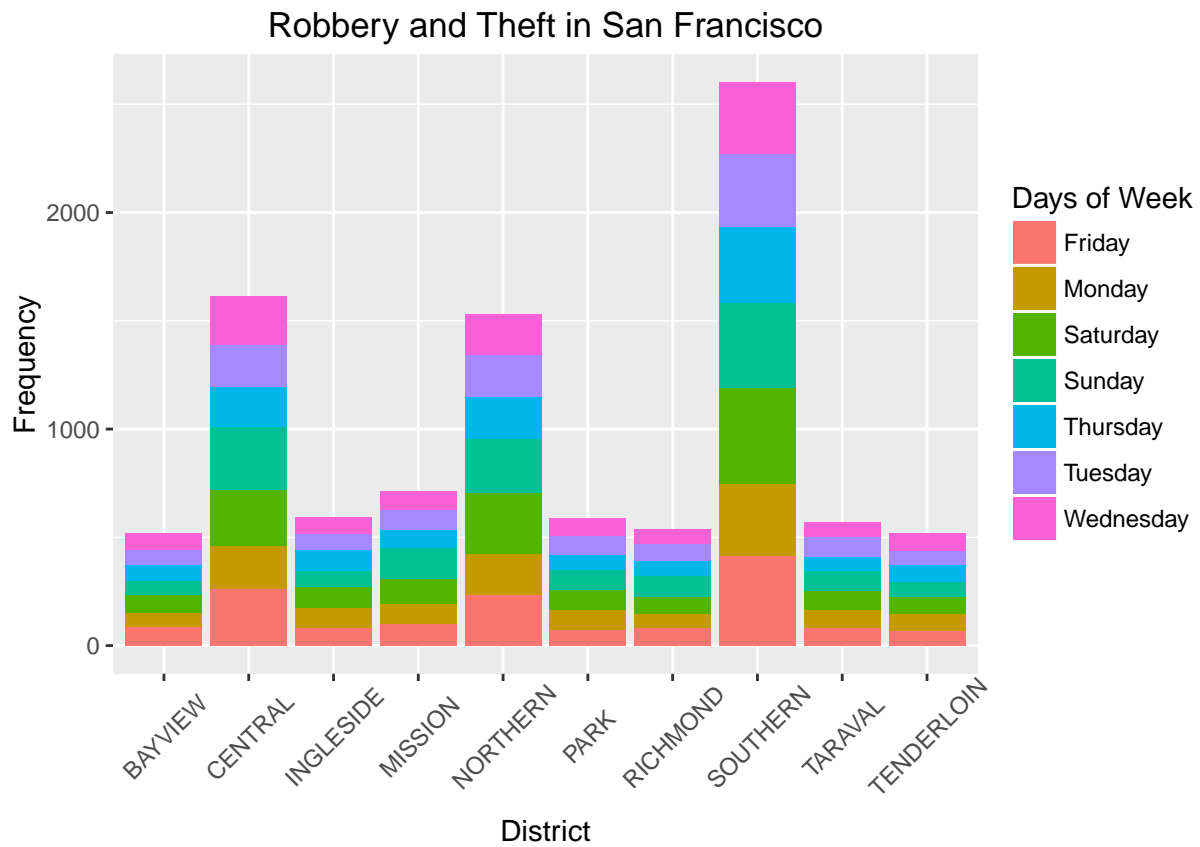
```
sanfrancisco_center <- filter(sanfrancisco, PdDistrict=="CENTRAL")
plot_incidents <- ggplot(sanfrancisco_center, aes(x=Category, fill=DayOfWeek))
plot_incidents <- plot_incidents + geom_bar()
plot_incidents <- plot_incidents + theme(axis.text.x=element_text(angle=45,vjust=.6))
plot_incidents <- plot_incidents + labs(x = "Crime Type", y = "Frequency")
plot_incidents <- plot_incidents + ggtitle("Crime Type in Central San Francisco")
plot_incidents <- plot_incidents + scale_fill_discrete(name = "Days of Week")
plot_incidents
```

Crime Type in Central San Francisco

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, fill=DayOfWeek))
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 + scale_fill_discrete(name = "Days of Week")
plot_theft <- plot_theft + theme(axis.text.x=element_text(angle=45,vjust=.6))
plot_theft
```

Robbery and Theft in San Francisco

In conclusion - we can see that Robbery is largely represented in the Central District as a percentage of it's crime.