Crime Analytics: Visualization of Incident Reports

Carolyn

September 18, 2016

In this analysis, I'm trying to show that in San Francisco summer 2014, SOUTHERN District has most crimes, while in city center larcency/theft is the most common incident. In fact, the robberies or theft are most common in the CENTRAL District. Below are the steps.

1. Setting the working directory.

2. Read in the data for San Francisco summer 2014 incidents.

```
sf_14summer_incidents<-read.csv('sanfrancisco_incidents_summer_2014.csv')
```

3. Load the libraries in need.

```
library(sqldf)
```

```
## Loading required package: gsubfn
```

```
## Loading required package: proto
```

```
## Loading required package: RSQLite
```

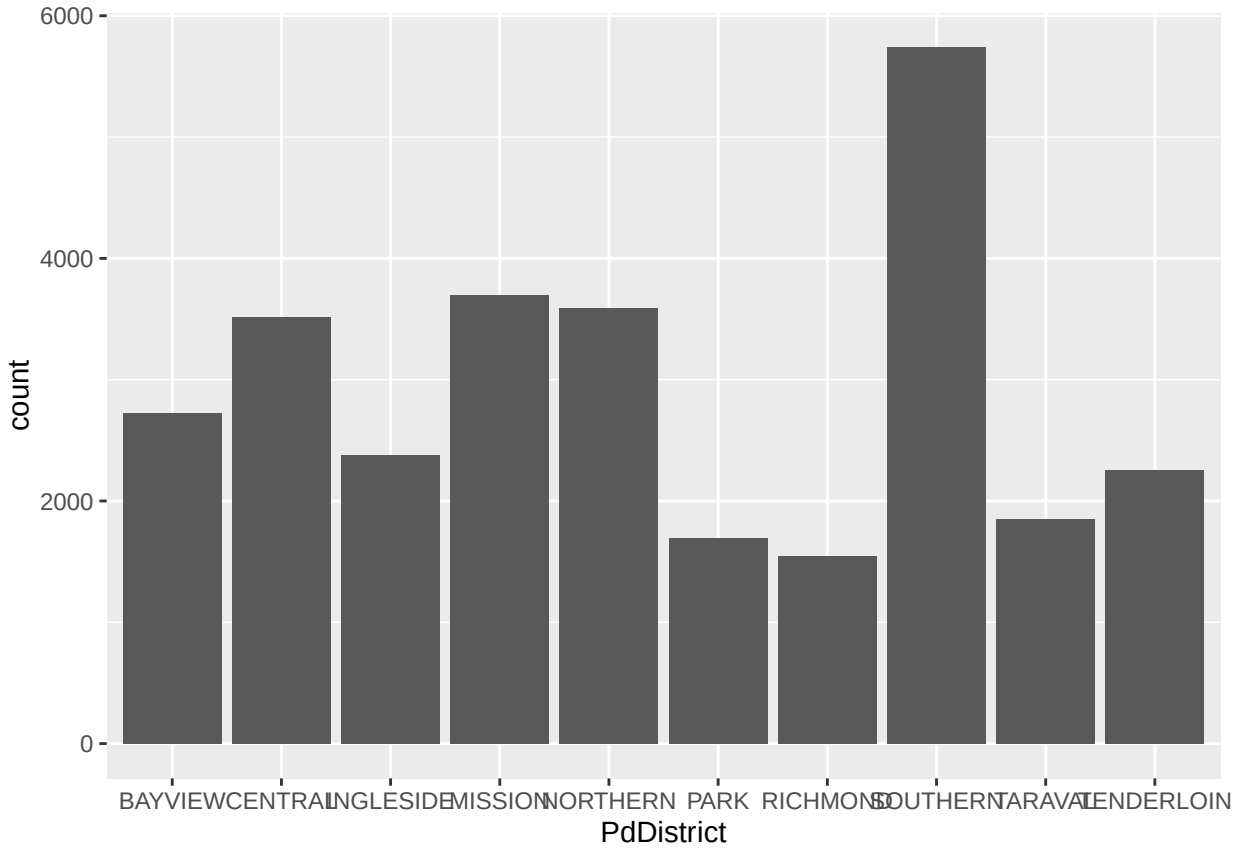
```
## Loading required package: DBI
```

```
library(ggplot2)
```

4. Analysis and Plots

First let's look at the totoal number of incidents by district.

```
qplot(PdDistrict,data = sf_14summer_incidents)
```



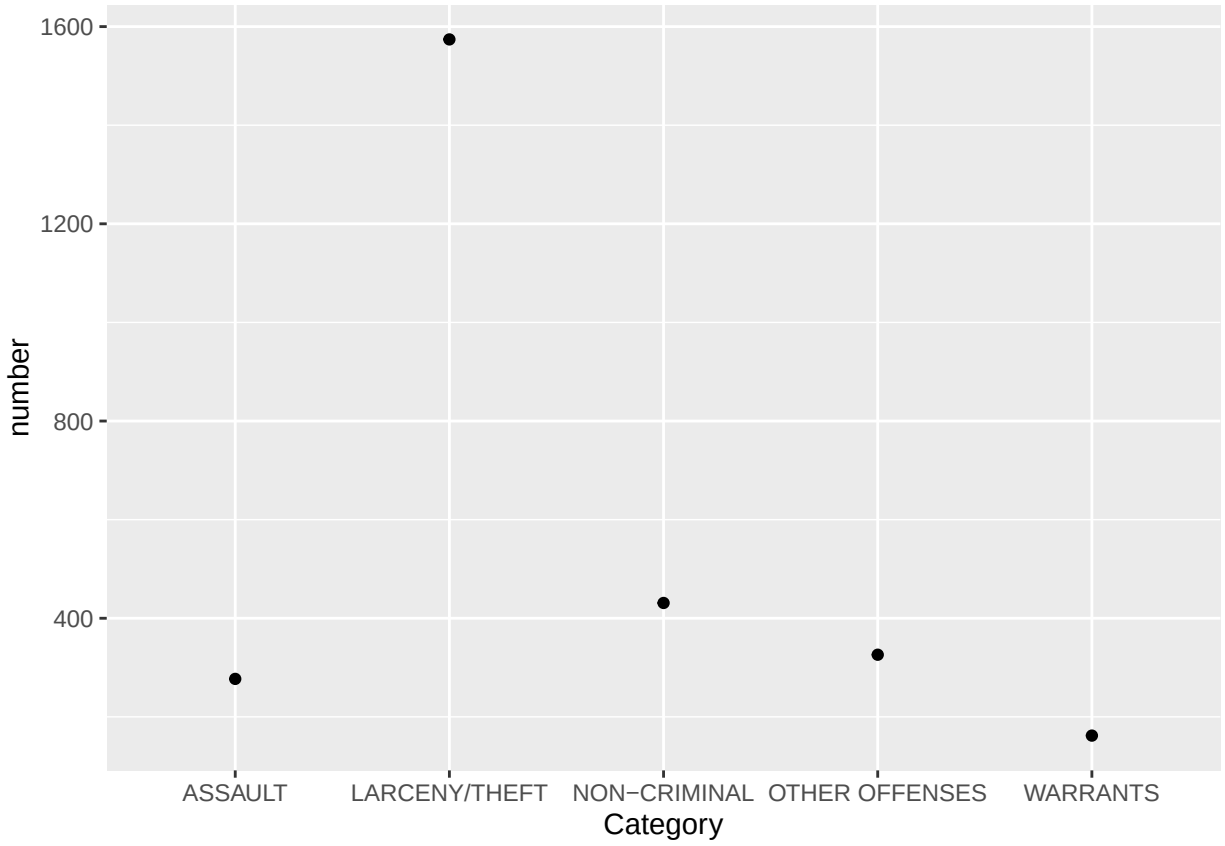
We can see the SOUTHERN District has the highest number of incidents.

Now let's look at city center, which is the CENTRAL District. What incidents are most common in this district? I will pick the top 5 most common incident categories.

```
central<-sqldf("select Category, count(*) as number from sf_14summer_incidents where PdDistrict = 'CENTR
```

```
## Loading required package: tcltk
```

```
central10<-central[1:5,]
qplot(Category, number, data = central10)
```



We can see from this plot, in city center, larcency/theft is the most commom incident.

Next, we need to check in what areas or neighborhoods are robberies or thefts most common? First, we need to take a look at all the incident categories and decide what categories can be thought of as robberies or theft.

```
table(sf_14summer_incidents$Category)
```

```
##
##                      ARSON                    ASSAULT
##                         63                       2882
##                    BRIBERY                   BURGLARY
##                          1                          6
##         DISORDERLY CONDUCT DRIVING UNDER THE INFLUENCE
##                         31                        100
##               DRUG/NARCOTIC                DRUNKENNESS
##                       1345                        147
##                EMBEZZLEMENT                  EXTORTION
##                         10                          7
##             FAMILY OFFENSES      FORGERY/COUNTERFEITING
##                         10                         18
##                      FRAUD                   GAMBLING
##                        242                          1
##                  KIDNAPPING              LARCENY/THEFT
##                        117                       9466
##                 LIQUOR LAWS                  LOITERING
##                         42                          3
```

```
##              MISSING PERSON                NON-CRIMINAL
##                        1266                        3023
##              OTHER OFFENSES     PORNOGRAPHY/OBSCENE MAT
##                        3567                           1
##                PROSTITUTION                    ROBBERY
##                         112                         308
##                    RUNAWAY             SECONDARY CODES
##                          61                         442
##             STOLEN PROPERTY                     SUICIDE
##                           8                          14
##              SUSPICIOUS OCC                    TRESPASS
##                        1300                         281
##                   VANDALISM               VEHICLE THEFT
##                          17                        1966
##                    WARRANTS                 WEAPON LAWS
##                        1782                         354
```

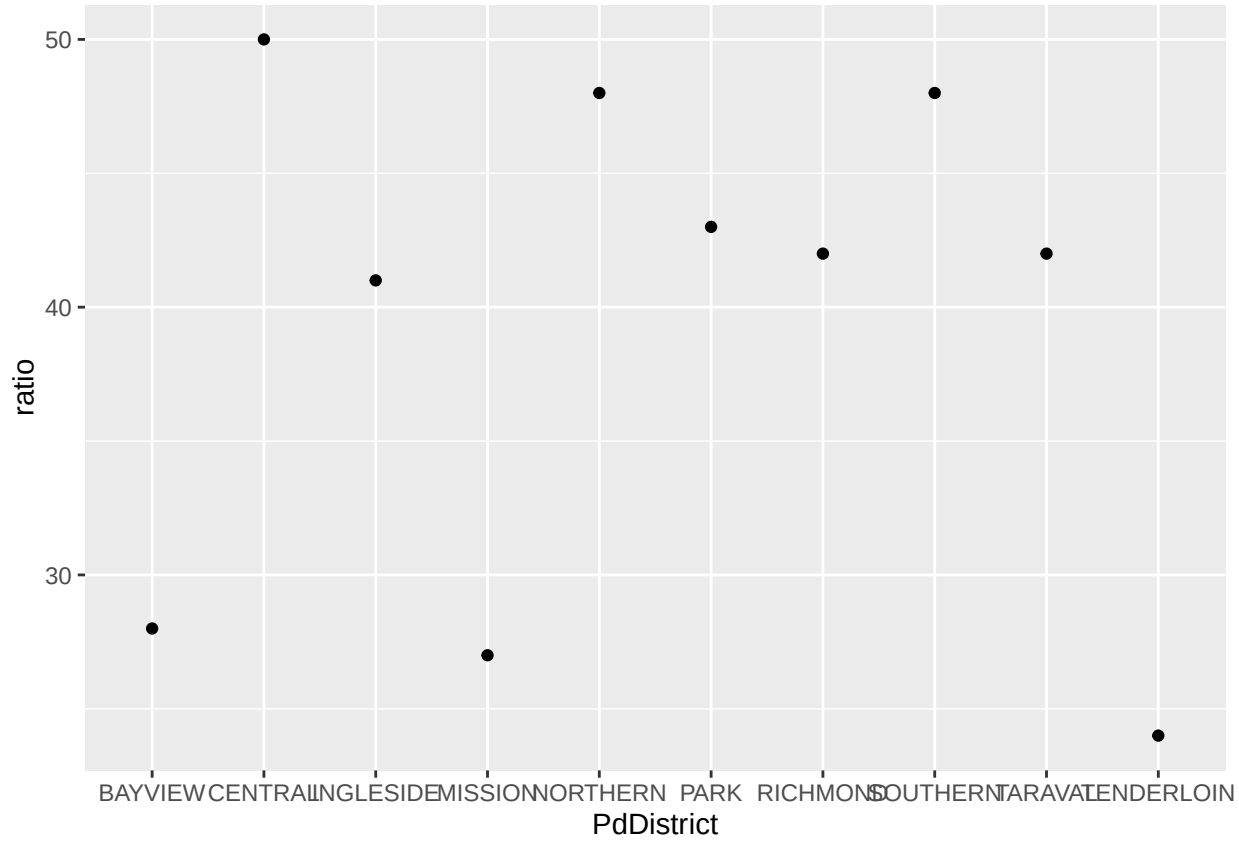
This is a quick list of total number of incidents by category. Of all these categories, we can choose and combine LARCENY/THEFT, ROBBERY, STOLEN PROPERTY, VEHICLE THEFT to represent robberies or theft.

Use sql to summarize number of robberies or theft by district and calculate the ratio of this kind of incidents over all the incidents in that district.

```
RobTheft<-sqldf("select PdDistrict, count(*) as number_RT from sf_14summer_incidents where Category in
SumTotal<-sqldf("select PdDistrict, count(*) as total from sf_14summer_incidents group by PdDistrict")
ratio<-sqldf("select a.PdDistrict, number_RT*100/total as ratio from RobTheft as a join SumTotal as b wh
```

Now we can plot the ratio.

```
qplot(PdDistrict, ratio, data = ratio)
```



As we can see, robberies or theft is most common in the CENTRAL District at around 50%, which is in line with our previous oberservation where larcency/theft is the most common incident in city center.