Title: DataViz Assignment4: Exploratory Data Visualizations

Author: Ruochen Qiu Date: 10/15/2017

This report including five charts which illustrated the relationship between Chicago public school ACT test performance and their drop out rate, as well as the distribution of suspect crime which involve in gang activities and 2016 summer Chicago crime analysis.

```
library(tidyverse)
library(rmarkdown)
library(ggplot2)
library(readr)
library(haven)
library(dplyr)
library(tidyr)
library(stringr)
library(wesanderson)
library(RColorBrewer)
library(ggthemes)
library(plotrix)
library(raster)
library(ggmap)
library(cowplot)
```

Chart 1: Public School Performance 2012

Data loading and cleaning.

```
cps <- read_csv("cps1.csv")
cps <- cps %>%
  filter(`Overall Rating` != "Not Enough Data")

cps_plot <- ggplot(data = cps, aes(x=`Average Score ACT 2012`, y=`One-Year Dr
opout Rate 2012 - Percent`, color=`Overall Rating`))+geom_point(alpha=0.6, si
ze =4)+
  scale_y_continuous(breaks=seq(0, 0.3, 0.02))+theme_minimal()+
  scale_color_manual(values=wes_palette(n=3, name="GrandBudapest"))+
  theme(legend.title = element_text(size = 9),
        legend.position = c(0.9, 0.8),
        legend.text = element_text(size = 9))+
  labs(title = "Public School Performance 2012",
       subtitle = "   The level 1 public high schools had higher average ACT
score and lower drop out rate compare with other two levels,
while none of the Level 3 school had average ACT score higher than 20 even so
me of them had drop out rate lower than 0.02.
",caption = "Source:Chicago Public Schools Progress Report", x = "Average ACT
 Score", y = "One Year Drop out Rate")+
  theme(plot.title = element_text(hjust = 0.5))
```

Public School Performance 2012

The level 1 public high schools had higher average ACT score and lower drop out rate compare with oth
while none of the Level 3 school had average ACT score higher than 20 even some of them had drop out
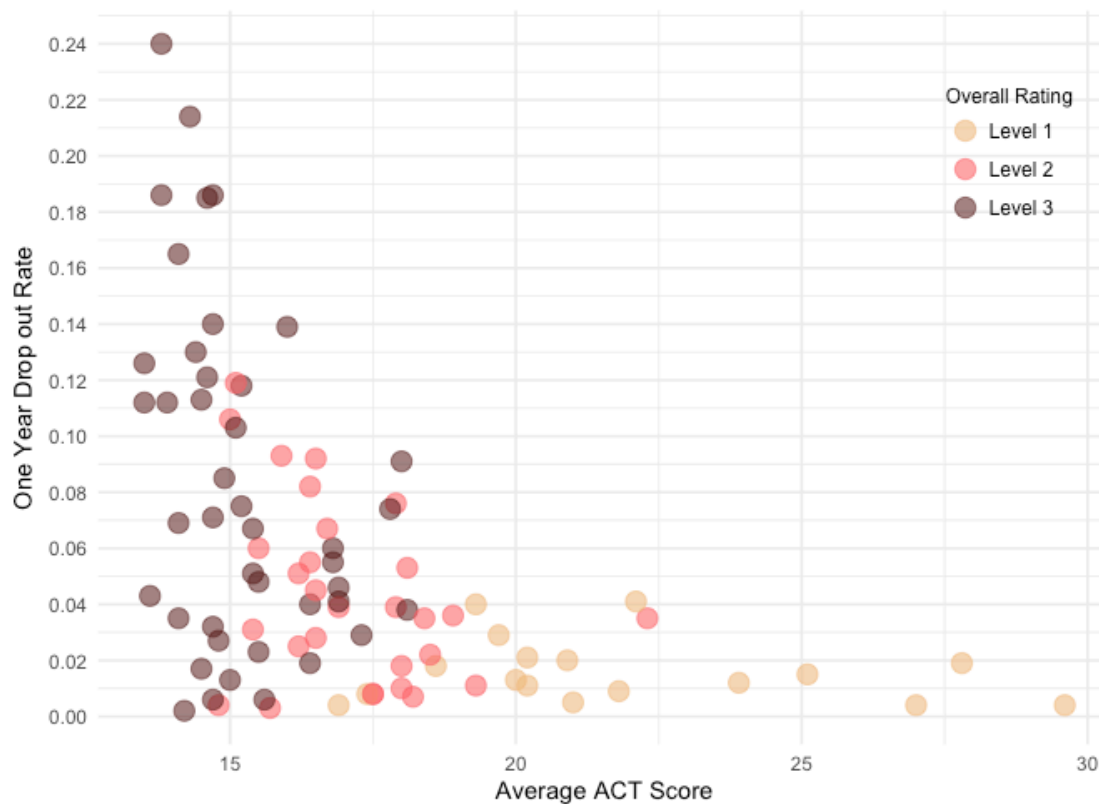
Source:Chicago Public Schools Progress Report

The full range of the ACT score for this data set is 0-0.25. As mentioned above, the data showing that public high schools in higher level had higher average ACT score and lower drop out rate, while the schools in lower level had average lower ACT scores **no matter** there drop out rate is.

Chart 2: 2016 Chicago Summer Investigatory Stop by Age

Data loading and cleaning.

```
isr <- read_csv("isr.csv")

isr_lim <- isr %>% subset(DATE >="7/1/16" & DATE <= "8/31/16")
isr_lim <- isr_lim %>%
  filter(AGE <= 90)
isr_lim <- isr_lim %>%
  add_count(AGE)
isr_lim <- isr_lim %>%
  filter(!is.na(`Contact Type`))

isr_plot <- ggplot(data = isr_lim, aes(x=AGE, fill=`Contact Type`))+geom_bar(
alpha=0.8)+
  scale_fill_manual(values = wes_palette("Chevalier"))+
```

```
    scale_x_continuous(breaks=seq(16, 90, 2)) + theme_minimal()+
  theme(legend.title = element_text(size = 9),
        legend.position = c(0.8, 0.9),
        legend.text = element_text(size = 9))+
labs(title = "2016 Chicago Summer Investigatory Stop by Age",
    subtitle = "The data showing that, from July, 2016 to August, 2016, youn
g people who aged from 18 to 24 are more likely
    to be stopped compare with other age group, while they also have higher
probability involve in gang activities.",
    caption = "Source:Chicago Police Investigatory Stop Reports
    ", x = "Age", y = "Investigatory Stop Count") +
      theme(plot.title = element_text(hjust = 0.5))
```
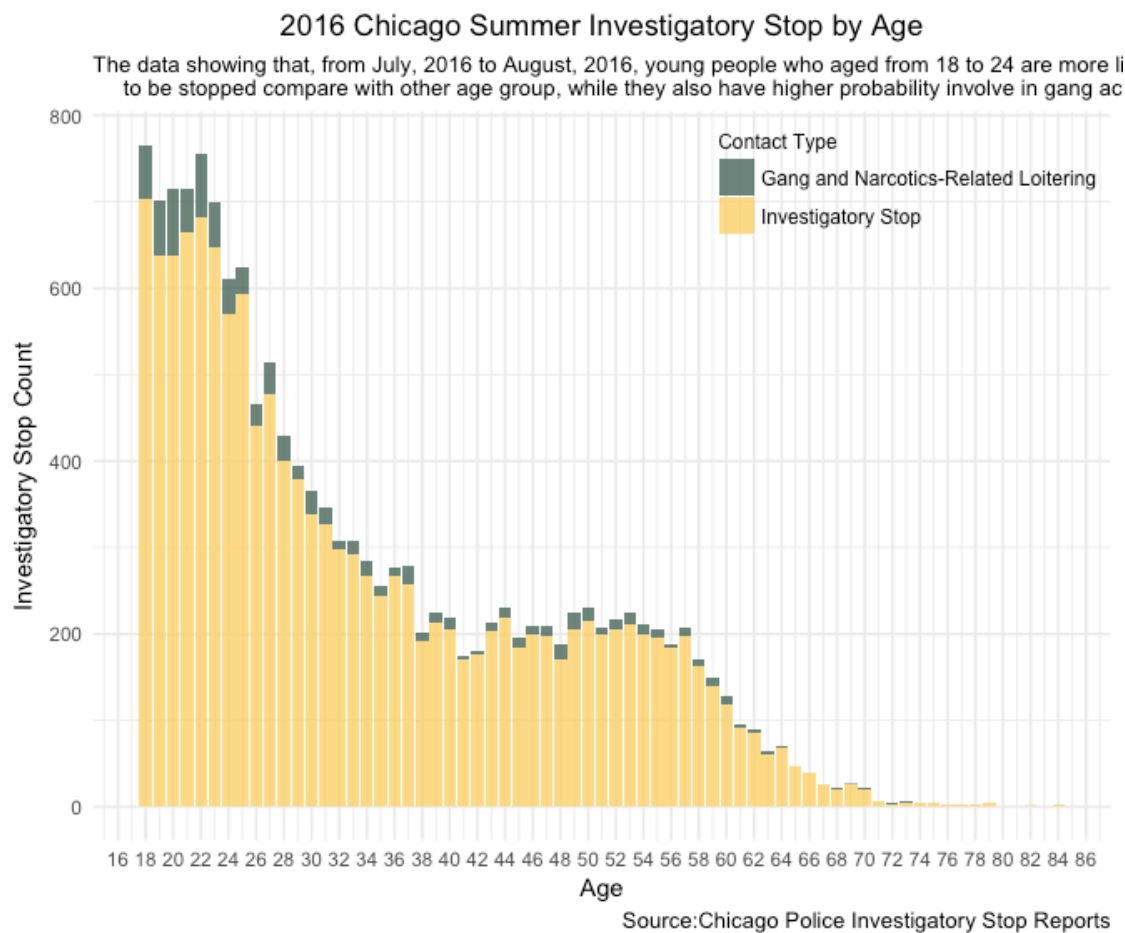

In this chart, only the investigatory stop data been studied while the legend shows weather the suspects involve in gang activities. It's pretty obvious that young people aged from 18-24 had higher possibility to be stopped while they also had higher likelihood to involving in **gang activities**.

## Chart 3: 2016 Chicago Crime by Hours

Data loading and cleaning.

```r
cri_lim <- read_csv("cri_lim_trim.csv")
cri_lim <- arrange(cri_lim,desc(value))

cri_plot<- ggplot(cri_lim, aes(x=Hour, y=`Primary Type`, fill= value)) + geom
_tile(colour = "white",alpha=0.9) +
   theme_minimal() +
  scale_fill_gradientn(colours = brewer.pal(9, 'YlOrRd'))+
  scale_x_discrete(expand = c(0, 0)) +
  scale_y_discrete(expand = c(0, 0)) +
  coord_equal() +
  theme(legend.position = "bottom", legend.direction = "horizontal",
             legend.box = "horizontal",
       legend.title = element_text(size = 9),
       legend.text = element_text(size = 9))+
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
+
  labs(title = "2016 Chicago Crime by Hours",
       subtitle = "The data showing that, in 2016, theft mostly happened duri
ng 12pm-18pm and battery are mainly
happened during mid-night.",
       caption = "Source:Chicago Data Portal
       ", x = "Hours", y = "Crime Type")+
       theme(plot.title = element_text(hjust = 0.5))
```

2016 Chicago Crime by Hours

The data showing that, in 2016, theft mostly happened during 12pm-18pm and battery are mainly happened during mid-night.
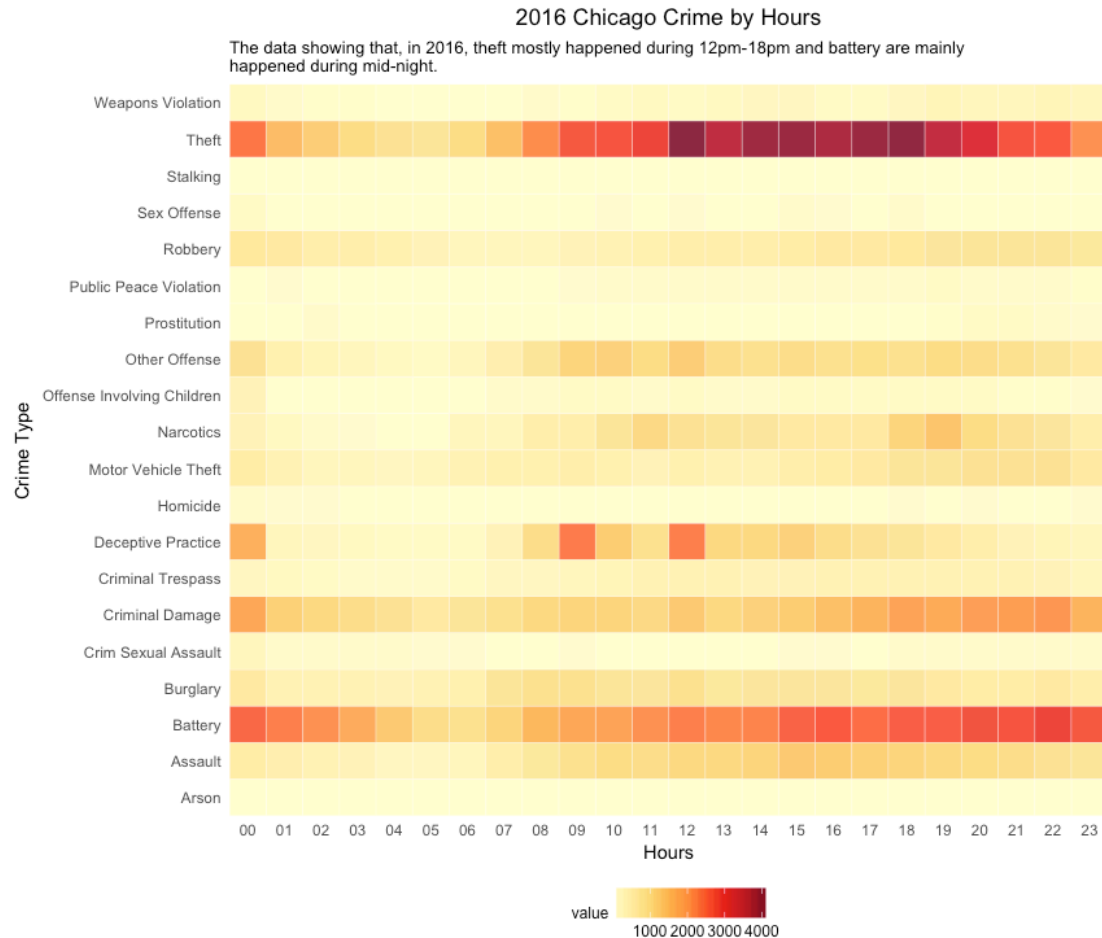
Source:Chicago Data Portal

This chart shows the most frequently appeared crime incidents during 24 hours in the summer of 2016. The dark red color indicating the highest volume while the light yellow shows the lowest volume. From the chart, we can tell that theft had significantly higher incidence rate compare with other crime types and it's more likely happening **during afternoon**. Battery had a similar tendency while deceptive practice more likely to appear at morning.

Chart 4: Comparison between Investigation Stop and Crime Offense by District

```
cri <- read_csv("cri.csv")

cri <- cri %>% subset(Date >="1/1/16" & Date <= "12/31/16")
isr <- isr %>% subset(DATE >="1/1/16" & DATE <= "12/31/16")

cri_n <-cri %>%
  group_by(District) %>%
  summarise(IS = n())

isr_n <- isr %>%
  group_by(DISTRICT) %>%
```

```
  summarise(Cr = n())

names(cri_n)[names(cri_n)=="District"] <- "DISTRICT"

cri_n <- cri_n %>%
  filter(!is.na(DISTRICT))
isr_n <- isr_n %>%
  filter(!is.na(DISTRICT))

join <- left_join(cri_n, isr_n, by = "DISTRICT")

Investigation_Stop <- join$IS
Crime <- join$Cr
District <-join$DISTRICT

par(cex=0.85,bg="transparent")
par(mar=pyramid.plot(Investigation_Stop, Crime, top.labels=c("Investigation S
top", "",  "Crime"),labels=District,main="Comparison between Investigation St
op and Crime Offense by District
Source:Chicago Police Investigatory Stop Reports", lxcol="#d9a744", rxcol="#5
75e70", unit="",  gap=0 ))
```
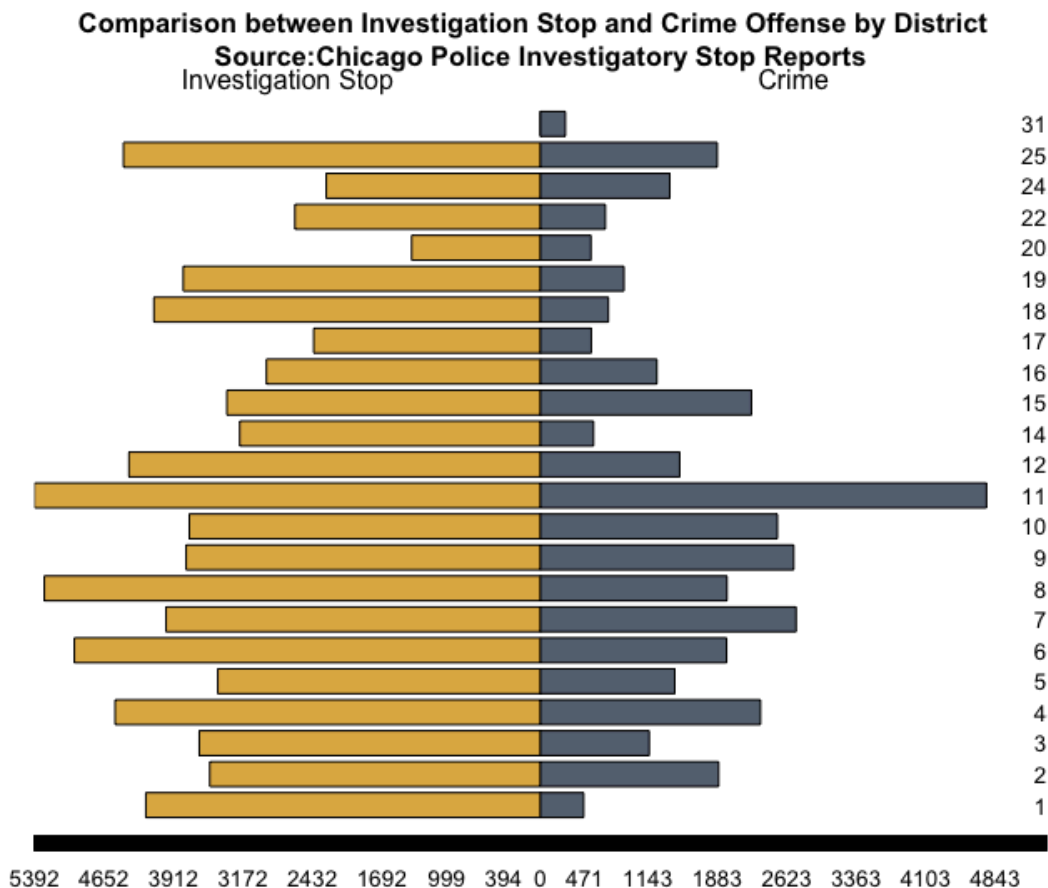

Comparison between Investigation Stop and Crime Offense by District
Source:Chicago Police Investigatory Stop Reports

This chart compared the frequency of investigatory stop with the frequency of real crime been found among districts. District 31 is the only one that investigatory stop less than crime incidents, while district 11 had a pretty close rate in both side. The police officers in district 1 had relatively high stop checking rate, compare with others.

Chart 5: 2016 Summer Chicago Severe Crime Map

```
cri_offense<- cri %>% subset(Date >="7/1/16" & Date <= "8/31/16")
cri_offense$Location [cri_offense$Location  == ''] <- NA
cri_offense<- na.omit(cri_offense)

cri_offense$Longitude <- round(as.double(cri_offense$Longitude), 3)
cri_offense$Latitude <- round(as.double(cri_offense$Latitude), 3)

cri_offense <- cri_offense %>%  filter(`Primary Type` =="WEAPONS VIOLATION"|`
Primary Type` =="SEX OFFENSE"|`Primary Type` =="ROBBERY"|`Primary Type` =="HO
MICIDE")

cri_offense$`Primary Type` <- factor(cri_offense$`Primary Type`, levels = c("
ROBBERY","SEX OFFENSE","WEAPONS VIOLATION","HOMICIDE"))

cri_offense<- cri_offense %>%
  group_by(`Longitude`,`Latitude`,`Primary Type`) %>%
  summarise(TOTAL = n())

chicago <- get_stamenmap(bbox = c(left = -87.885169, bottom=41.643919,
                                  right = -87.523984, top = 42.023022),
                         zoom=12,maptype="toner")

chicago <- ggmap(chicago, extent ="device")
map<- chicago+geom_point( data = cri_offense, aes(x = Longitude, y = Latitude
, color =`Primary Type`,size=TOTAL), alpha=1)+guides(size=FALSE)+ scale_color
_manual(values=wes_palette(n=4, name="Moonrise2")) +
  theme(legend.position = "bottom", legend.direction = "horizontal",
        legend.box = "horizontal",
        legend.title = element_text(size = 8),
        legend.text = element_text(size = 8))+
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
+
  labs(title = "2016 Summer Chicago Severe Crime Map",
       subtitle = "The data showing that the residents live in suburb and
  south area had higher possibility involve in homicide
  compare with downtown and north area during summer.",
       caption = "Source:Chicago Data Portal")+
       theme(plot.title = element_text(hjust = 0.5))
```
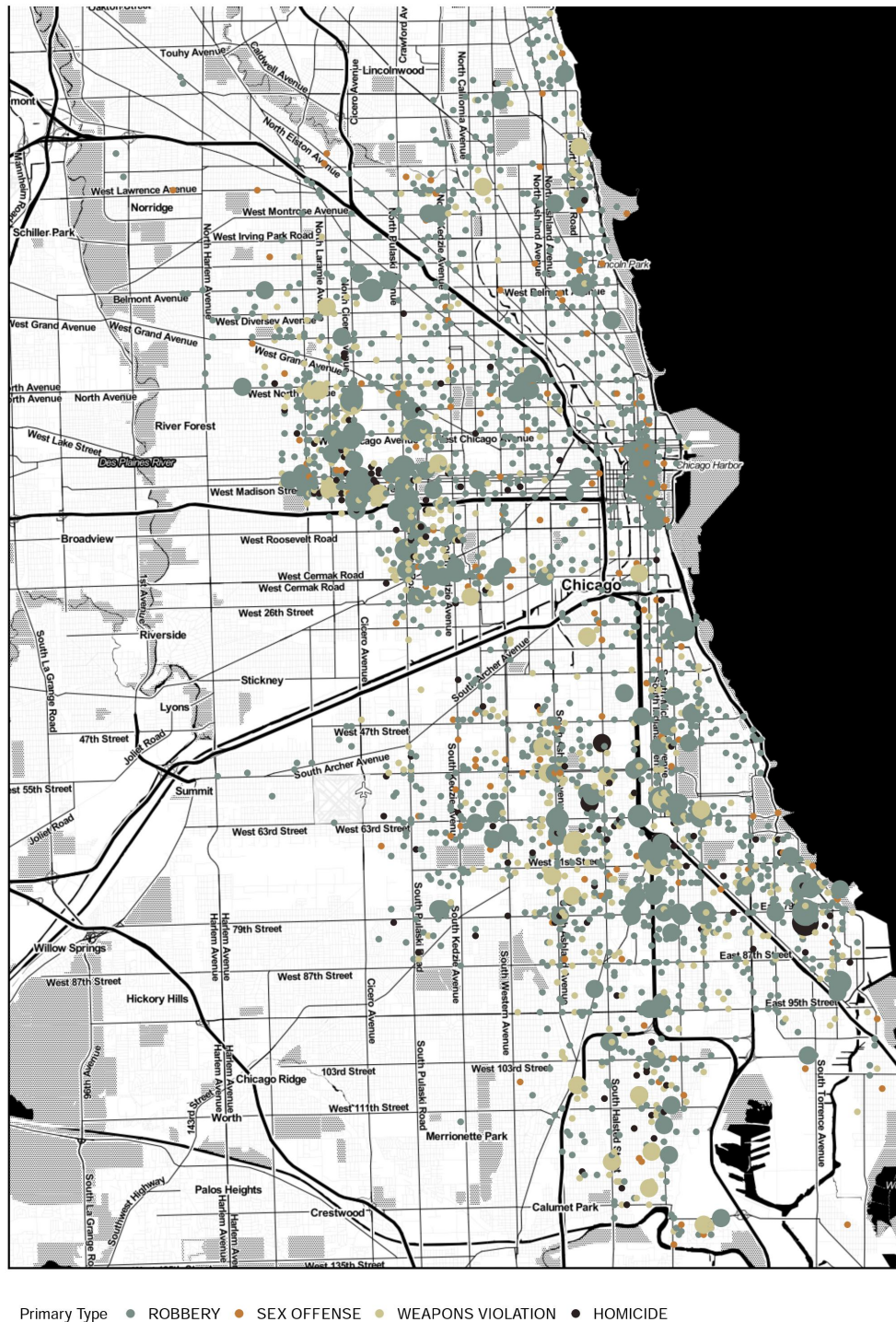
2016 Summer Chicago Severe Crime Map

The data showing that the residents live in suburb and
south area had higher possibility involve in homicide
compare with downtown and north area during summer.

Primary Type   ● ROBBERY   ● SEX OFFENSE   ● WEAPONS VIOLATION   ● HOMICIDE

Source:Chicago Data Portal

This map is focus on the severe crime incidents appeared during summer 2016 at Chicago.
The point size indicating the frequency of the incidents while the four crime types been

denoted with four different colors. It's easy to tell that robbery and weapons violence are roughly spread around the whole city, while homicide incidents are more likely to appear in **southern** part of the city.