

Chicago Crimes

Loading library

```
library(data.table)
library(ggplot2)
library(ggmap)
```

Reading data

```
setwd("/work/Rproject/Data_Incubator")

data <- fread("Crimes_-_2001_to_present.csv")
```

```
##
Read 0.0% of 6107396 rows
Read 4.6% of 6107396 rows
Read 7.7% of 6107396 rows
Read 11.0% of 6107396 rows
Read 14.7% of 6107396 rows
Read 19.6% of 6107396 rows
Read 25.9% of 6107396 rows
Read 33.6% of 6107396 rows
Read 41.8% of 6107396 rows
Read 42.9% of 6107396 rows
Read 50.1% of 6107396 rows
Read 53.9% of 6107396 rows
Read 63.0% of 6107396 rows
Read 65.0% of 6107396 rows
Read 73.5% of 6107396 rows
Read 80.6% of 6107396 rows
Read 85.5% of 6107396 rows
Read 94.3% of 6107396 rows
Read 6107396 rows and 22 (of 22) columns from 1.339 GB file in 00:00:34
```

```
row.has.na <- apply(data, 1, function(x){any(is.na(x))})
clean_data <- data[!row.has.na]

houseprice <- fread("houseprice.csv")

map <- get_map("Chicago", zoom = 11, color = "bw")
```

Crimes Distribution

I estimate the density of crimes from the longitude and latitudes and use that to plot contours. This shows patterns and hotspots of crimes in Chicago.

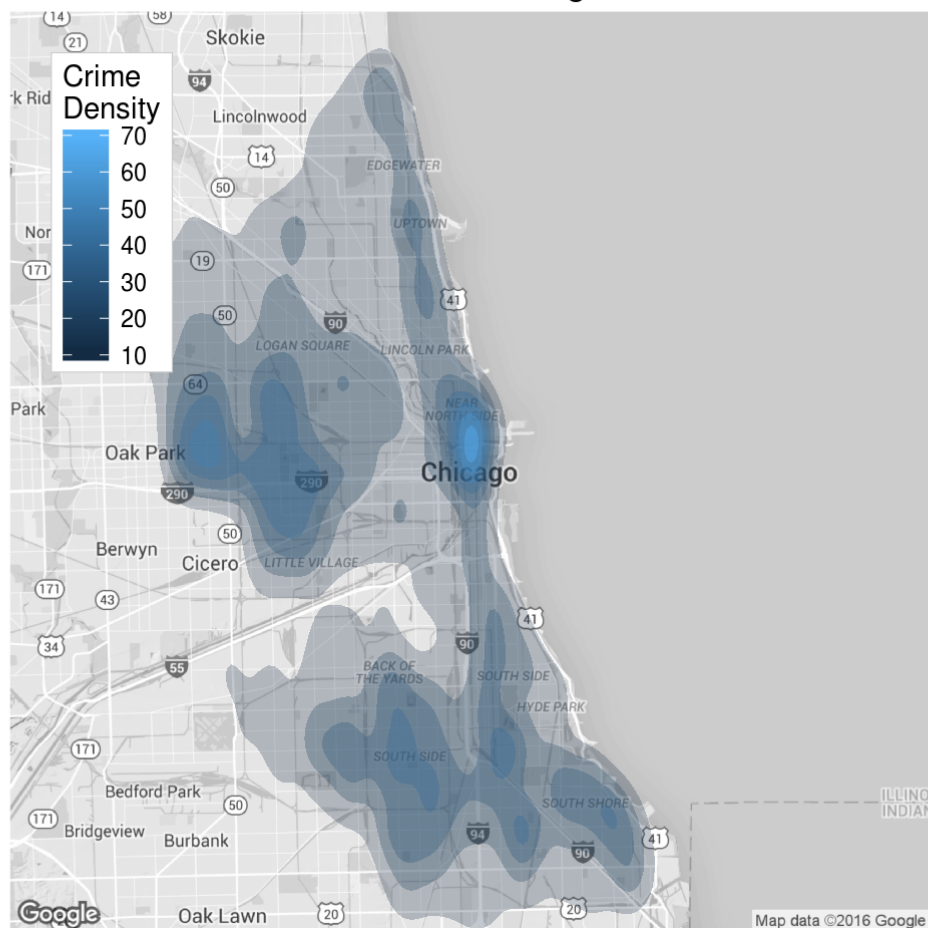
```

set.seed(26)
plot_data <- clean_data[,.(`Primary Type`,Longitude, Latitude)][sample(.N, 100000)]

contours <- stat_density2d(
  aes(x = Longitude, y = Latitude, fill = ..level.., alpha=..level..),
  size = 0.1, data = plot_data, n=200,
  geom = "polygon")
ggmap(map, extent='device', legend="topleft") + contours +
  scale_alpha_continuous(range=c(0.25,0.4), guide='none') +
  scale_fill_gradient('Crime\nDensity')+
  ggtitle('Crime in Chicago')

```

Crime in Chicago

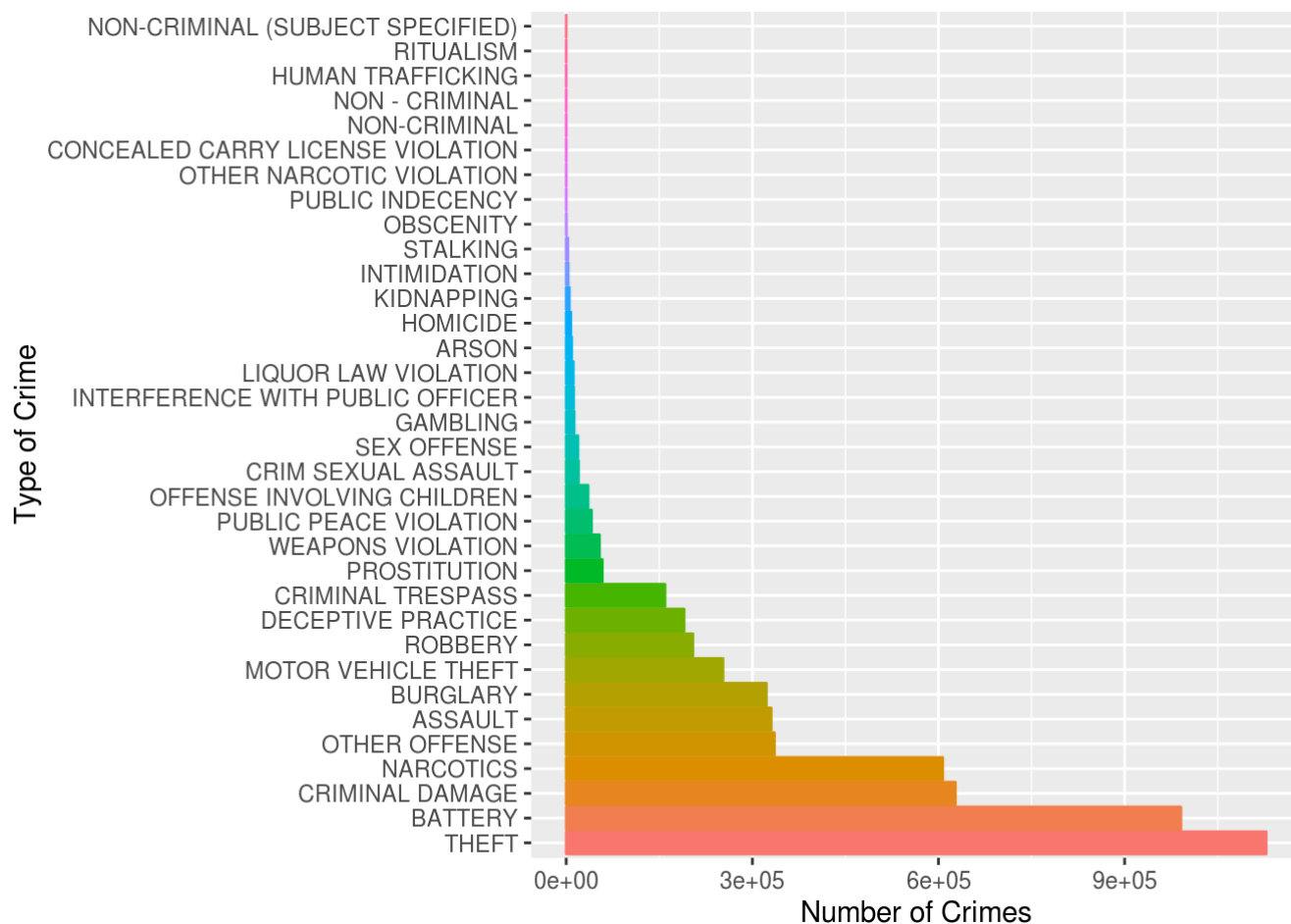


Types of Crime

I draw a bar plot to show number of crimes in each category. Bar plots indicate that the top crime category is Theft.

```
crimes <- clean_data[,.(.N), by=.`Primary Type`]]
crimes[,Category := reorder(`Primary Type`, -N)]

ggplot(crimes) +
  geom_bar(aes(x=Category, y=N, color = Category, fill = Category), stat="identity")+
  coord_flip() +
  ylab("Number of Crimes") +
  xlab("Type of Crime") +
  theme(legend.position="None")
```



Top 4 Crimes

I select top 4 crime types and plot the map density for each crime

```

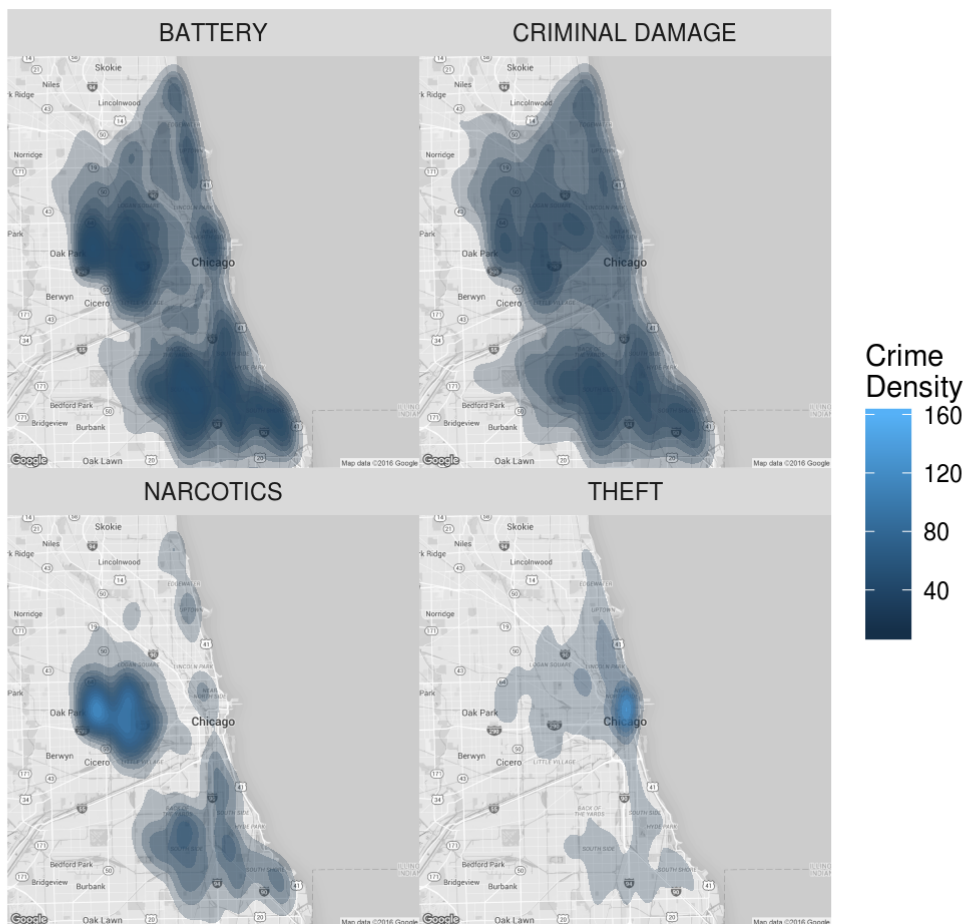
crimes <- crimes[order(-N)]
top_crimes <- head(crimes$Category,4)

plot_data <- clean_data[`Primary Type` %in% top_crimes][,.(`Primary Type`,Longitude, Latitude)]
[sample(.N, 100000)]

contours <- stat_density2d(
  aes(x = Longitude, y = Latitude, fill = ..level.., alpha=..level..),
  size = 0.1, data = plot_data, n=200,
  geom = "polygon")
ggmap(map, extent='device') + contours +
  scale_alpha_continuous(range=c(0.25,0.4), guide='none') +
  facet_wrap(~`Primary Type`) +
  scale_fill_gradient('Crime\nDensity')+
  ggtitle('Crime in Chicago')

```

Crime in Chicago



Crime by Day of Week and Hour

Convert date to day of week and Hour. I only consider top 4 crimes in Chicago for further calculation.

```

plot_data <- clean_data[`Primary Type` %in% top_crimes][,.(Date,`Primary Type`,Longitude, Latitude)]
plot_data[order(plot_data$Date)][sample(.N, 100000)]
date_time <- strptime(plot_data$Date, '%m/%d/%Y %I:%M:%S %p', tz="CDT")
plot_data$Weekday <- as.numeric(format(date_time, '%w')) # Sunday is 0
plot_data$Hour <- as.numeric(format(date_time, '%H'))
plot_data$Hour <- as.factor(plot_data$Hour)
plot_data$Weekday <- as.factor(plot_data$Weekday)
levels(plot_data$Weekday) <- c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
"Saturday")

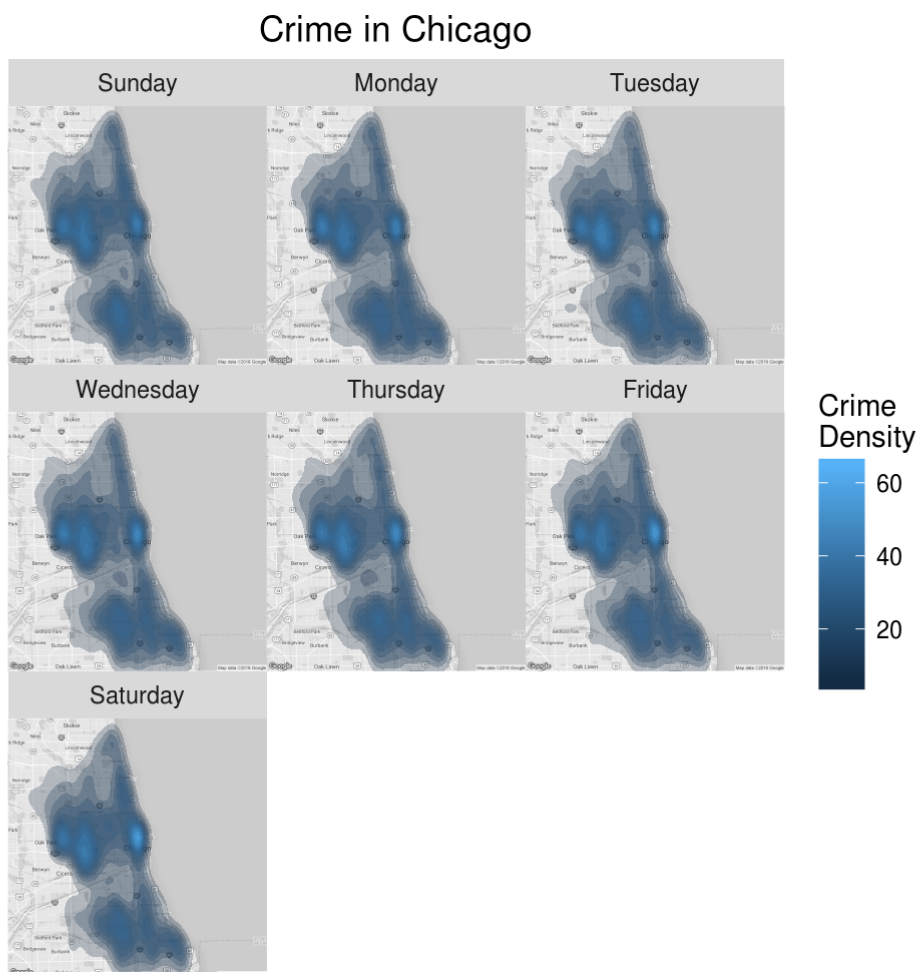
```

Distribution of crimes and top crime (theft) on different day of week

```

# Weekdays
contours <- stat_density2d(
  aes(x = Longitude, y = Latitude, fill = ..level.., alpha=..level..),
  size = 0.1, data = plot_data, n=200,
  geom = "polygon")
ggmap(map, extent='device') + contours +
  scale_alpha_continuous(range=c(0.25,0.4), guide='none') +
  facet_wrap(~Weekday) +
  scale_fill_gradient('Crime\nDensity')+
  ggtitle('Crime in Chicago')

```

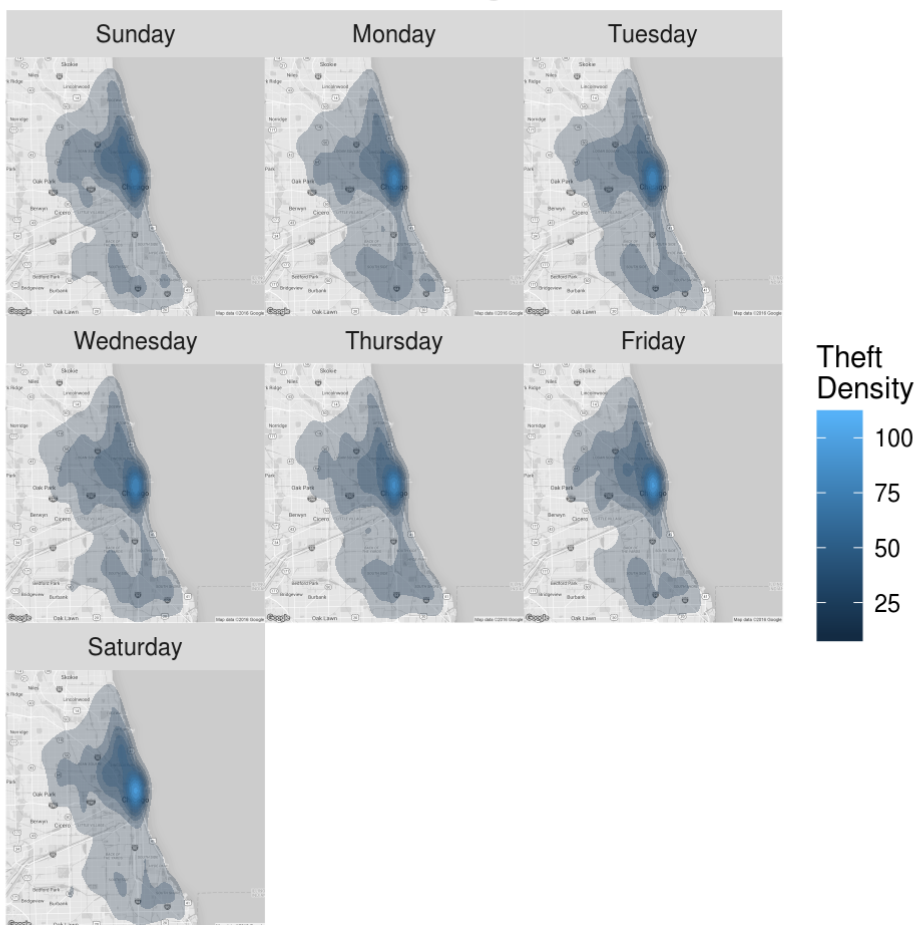


```

contours <- stat_density2d(
  aes(x = Longitude, y = Latitude, fill = ..level.., alpha=..level..),
  size = 0.1, data = plot_data[Primary Type=='THEFT'], n=200,
  geom = "polygon")
ggmap(map, extent='device') + contours +
  scale_alpha_continuous(range=c(0.25,0.4), guide='none') +
  facet_wrap(~Weekday) +
  scale_fill_gradient('Theft\nDensity')+
  ggtitle('Theft in Chicago')

```

Theft in Chicago

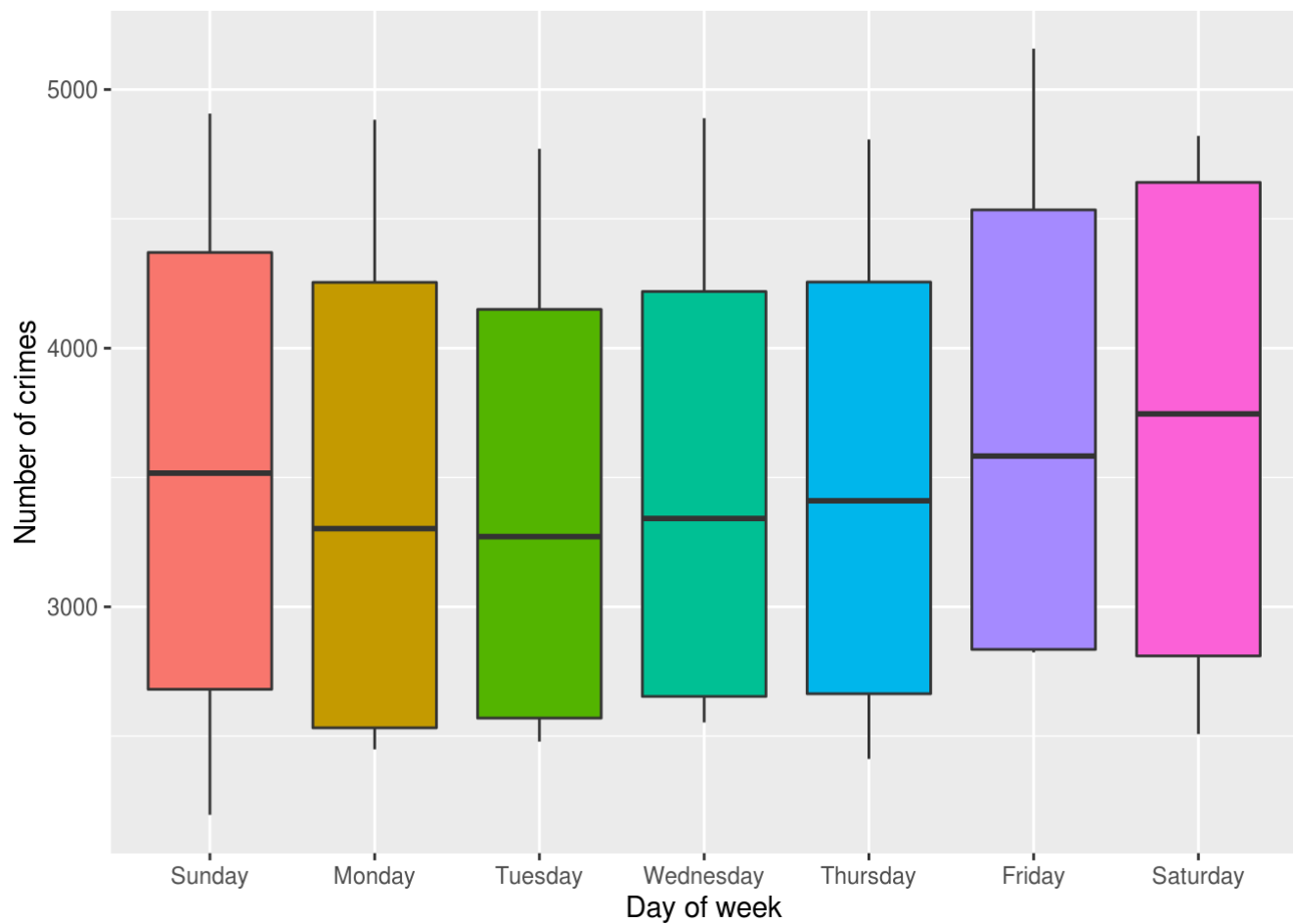


Number of crimes and day of week

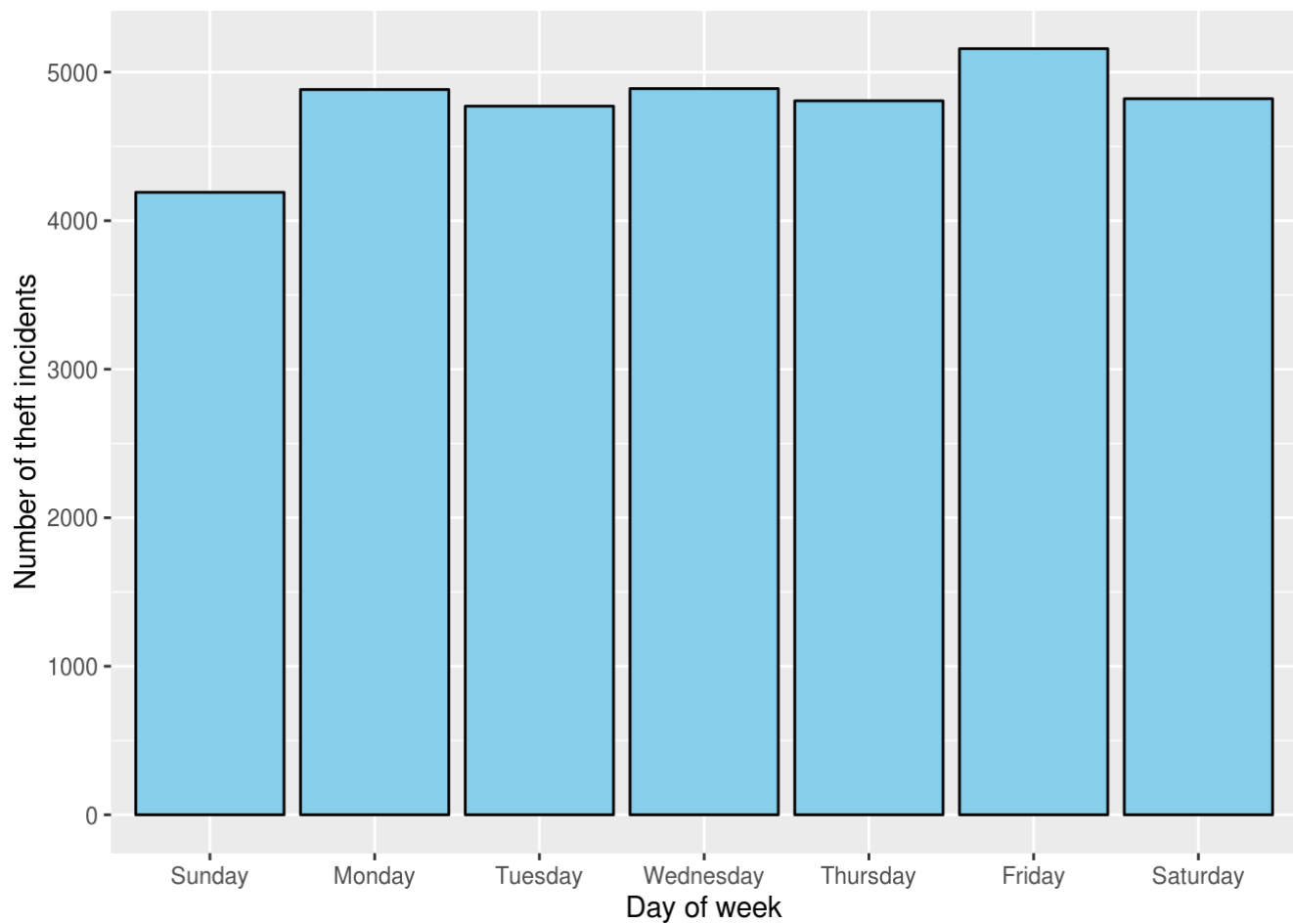
```

plot_weekday <- plot_data[,.(.N),by=(Primary Type, Weekday)]
ggplot(data = plot_weekday, aes(x=Weekday, y=N, fill = Weekday)) +
  geom_boxplot() +
  theme(legend.position="None") +
  xlab("Day of week")+
  ylab("Number of crimes")

```



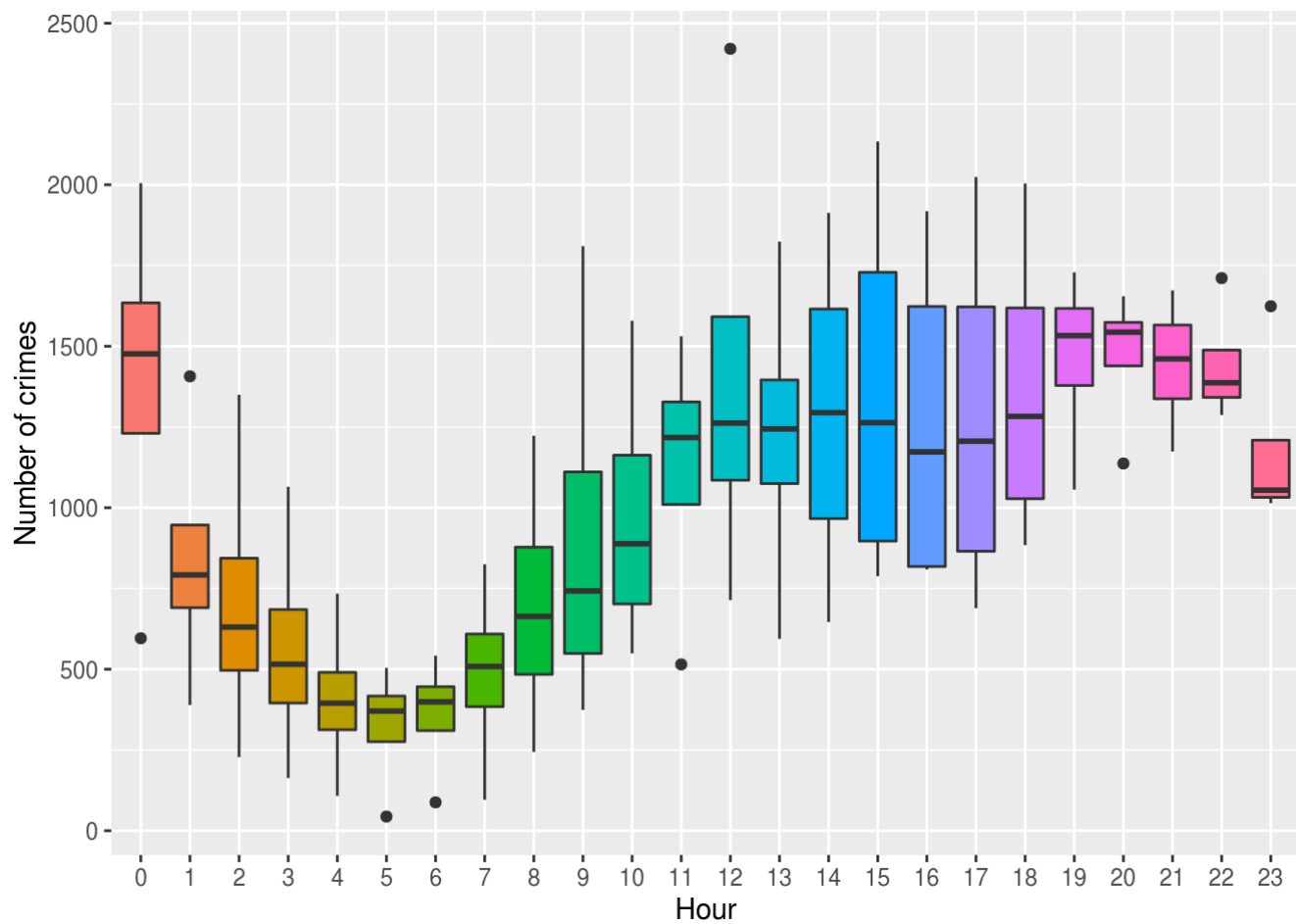
```
ggplot(data = plot_data[Primary Type=='THEFT'], aes(x=Weekday)) +  
  geom_bar(colour="black", fill="skyblue") +  
  theme(legend.position="None") +  
  xlab("Day of week")+  
  ylab("Number of theft incidents")
```



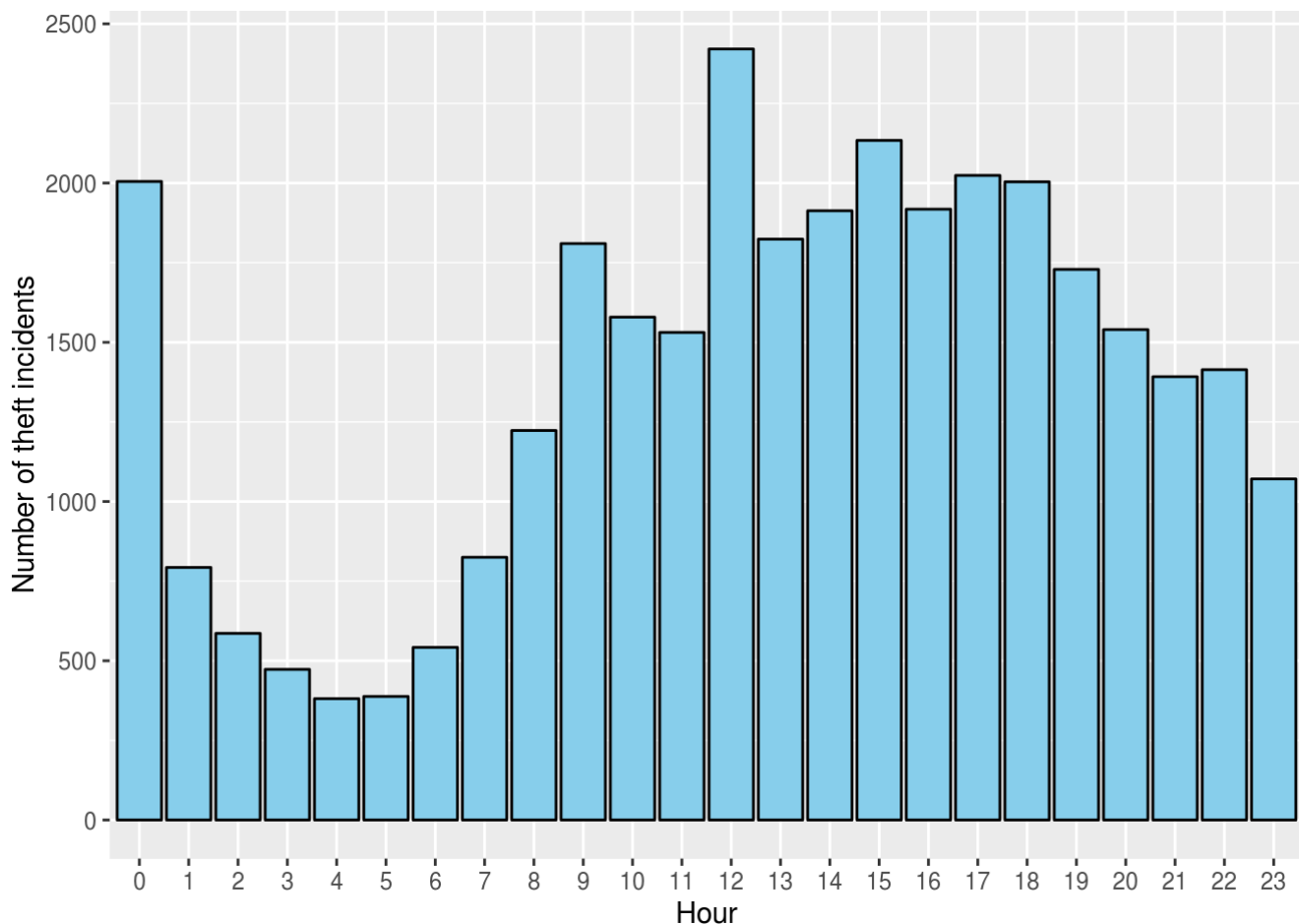
Number of crimes and hour

```
plot_Hour <- plot_data[,.(.N),by=.`Primary Type`, Hour]]

ggplot(data = plot_Hour, aes(x=Hour, y=N, fill = Hour)) +
  geom_boxplot() +
  theme(legend.position="None") +
  xlab("Hour")+
  ylab("Number of crimes")
```

```
ggplot(data = plot_data[`Primary Type`=='THEFT'], aes(x=Hour)) +
  geom_bar(colour="black", fill="skyblue") +
  theme(legend.position="None") +
  xlab("Hour")+
  ylab("Number of theft incidents")
```



Hyde Park

I am curious to know the distribution of crimes in my neighborhood, Hyde Park Chicago, to find out which area is safe. In addition, I plot the housing price in the map to compare the patterns of crime and housing price and find out if these two variables are depend on each other. People believe that the price of the house is higher in the safe area with lower crime rate but I would like to check this with real data. Unfortunately, I couldn't download a big data set from Zillow to see the patterns in houseing price.

```
map <- get_map("Chicago Hyde park", zoom = 14, color = "bw")
plot_data <- clean_data[Latitude < 41.816881 & Latitude > 41.773112 & Longitude > -87.621771] #[sample(.N,100000)]

contours <- stat_density2d(
  aes(x = Longitude, y = Latitude, fill = ..level.., alpha=..level..),
  size = 0.1, data = plot_data, n=200,
  geom = "polygon")

ggmap(map, extent='device') + contours +
  scale_alpha_continuous(range=c(0.25,0.4), guide='none') +
  scale_fill_gradient('Crime\nDensity') +
  ggtitle('Crime and Houseing Price in Hyde Park Chicago') +
  geom_point(data = houseprice, aes(x=lon, y=lat, colour=log(price))) +
  scale_colour_gradient(low="red", high = "black", 'log(price'))
```

Crime and Housing Price in Hyde Park Chicago

