Crimes in Los Angeles

Introduction

City of Los Angeles or "The Birthplace of Jazz" one of the most populous city in the United States of America with the population estimated over four million. With the city of this size it is worth the effort to explore the crime rate in this city.

The current project is aimed to explore the crime rate between 2017 until the recent update in 2018. The dataset used in this project is found in this link which is provided by Los Angeles Police Department.

Date Preparation

```
library(data.table)
library(tidyverse)
library(ggthemes)
library(ggmap)
library(maps)
library(mapdata)
library(lubridate)
library(stringr)
library(ggrepel)
library(xts)
library(varhandle)

crime_la <- as.data.frame(fread("Crime_Data_from_2010_to_Present.csv", na.strings = c("NA")))

glimpse(crime_la)
```

```
## Observations: 1,805,537
## Variables: 26
## $ `DR Number`              <int> 1208575, 102005556, 418, 101822289, 4...
## $ `Date Reported`          <chr> "03/14/2013", "01/25/2010", "03/19/20...
## $ `Date Occurred`          <chr> "03/11/2013", "01/22/2010", "03/18/20...
## $ `Time Occurred`          <int> 1800, 2300, 2030, 1800, 2300, 1400, 2...
## $ `Area ID`                <int> 12, 20, 18, 18, 21, 1, 11, 16, 19, 9,...
## $ `Area Name`              <chr> "77th Street", "Olympic", "Southeast"...
## $ `Reporting District`     <int> 1241, 2071, 1823, 1803, 2133, 111, 11...
## $ `Crime Code`             <int> 626, 510, 510, 510, 745, 110, 510, 51...
## $ `Crime Code Description` <chr> "INTIMATE PARTNER - SIMPLE ASSAULT", ...
## $ `MO Codes`               <chr> "0416 0446 1243 2000", "", "", "", "0...
## $ `Victim Age`             <int> 30, NA, 12, NA, 84, 49, NA, NA, NA, 2...
## $ `Victim Sex`             <chr> "F", "", "", "", "M", "F", "", "", ""...
## $ `Victim Descent`         <chr> "W", "", "", "", "W", "W", "", "", ""...
## $ `Premise Code`           <int> 502, 101, 101, 101, 501, 501, 108, 10...
## $ `Premise Description`    <chr> "MULTI-UNIT DWELLING (APARTMENT, DUPL...
## $ `Weapon Used Code`       <int> 400, NA, NA, NA, NA, 400, NA, NA, NA,...
## $ `Weapon Description`     <chr> "STRONG-ARM (HANDS, FIST, FEET OR BOD...
## $ `Status Code`            <chr> "AO", "IC", "IC", "IC", "IC", "AA", "...
## $ `Status Description`     <chr> "Adult Other", "Invest Cont", "Invest...
## $ `Crime Code 1`           <int> 626, 510, 510, 510, 745, 110, 510, 51...
## $ `Crime Code 2`           <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
```

```
## $ `Crime Code 3`           <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ `Crime Code 4`           <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ Address                  <chr> "6300    BRYNHURST                    ...
## $ `Cross Street`           <chr> "", "15TH", "", "WALL", "", "", "AVEN...
## $ Location                 <chr> "(33.9829, -118.3338)", "(34.0454, -1...
```

The data used in this project contains 1.8 millions observation and 26 variables. The dataset date range from 2010 up until recent 22/08/2018. The description of the dataset can be seen below:

- *DR Number*: Division of Records Number: Official file number made up of a 2 digit year, area ID, and 5 digits
- *Date Reported*: formated in MM/DD/YYYY
- *Date Occurred*: formated in MM/DD/YYYY
- *Time Occurred*: In 24 hour military time
- *Area ID*: The LAPD 21 Community Police Station sequentially numbered from 1-21
- *Area Name*: The name of the 21 Geaographic Areas or Patrol Divisions
- *Reporting District*: A four-digit code represents a sub-area within a Geographic Area
- *Crime Code*: Indicated the crime commited
- *Crime Code Description*: Description from Crime Code.
- *MO Codes*: Modus Operandi: Acitivities associated witht he suspect in commision of the crime.
- *Victim Age*: Age in two character numeric
- *Victim Sex*: F- Female, M-Male, X-Unknown
- *Victim Descent*: A-Other Asian, B-Black, C-Chinese, D-Cambodian, F-Filipino, G-Guamanian, H-Hispanci/Latin/Mexican, I-American Indian/Alaskan Native, J-Japanese, K-Korean, L-Laotian, O-Other, P-Pacific Islander, S-Somoan, U-Hawaiian, V-Vietnamese, W-White, X-Unknown, Z-Asian Indian
- *Premise Code*: The type of structure, vechicle, or location where the crime took place
- *Premise Description*: Defines the Premise Code provided
- *Weapon Used Code*: The type of weapon used in the crime
- *Weapon Description*: Defines of the Weapon Used Code
- *Status Code*: Status of the case
- *Status Description*: Defines the Status Code
- *Crime Code 1*: Indicates the crime commited. Crime code 1 is the primary and most serious one. Crime Code 2, 3, and 4 are respectively less serious offenses
- *Crime Code 2*: May contain a code for an additional crime, less serious than Crime Code 1
- *Crime Code 3*: May contain a code for an additional crime, less serious than Crime Code 1
- *Crime Code 4*: May contain a code for an additional crime, less serious than Crime Code 1
- *Address*: Street address of crime incident rounded to the nearest hundredblock to maintain anonymity
- *Cross Street*: Cross street of rounded Address
- *Location*: The location where the crime indident occurred. XY indicated latitudes and longitude.

Data cleaning

Before the analysis, a simple data analysis such as convert data into corrected data type, recode the variable into readable format and select relevant variables is conducted as shown below:

```
#select relevant variables
crime_la_selected <- select(crime_la, `Date Occurred`, `Time Occurred`, `Area Name`, `Crime Code Descri

#convert the date into date type
crime_la_selected$`Date Occurred` <- mdy(crime_la_selected$`Date Occurred`)

#Separate latitude and longitude
location <- crime_la_selected$Location %>% # take coord as string
  str_replace_all("[()]", "") %>% # replace parantheses
  str_split_fixed(", ", n=2) %>% # split up based on comma and space after
  as.data.frame %>% # turn this to a data frame
  transmute(lat=V1, long=V2) # rename the variables

#combine the lat and long then remove the location
crime_la_selected <- cbind(crime_la_selected, location)

crime_la_selected <- subset(crime_la_selected, select = -c(Location))

#select only 2017 and 2018
crime_selected_years <- filter(crime_la_selected, `Date Occurred` >= as_date("2017-01-01"), `Date Occur

#remove these data frames to same memory
rm(crime_la, crime_la_selected, location) #remove these data frames to same memory

#separate date into year, month and day.
crime_selected_years$year <- year(crime_selected_years$`Date Occurred`)
crime_selected_years$month <- month(crime_selected_years$`Date Occurred`)
crime_selected_years$days <- day(crime_selected_years$`Date Occurred`)

#Recode the variable into readable format
crime_selected_years$`Victim Sex` <- recode(crime_selected_years$`Victim Sex`, 'F' = 'Female', 'M' = 'M

crime_selected_years$`Victim Descent` <- recode(crime_selected_years$`Victim Descent`, "A" = "Other Asi

#convert the character into factor
character_vars <- lapply(crime_selected_years, class) == "character"
crime_selected_years[, character_vars] <- lapply(crime_selected_years[, character_vars], as.factor)

glimpse(crime_selected_years)
```

```
## Observations: 369,945
## Variables: 15
## $ `Date Occurred`          <date> 2017-07-20, 2017-07-21, 2017-04-21, ...
## $ `Time Occurred`          <int> 2000, 1000, 1930, 1700, 745, 1, 730, ...
## $ `Area Name`              <fct> West Valley, West Valley, Rampart, Ra...
## $ `Crime Code Description` <fct> BURGLARY FROM VEHICLE, BURGLARY FROM ...
## $ `Victim Age`             <int> 55, 20, 16, 16, 16, 16, 16, 16, 16, 2...
## $ `Victim Sex`             <fct> Male, Male, , , , , , , , Male, , , ,...
## $ `Victim Descent`         <fct> Other, Other, , , , , , , , Black, , ...
## $ `Premise Description`    <fct> , , STREET, STREET, STREET, STREET, S...
```

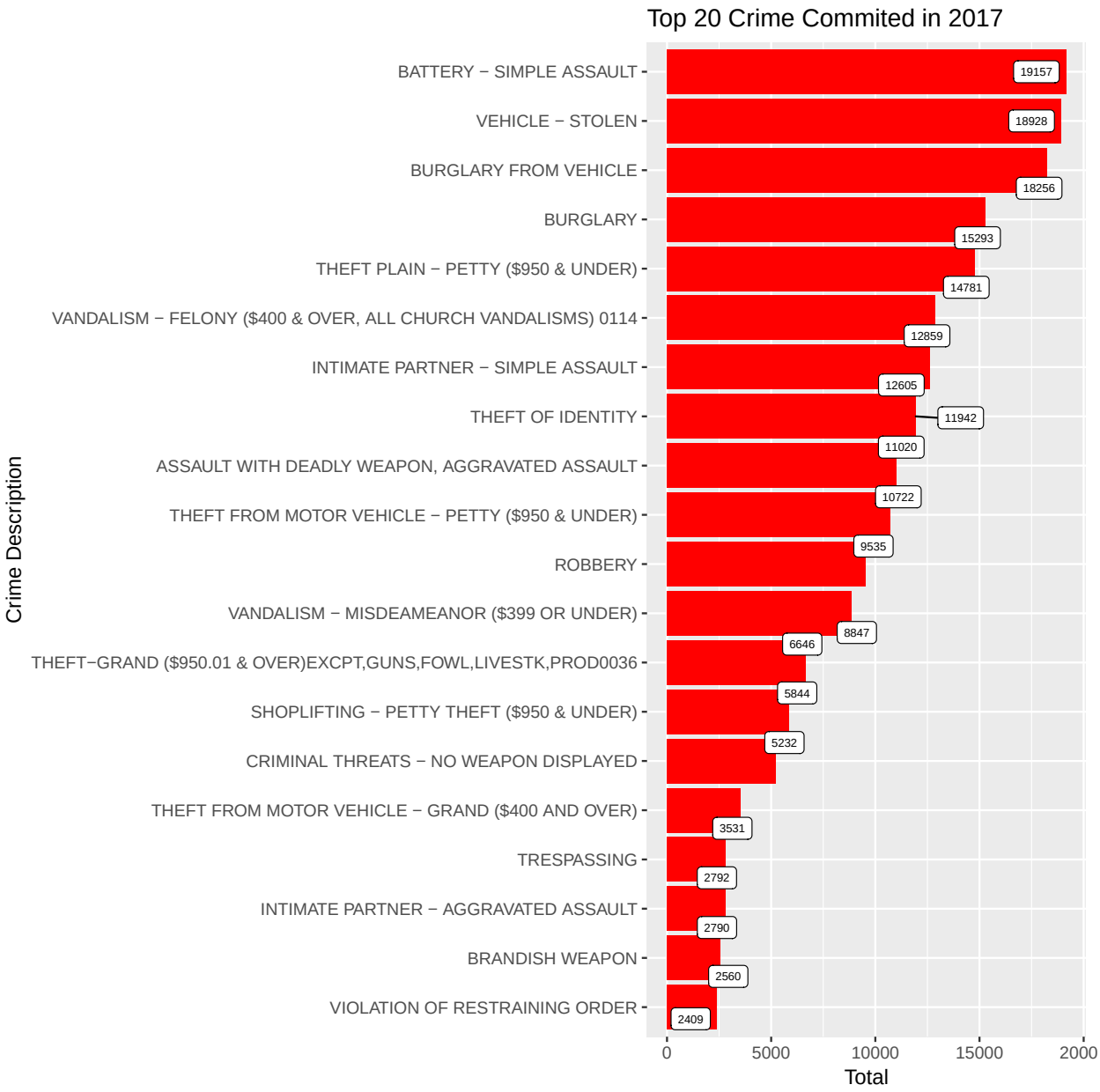
```
## $ `Weapon Description`     <fct> , , , , , , , , , , , , , , , , , , , ,...
## $ `Status Description`     <fct> Invest Cont, Invest Cont, Invest Cont...
## $ lat                      <fct> , , 34.0886, 34.0512, 34.0328, 34.067...
## $ long                     <fct> , , -118.2979, -118.2787, -118.2915, ...
## $ year                     <dbl> 2017, 2017, 2017, 2017, 2017, 2017, 2...
## $ month                    <dbl> 7, 7, 4, 2, 4, 4, 4, 3, 5, 6, 1, 2, 3...
## $ days                     <int> 20, 21, 21, 11, 25, 7, 8, 6, 11, 6, 2...
```

After the data cleaning process, only 369.945 observation and 15 variables are selected.

Total Crime in 2017 and 2018

Lets look at the top 20 of crime that have been comminted in 2017.

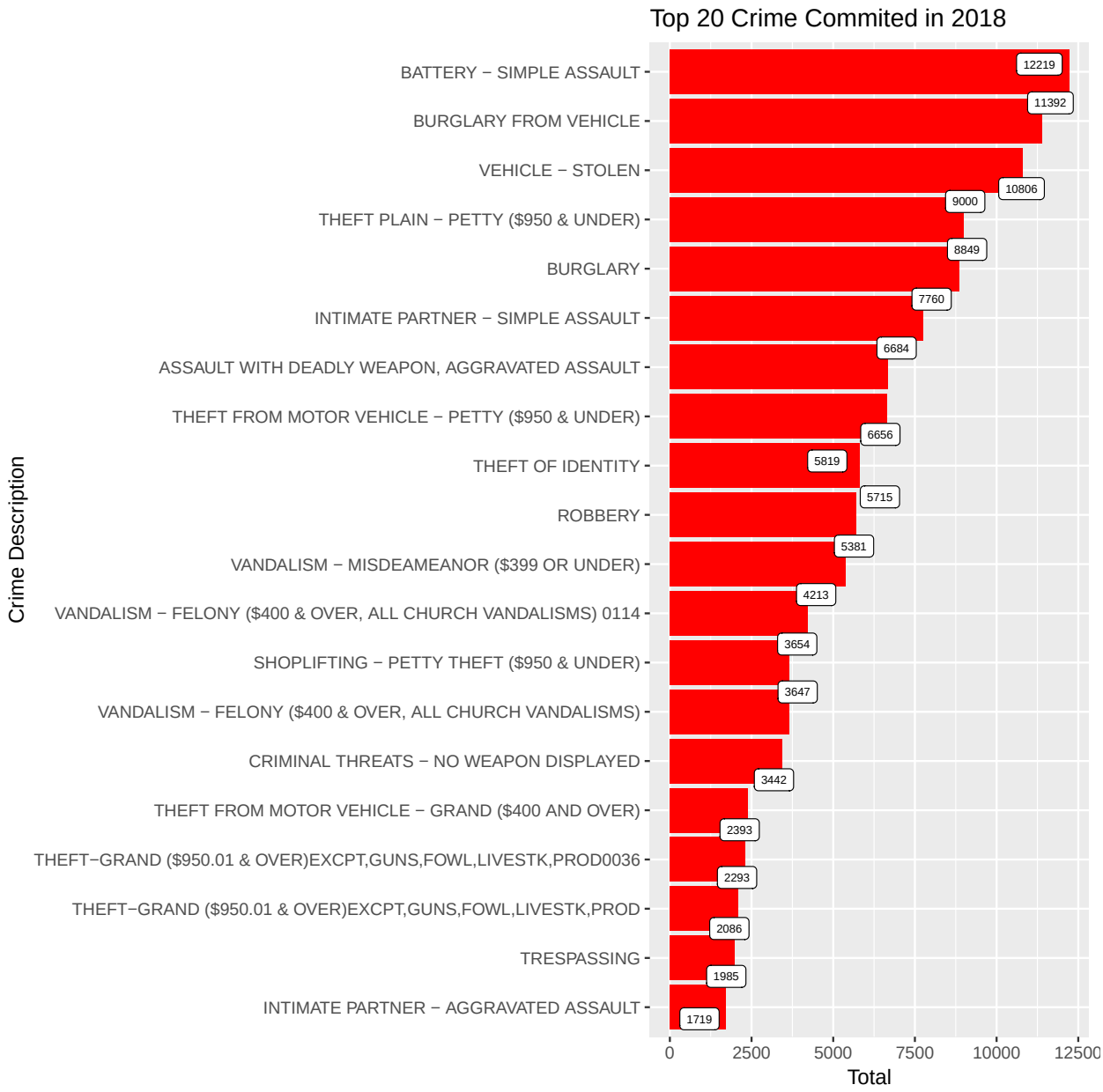
```
year_2017 <- crime_selected_years %>%
  filter(year == "2017")

group <- year_2017 %>%
  group_by(`Crime Code Description`) %>%
  summarise(total = n()) %>%
  distinct() %>%
  top_n(20)

group %>%
  ggplot(aes(reorder(`Crime Code Description`, total), y = total)) +
  geom_col(fill = "red") +
  geom_label_repel(aes(label = total), size = 2) +
  coord_flip() +
  labs(title = "Top 20 Crime Commited in 2017",
       x = "Crime Description",
       y = "Total")
```



How about in year 2018:

```
year_2018 <- crime_selected_years %>%
  filter(year == "2018")

group <- year_2018 %>%
  group_by(`Crime Code Description`) %>%
  summarise(total = n()) %>%
  distinct() %>%
  top_n(20)

group %>%
  ggplot(aes(reorder(`Crime Code Description`, total), y = total)) +
  geom_col(fill = "red") +
  geom_label_repel(aes(label = total), size = 2) +
```

```
  coord_flip() +
  labs(title = "Top 20 Crime Commited in 2018",
       x = "Crime Description",
       y = "Total")
```



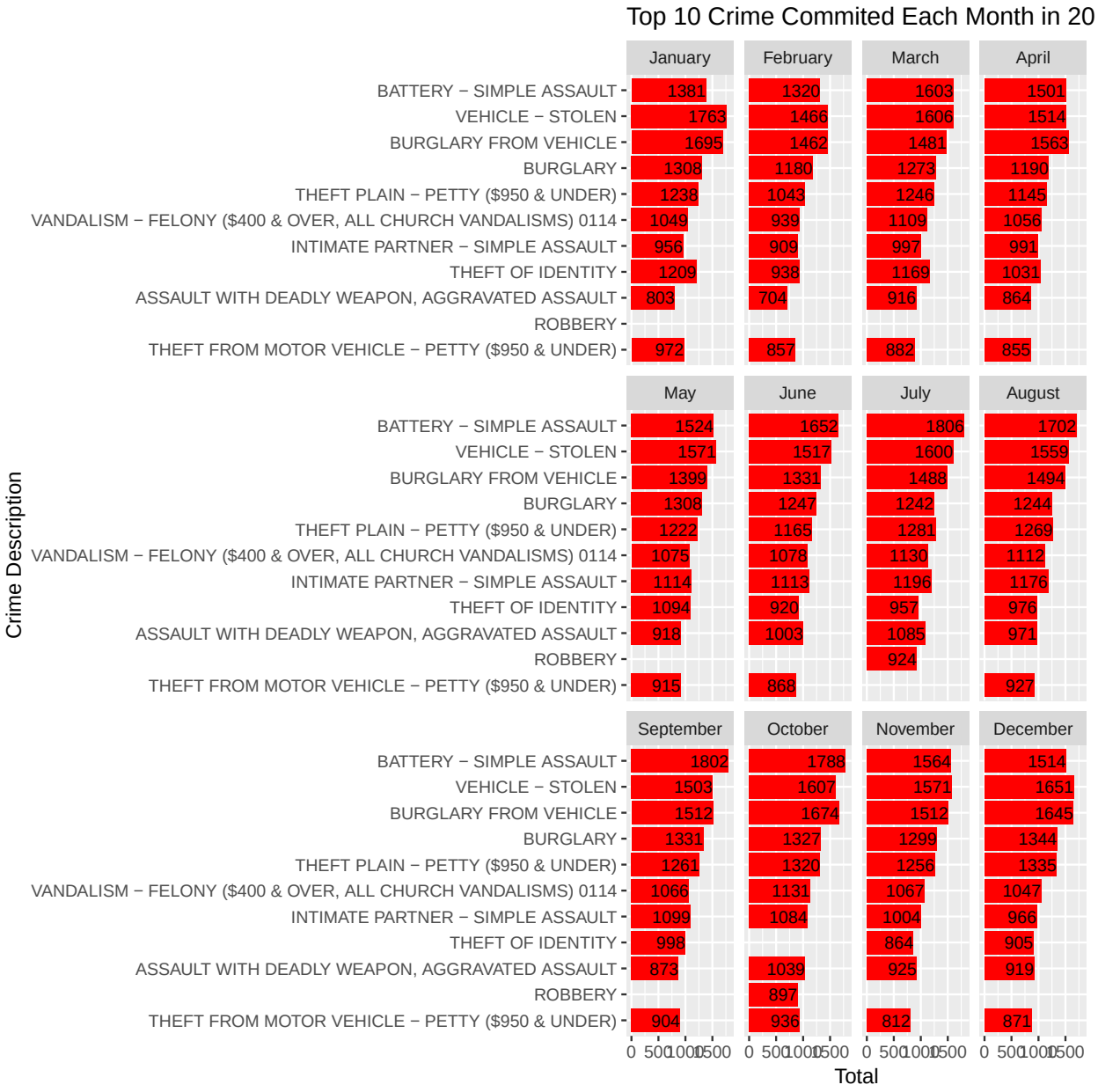
There are some difference of crime commited between 2017 and 2018.

Month

Let see which month have the highest crime.

```
#recode month
year_2017$month <- recode(year_2017$month, "1" = "January", "2" = "February", "3" = "March", "4" = "Apri
```

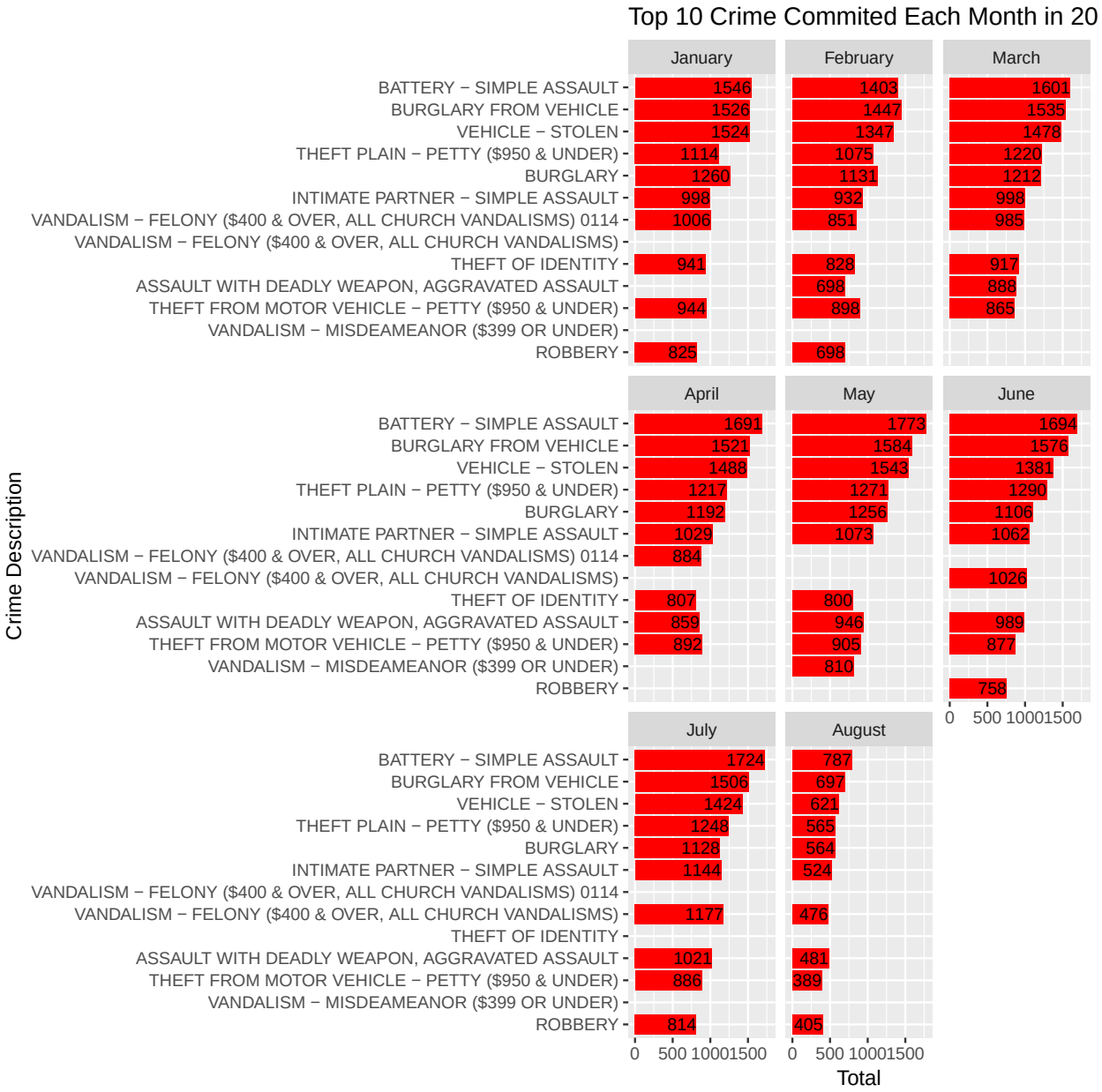
```r
#reorder the factor
year_2017$month <- ordered(year_2017$month, levels = c('January', 'February', 'March', 'April', 'May',

#summary the top 10 crime on eacth month
month <- year_2017 %>%
  group_by(month, `Crime Code Description`) %>%
  summarise(total = n()) %>%
  distinct() %>%
  top_n(10)

month %>%
  ggplot(aes(reorder(`Crime Code Description`, total), y = total)) +
  geom_col(fill = "red") +
  geom_text(aes(label=total), color='black', hjust = 1, size = 3) +
  coord_flip() +
  facet_wrap(~ month) +
  labs(title = "Top 10 Crime Commited Each Month in 2018",
       x = "Crime Description",
       y = "Total")
```



In 2018:

```
#recode month
year_2018$month <- recode(year_2018$month, "1" = "January", "2" = "February", "3" = "March", "4" = "Apri

#reorder the factor
year_2018$month <- ordered(year_2018$month, levels = c('January', 'February', 'March', 'April', 'May',

#summary the top 10 crime on eacth month
month <- year_2018 %>%
  group_by(month, `Crime Code Description`) %>%
  summarise(total = n()) %>%
  distinct() %>%
  top_n(10)
```

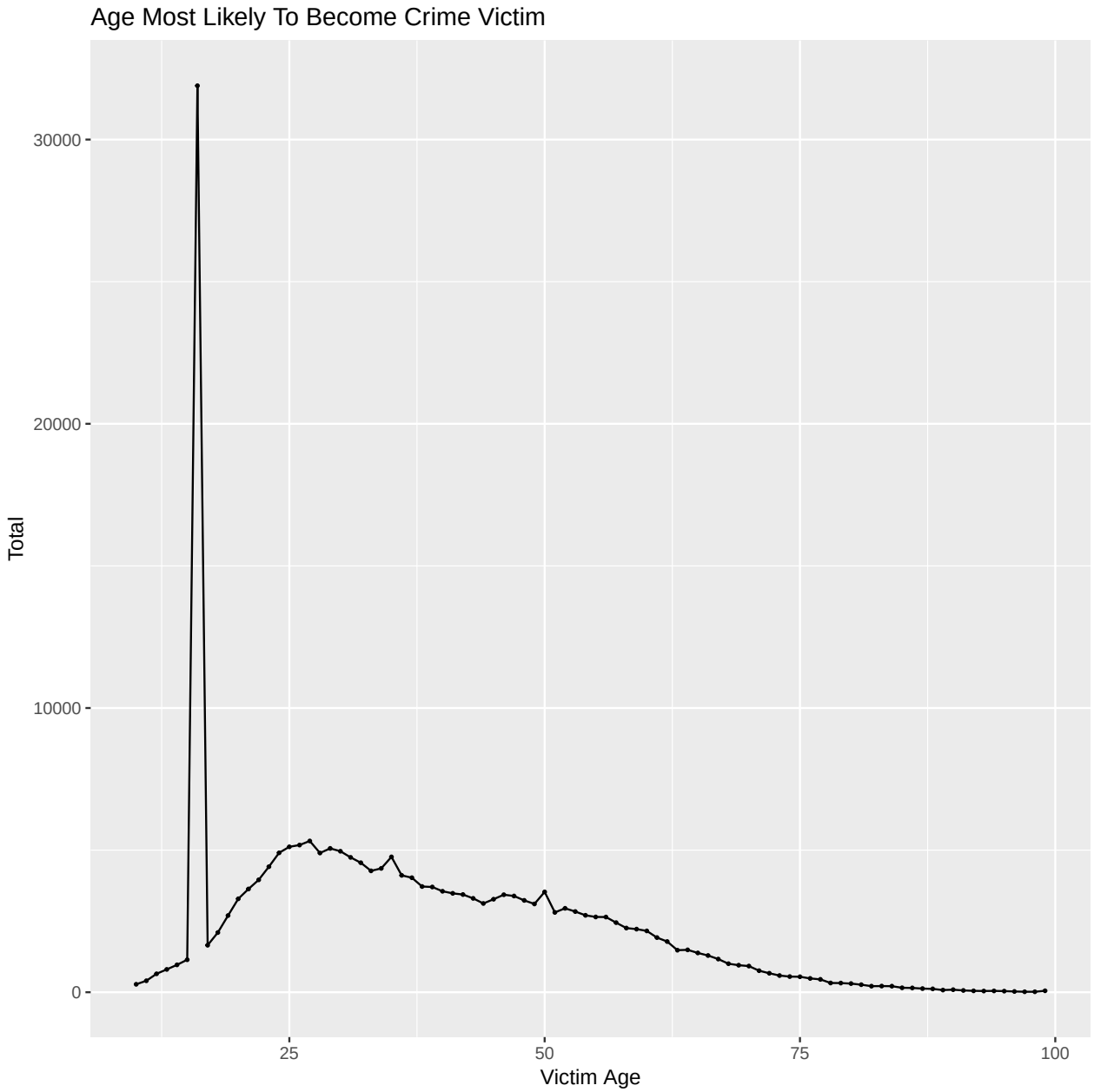
```
month %>%
  ggplot(aes(reorder(`Crime Code Description`, total), y = total)) +
  geom_col(fill = "red") +
  geom_text(aes(label=total), color='black', hjust = 1, size = 3) +
  coord_flip() +
  facet_wrap(~ month) +
  labs(title = "Top 10 Crime Commited Each Month in 2018",
       x = "Crime Description",
       y = "Total")
```



As you can see, there is a increasing of crime in 2018 compared to 2017. Next we are going to examin the age most likely to become victim of crime.

Age group

```
age <- year_2017 %>%
  group_by(`Victim Age`) %>%
  summarise(total = n()) %>%
  na.omit()

age %>%
  ggplot(aes(x = `Victim Age`, y = total)) +
  geom_line(group = 1) +
  geom_point(size = 0.5) +
  labs(title = "Age Most Likely To Become Crime Victim",
       x = "Victim Age",
       y = "Total")
```

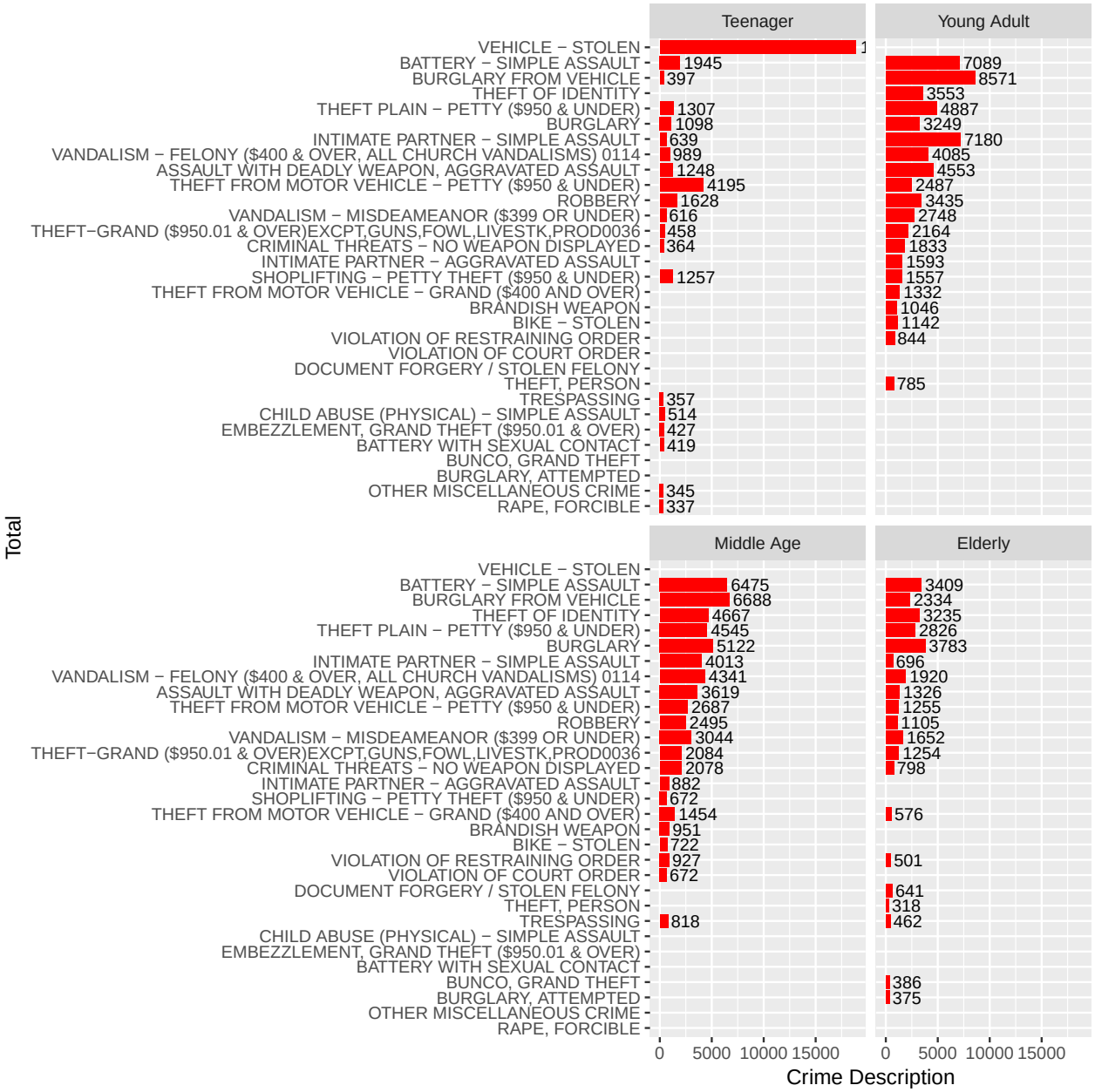


As shown above, the age group below 25 are most likely to become victim of crime in 2017. The huge spike is represent age 16.

Next I'm going to factor the age into different group and examine which crime are targeted to different age group. I going to cut the age group into teenager (10-18), young adult, (19 - 35), middle age (36-55) and elderly (56 above)

```
year_2017$age_group <- cut(year_2017$`Victim Age`, breaks = c(-Inf, 19, 35, 55, Inf), labels = c("Teenag

age.group <- year_2017 %>%
  group_by(age_group, `Crime Code Description`) %>%
  summarise(total = n()) %>%
  top_n(20) %>%
  na.omit()
```

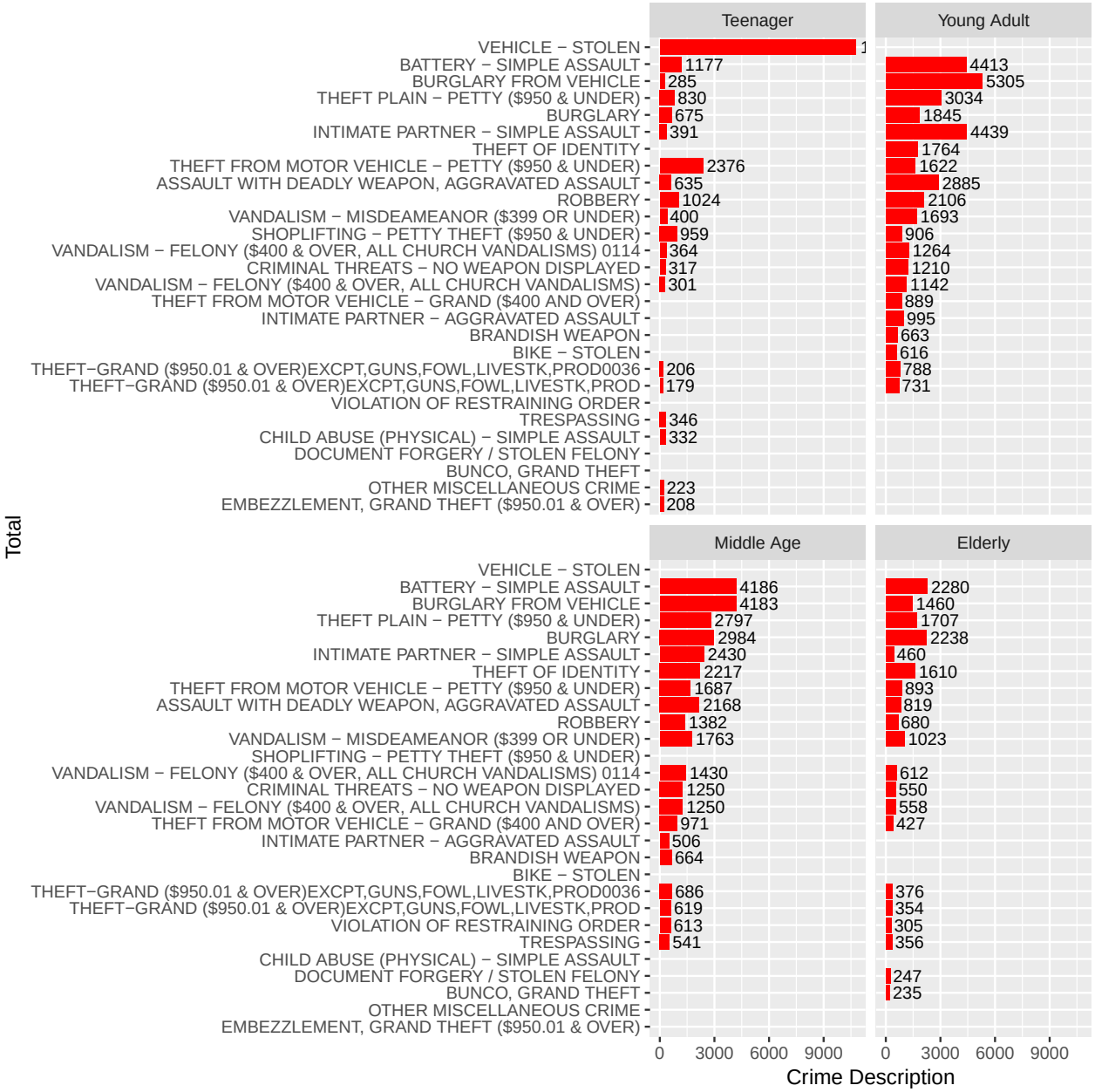
```
age.group %>%
  ggplot(aes(reorder(x = `Crime Code Description`, total), y = total)) +
  geom_col(fill = 'red') +
  geom_text(aes(label=total), color='black', hjust = -0.1, size = 3) +
  coord_flip() +
  facet_wrap(~ age_group) +
  labs(x = 'Total',
       y = "Crime Description")
```



In 2018:

```
year_2018$age_group <- cut(year_2018$`Victim Age`, breaks = c(-Inf, 19, 35, 55, Inf), labels = c("Teenag
```

```r
age.group <- year_2018 %>%
  group_by(age_group, `Crime Code Description`) %>%
  summarise(total = n()) %>%
  top_n(20) %>%
  na.omit()

age.group %>%
  ggplot(aes(reorder(x = `Crime Code Description`, total), y = total)) +
  geom_col(fill = 'red') +
  geom_text(aes(label=total), color='black', hjust = -0.1, size = 3) +
  coord_flip() +
  facet_wrap(~ age_group) +
  labs(x = 'Total',
       y = "Crime Description")
```

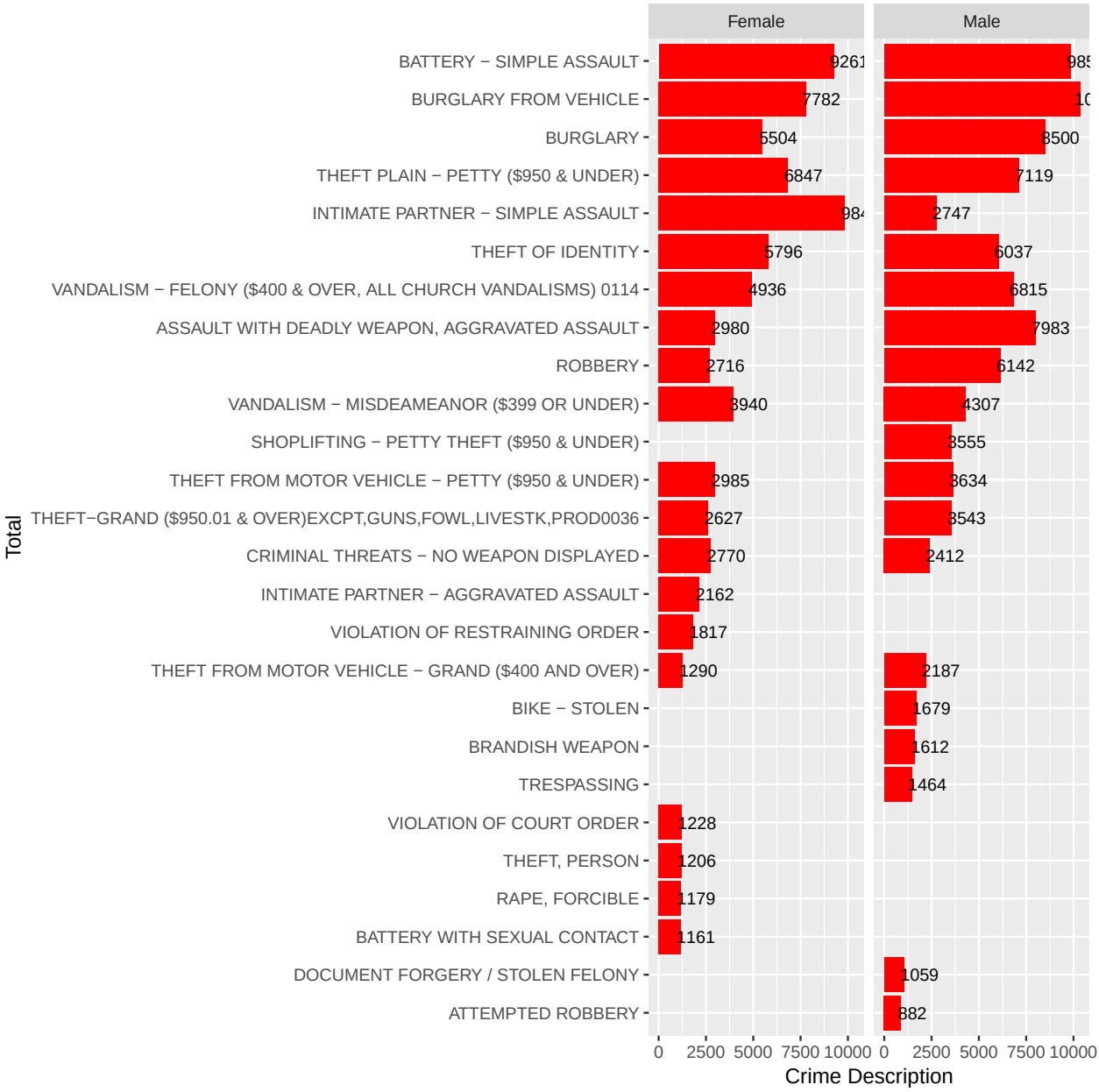


As you can see there are different crime target to different age group. How about gender difference.

Gender

```
gender <- year_2017 %>%
  group_by(`Victim Sex`, `Crime Code Description`) %>%
  summarise(total = n()) %>%
  filter(`Victim Sex` != "Unknown", `Victim Sex` != "H") %>%
  na.omit() %>%
  top_n(20)

gender <- gender[-c(1:30),]
```

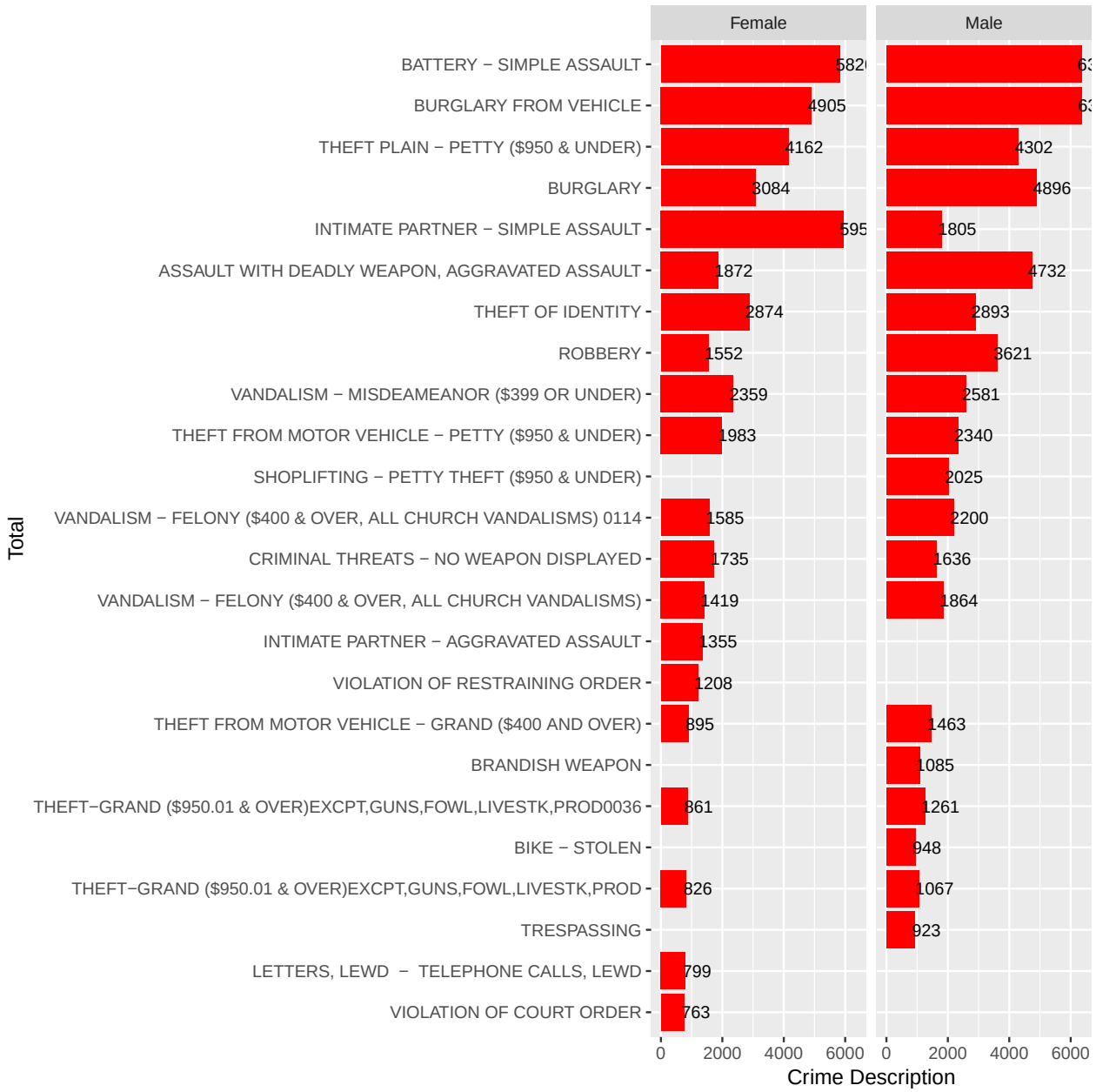
```
gender %>%
  ggplot(aes(reorder(x = `Crime Code Description`, total), y = total)) +
  geom_col(fill = 'red') +
  geom_text(aes(label=total), color='black', hjust = 0.1, size = 3) +
  coord_flip() +
  facet_wrap(~ `Victim Sex`) +
  labs(x = 'Total',
       y = "Crime Description")
```



As you can see both gender are likely to be victim of different kind of crime. Same result can be observe in 2018.

```
gender <- year_2018 %>%
  group_by(`Victim Sex`, `Crime Code Description`) %>%
```

```
  summarise(total = n()) %>%
  filter(`Victim Sex` != "Unknown" & `Victim Sex` != "N", `Victim Sex` != "H") %>%
  na.omit() %>%
  top_n(20)

gender <- gender[-c(1:21),]

gender %>%
  ggplot(aes(reorder(x = `Crime Code Description`, total), y = total)) +
  geom_col(fill = 'red') +
  geom_text(aes(label=total), color='black', hjust = 0.1, size = 3) +
  coord_flip() +
  facet_wrap(~ `Victim Sex`) +
  labs(x = 'Total',
       y = "Crime Description")
```

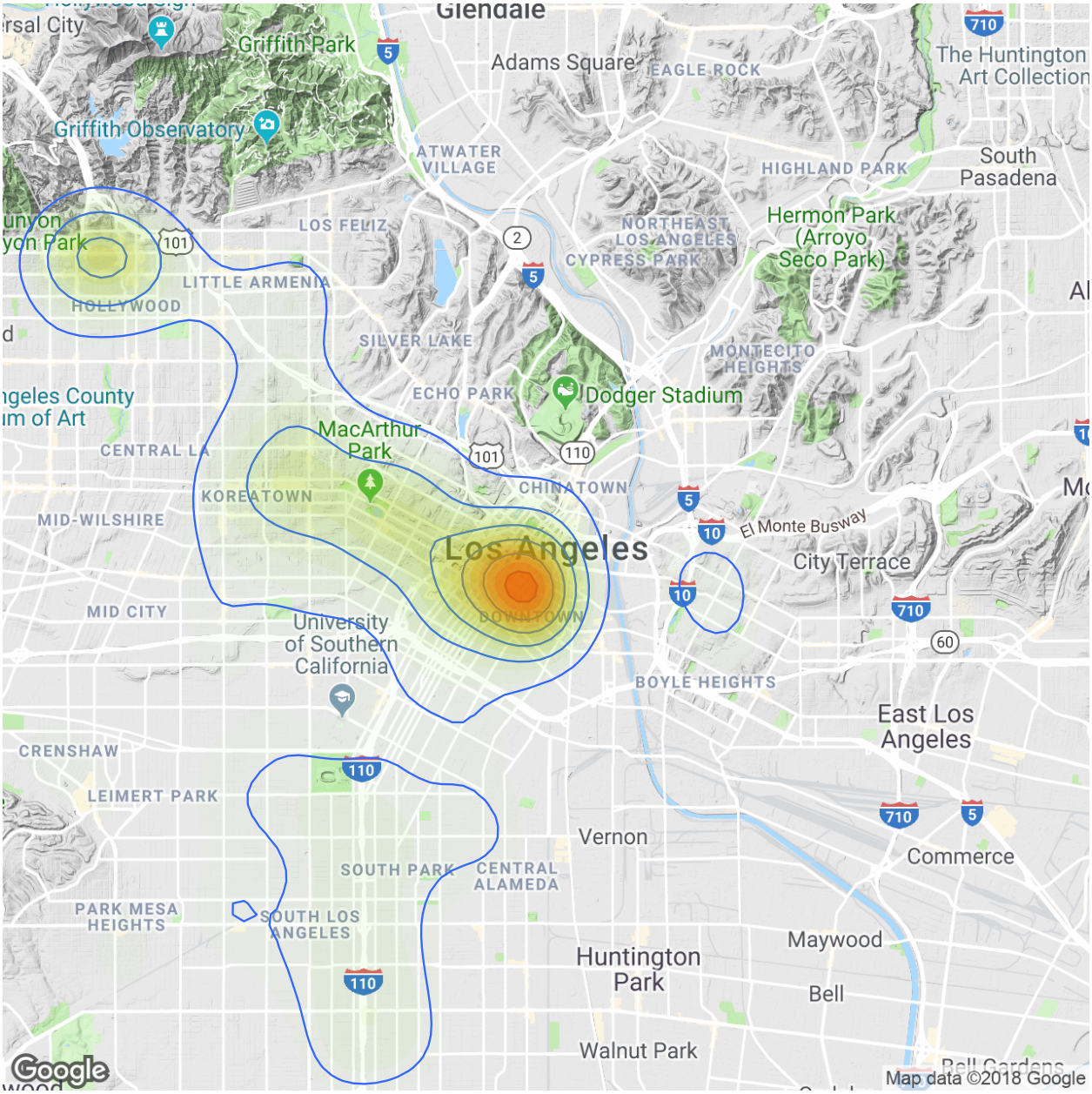


Map The Crime

Next we are going to map the crime. For the illustrate purpose, I'm going to map only the the highest crime commited in 2017 which are assualt and vehicle stolen and in 2018 are assualt and burglary of car.

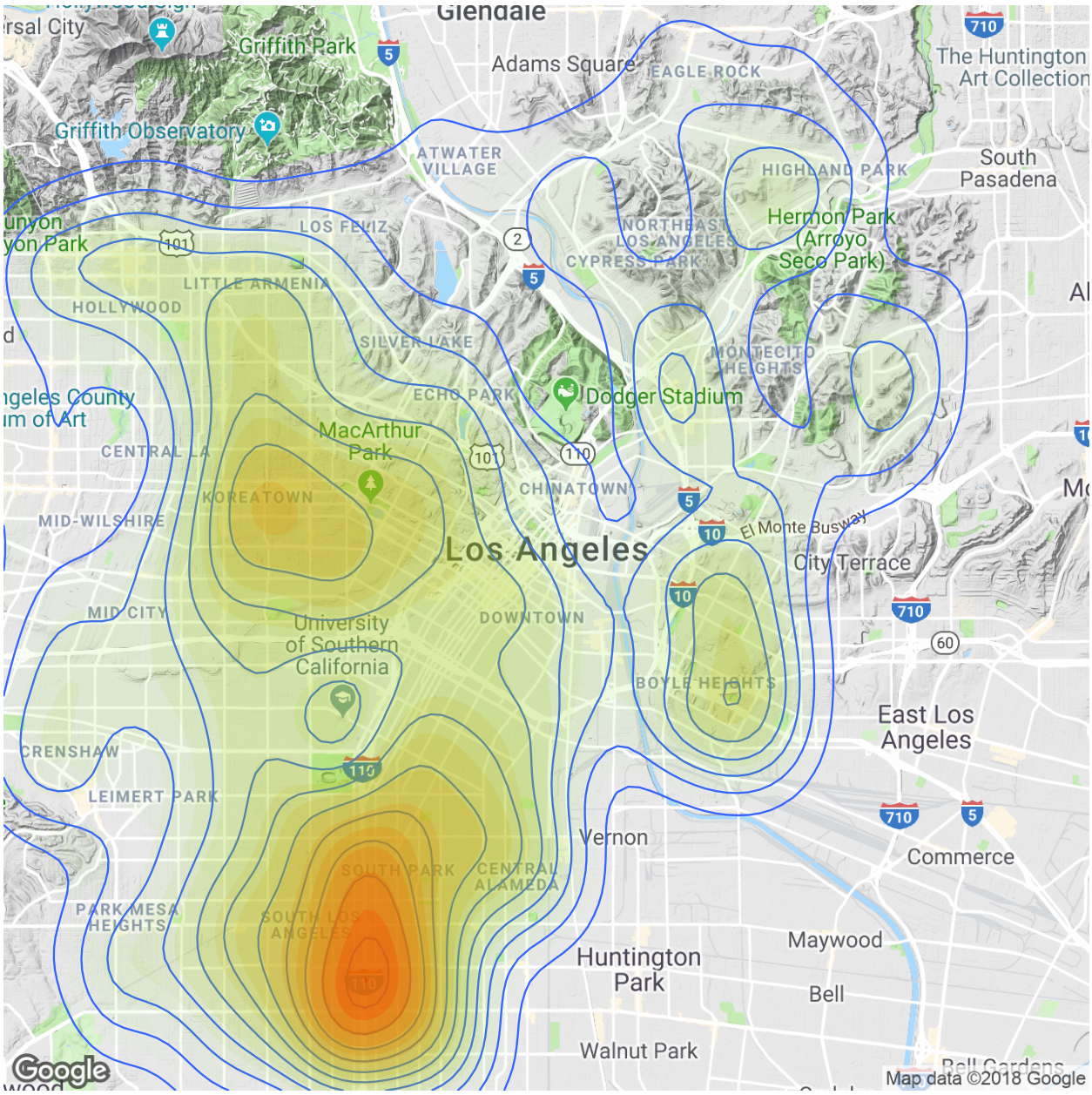
```
#get the map of LA
LA_map <- qmap(location = "Los Angeles", zoom = 12)

#unfactor variable
year_2017$lat <- unfactor(year_2017$lat)
year_2017$long <- unfactor(year_2017$long)

#select relevant variables
```

```
mapping <- year_2017 %>%
  select(`Crime Code Description`, long, lat) %>%
  filter(`Crime Code Description` == 'BATTERY - SIMPLE ASSAULT') %>%
  na.omit()

#mapping
LA_map + geom_density_2d(aes(x = long, y = lat), data = mapping) +
  stat_density2d(data = mapping,
    aes(x = long, y = lat, fill = ..level.., alpha = ..level..), size = 0.01,
    bins = 16, geom = "polygon") + scale_fill_gradient(low = "green", high = "red",
    guide = FALSE) + scale_alpha(range = c(0, 0.3), guide = FALSE)
```



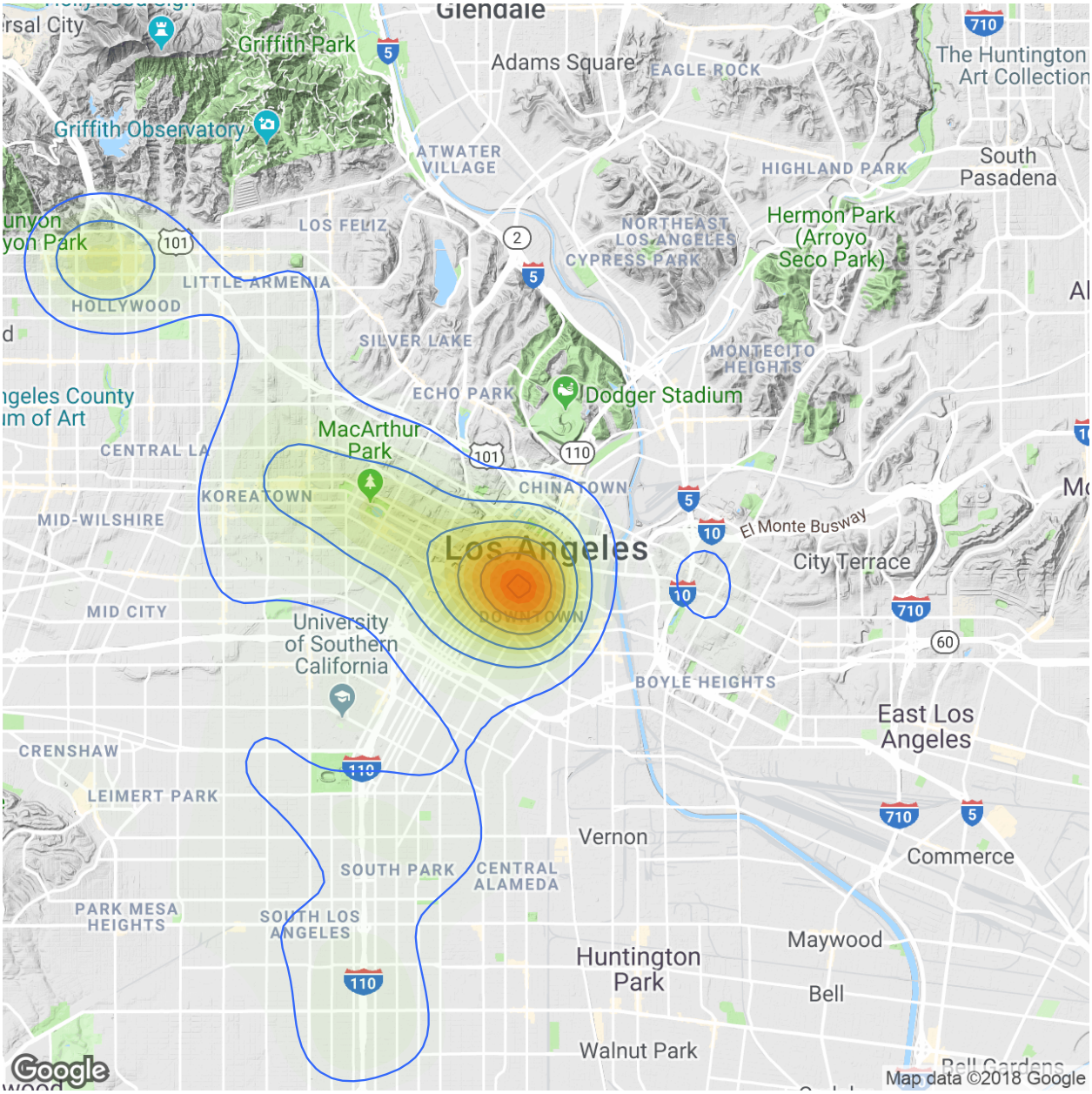
As you can see the battery assault is more likely to happen on Downtown Los Angeles.

```
mapping <- year_2017 %>%
  select(`Crime Code Description`, long, lat) %>%
  filter(`Crime Code Description` == 'VEHICLE - STOLEN') %>%
  na.omit()

LA_map + geom_density_2d(aes(x = long, y = lat), data = mapping) +
  stat_density2d(data = mapping,
    aes(x = long, y = lat, fill = ..level.., alpha = ..level..), size = 0.01,
    bins = 16, geom = "polygon") + scale_fill_gradient(low = "green", high = "red",
    guide = FALSE) + scale_alpha(range = c(0, 0.3), guide = FALSE)
```



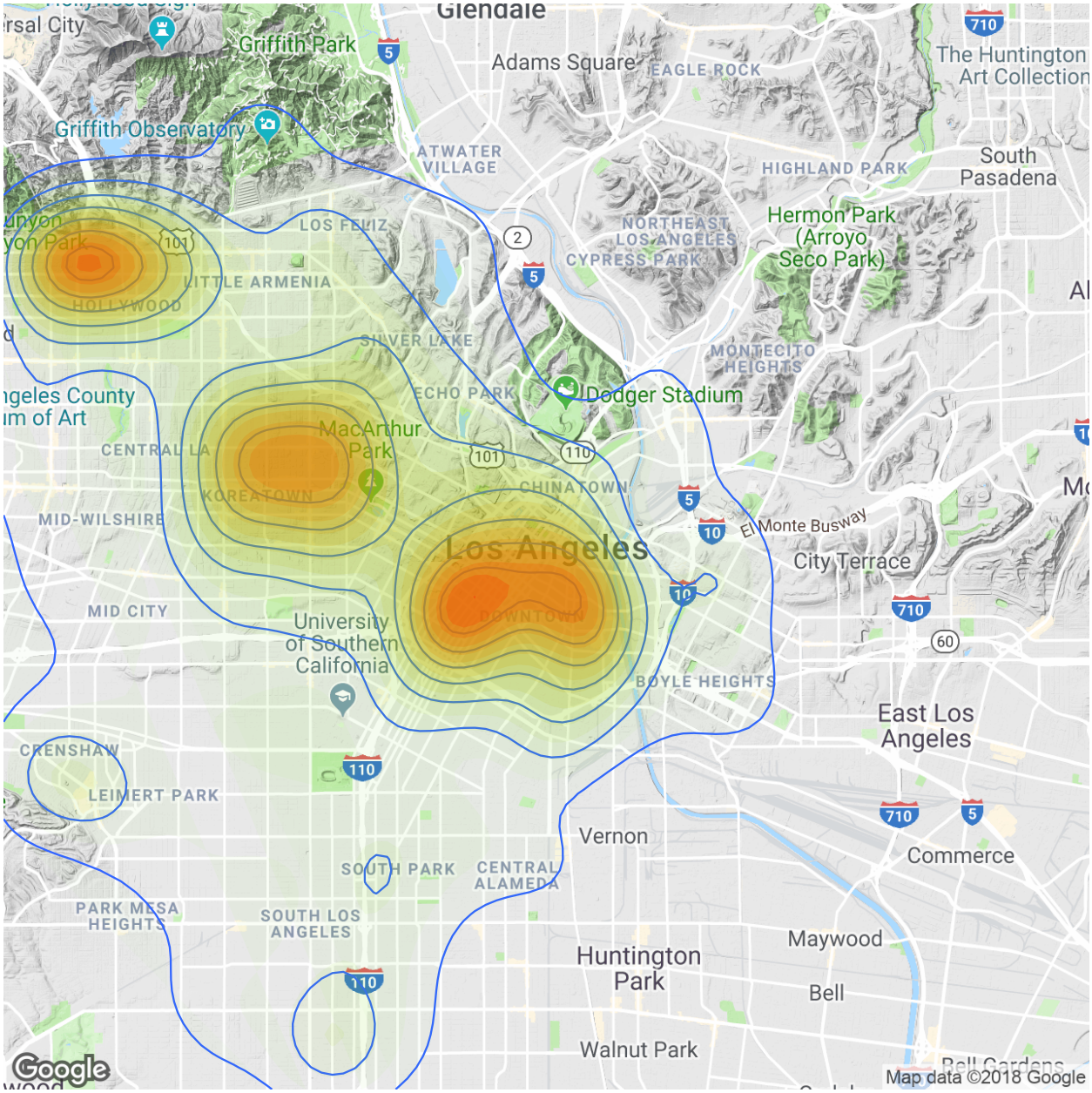
Interestingly, most vehicle are more likely to be stolen on South Los Angeles.

```r
#Convert factor to unmeric
year_2018$lat <- unfactor(year_2018$lat)
year_2018$long <- unfactor(year_2018$long)

#select relevant variables
mapping <- year_2018 %>%
  select(`Crime Code Description`, long, lat) %>%
  filter(`Crime Code Description` == 'BATTERY - SIMPLE ASSAULT') %>%
  na.omit()

#Mapping
LA_map + geom_density_2d(aes(x = long, y = lat), data = mapping) +
  stat_density2d(data = mapping,
    aes(x = long, y = lat, fill = ..level.., alpha = ..level..), size = 0.01,
    bins = 16, geom = "polygon") + scale_fill_gradient(low = "green", high = "red",
    guide = FALSE) + scale_alpha(range = c(0, 0.3), guide = FALSE)
```



In 2018, battery assualt are more likely to be happen in Downtown Los Angeles as well.

```
mapping <- year_2018 %>%
  select(`Crime Code Description`, long, lat) %>%
  filter(`Crime Code Description` == 'BURGLARY FROM VEHICLE') %>%
  na.omit()

LA_map + geom_density_2d(aes(x = long, y = lat), data = mapping) +
  stat_density2d(data = mapping,
    aes(x = long, y = lat, fill = ..level.., alpha = ..level..), size = 0.01,
    bins = 16, geom = "polygon") + scale_fill_gradient(low = "green", high = "red",
    guide = FALSE) + scale_alpha(range = c(0, 0.3), guide = FALSE)
```



However, the burgarly from vehicle are most likely happen in Hollywood, KoreaTown and Downtown Los Angeles.

Conclusion

This is just a simple demonstration of how to gain insight of the data and mapping the crime in Los Angeles.