

LA Arrest Visualization

Tuan Nguyen, Chi Nguyen

8/22/2019

Dependencies	Purposes
data.table	faster read large dataset
tidyverse	dplyr, tidyr, and ggplot
ggmap	maps API
maps	map API toolkits
mapdata	map data
ggrepel	manipulations of layers on maps
varhandle	unfactor function
gridExtra	show 2 maps side by side
options	number digits to display

```
library(data.table)
library(tidyverse)
library(ggmap)
library(maps)
library(mapdata)
library(ggrepel)
library(ggpubr)
library(varhandle)
library(gridExtra)
library(ggthemes)
options(digits=16, warn=-1)
ggmap::register_google(key='AIzaSyDbbC0QGOiirHtBsIXaehDgsxZdJVa_d0s')
```

1. Read the relevant dataset into a data frame
2. View the data frame

```
# Copy the dataset to a data frame
data_path = '~/Arrest-data-from-2010-to-present.csv'
la_arrests <- as.data.frame(fread(data_path))
View(la_arrests)
```

1. Select relevant variables in the dataset
2. Clean and wrangle data (variable names, unfilled values, latittue, longitude, date)
3. Factor variable \$Charge_Group_Description
4. Drop messy variable \$Location
5. Review the complete dataset

```

# Select relevant cols in the dataset
df <- select(la_arrests, `Arrest Date`, `Time`, `Age`, `Area Name`, `Sex Code`,
             `Charge Group Description`, `Arrest Type Code`, `Location`)

# Rename the multi-word variables so that each of them does not have space
setnames(df,
         old=c('Arrest Date', 'Area Name', 'Sex Code',
               'Charge Group Description', 'Arrest Type Code'),
         new=c('Arrest_Date', 'Area_Name', 'Sex_Code',
               'Charge_Group_Description', 'Arrest_Type_Code')))

# Extract latitude as double format from Location of the data frame
# Convert $Arrest_Date to Date format
df <- transform(df,
                Latitude = as.double(str_sub(word(Location, 2, 2), 2, -3)),
                Longitude = as.double(str_sub(word(Location, -1, -1), 2, -3)),
                Arrest_Date = as.Date(str_sub(df$Arrest_Date, 1, 10)))

# Fill empty values in Charge_Group_Description variable with 'Unknown'
df$Charge_Group_Description[df$Charge_Group_Description==''] <- 'Unknown'

# Factor and sort $Charge_Group_Description
df <- within(df, Charge_Group_Description
              <- factor(Charge_Group_Description,
                        levels=names(sort(table(Charge_Group_Description)))))

# Remove Location off the data frame
df <- select(df, -Location)
# Inspect the data frame
View(df)
glimpse(df)

```

```

## Observations: 1,276,160
## Variables: 9
## $ Arrest_Date           <date> 2019-06-22, 2019-06-22, 20...
## $ Time                  <int> 1630, 1010, 400, 302, 1240, 800, 40, 83...
## $ Age                   <int> 44, 8, 31, 23, 28, 13, 31, 40, 20, 14, ...
## $ Area_Name              <chr> "Pacific", "West Valley", "N Hollywood"...
## $ Sex_Code               <chr> "M", "M", "F", "F", "M", "M", "M", ...
## $ Charge_Group_Description <fct> Miscellaneous Other Violations, Unknown...
## $ Arrest_Type_Code       <chr> "M", "O", "M", "M", "F", "D", "F", "M", ...
## $ Latitude               <dbl> 33.9920, 34.1687, 34.1649, 34.2692, 33...
## $ Longitude              <dbl> -118.4201, -118.5579, -118.3965, -118.4...

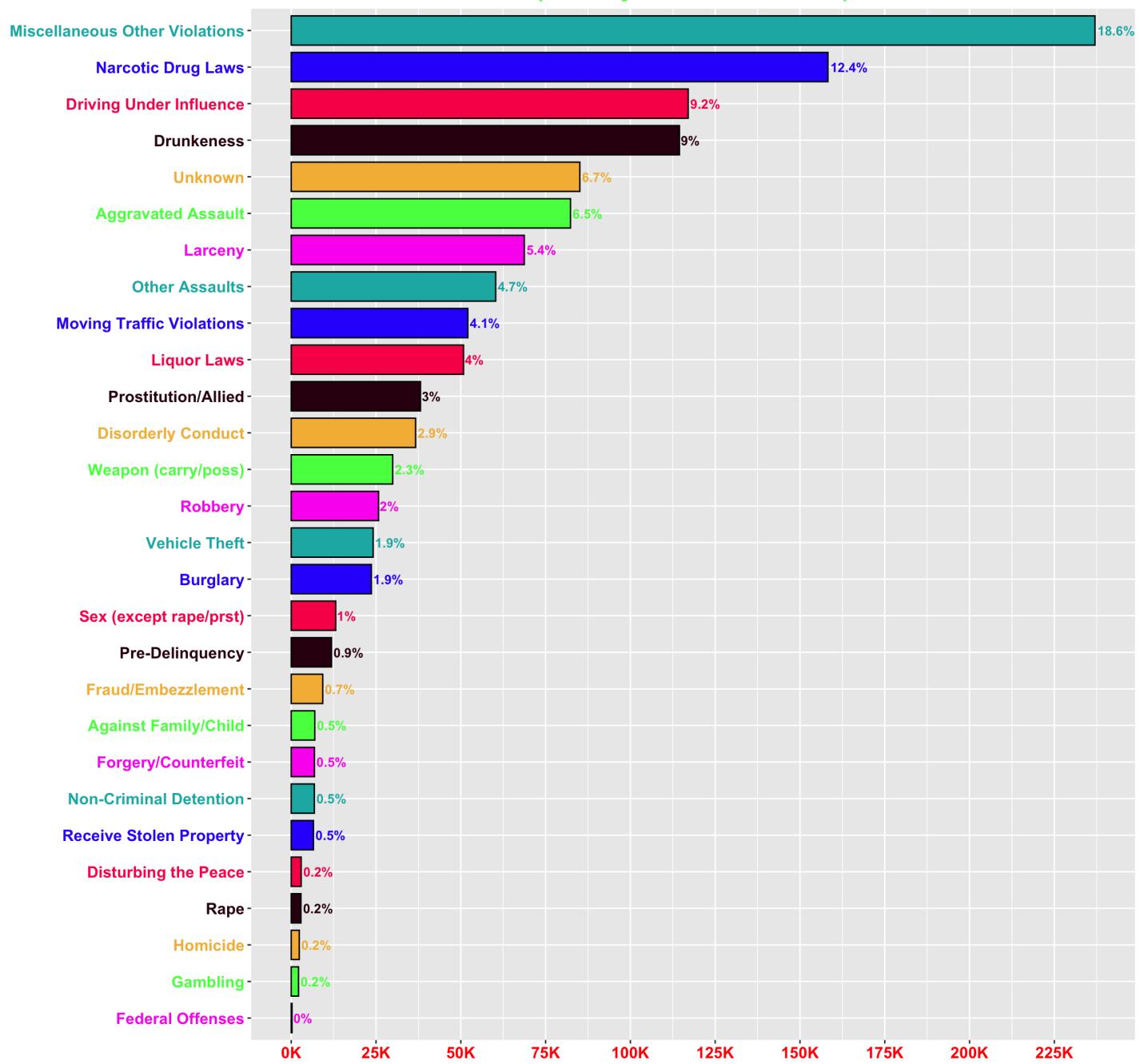
```

- Plot a bar chart for \$Charge_Group_Description in presentation of count and percentage in the same chart

```
p_colors <- c('#FC0BF0', '#52FD4D', '#F4B942', '#350113', '#F90359', '#320EFC', '#18B4B2')
p_colors <- c(rep(p_colors, 4))
breaks <- seq(0, 225000, 25000)
labels = paste0(seq(0, 225, 25), 'K')

ggplot(df, aes(x=Charge_Group_Description)) +
  geom_bar(width=0.8, color='black', fill=p_colors) +
  geom_bar(width=0.8, mapping=aes(x=Charge_Group_Description, y=..prop.., group=1),
            stat='count', fill=p_colors) +
  geom_text(aes(label=paste0(round(stat(..prop..)*100, 1), '%'), group=1),
            stat='count', size=3.5, hjust=-0.07, color=p_colors, fontface='bold') +
  scale_y_continuous(breaks = breaks, labels=labels) +
  coord_flip() +
  labs(title='Los Angeles Arrests By Year\n(January, 2010 - June, 2019)',
       caption='Source: https://data.lacity.org') +
  theme(axis.text.y=element_text(color=p_colors, size=12, face='bold'),
        axis.text.x=element_text(color='red', size=12, face='bold'),
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        plot.title=element_text(size=18, color='#52FD4D', face='bold', hjust=0.5))
```

Los Angeles Arrests By Year (January, 2010 - June, 2019)

Source: <https://data.lacity.org>

```

df_m <- subset(df, Sex_Code=='M', select=Charge_Group_Description)
df_m <- sort(table(df_m), decreasing=TRUE)
df_f <- subset(df, Sex_Code=='F', select=Charge_Group_Description)
df_f <- sort(table(df_f), decreasing=TRUE)
df_mf <- data.frame(Charge=character(), Gender=character(),
                     Counts=integer(), stringsAsFactors=FALSE)
charge_levels <- c()
for (i in 1:length(names(df_m))) {
  df_mf[nrow(df_mf) + 1, ] <- list(names(df_m)[i], 'Male', df_m[i])
  charge_levels[i] <- names(df_m)[i]
}
for (i in 1:length(names(df_f)))
  df_mf[nrow(df_mf) + 1, ] <- list(names(df_f)[i], 'Female', -1*df_f[i])
df_mf$Charge <- factor(df_mf$Charge, levels=charge_levels)
glimpse(df_mf)

```

```

## Observations: 56
## Variables: 3
## $ Charge <fct> Miscellaneous Other Violations, Narcotic Drug Laws, Drunk...
## $ Gender <chr> "Male", "Male", "Male", "Male", "Male", "Male", "Male", ...
## $ Counts <dbl> 201285, 128251, 101895, 92528, 65678, 63542, 47252, 44441...

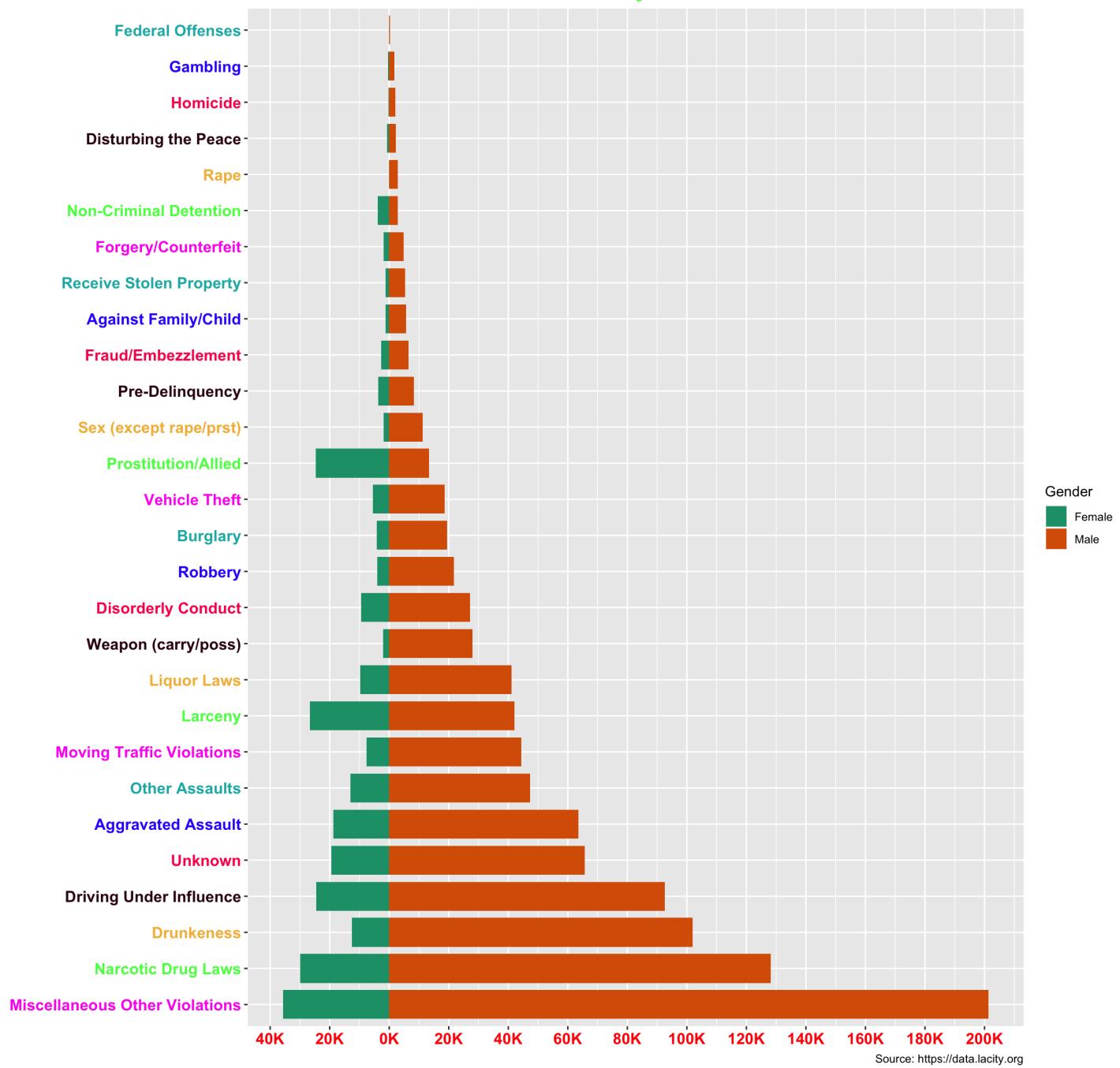
```

```

breaks <- seq(-200000, 200000, 20000)
labels = paste0(as.character(c(seq(200, 0, -20), seq(20, 200, 20))), 'K')
ggplot(df_mf, aes(x=Charge, y=Counts, fill=Gender)) +
  geom_bar(stat='identity', width=0.8) +
  scale_y_continuous(breaks=breaks, labels=labels) +
  coord_flip() +
  labs(title='Distributions of Los Angeles Arrests Among Male and Female',
       From January, 2010 to June, 2019',
       caption='Source: https://data.lacity.org') +
  theme(axis.text.y=element_text(color=p_colors, size=12, face='bold'),
        axis.text.x=element_text(color='red', size=12, face='bold'),
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        plot.title=element_text(size=18, color='#52FD4D', face='bold', hjust=0.5)) +
  scale_fill_brewer(palette='Dark2')

```

Distributions of Los Angeles Arrests Among Male and Female From January, 2010 to June, 2019

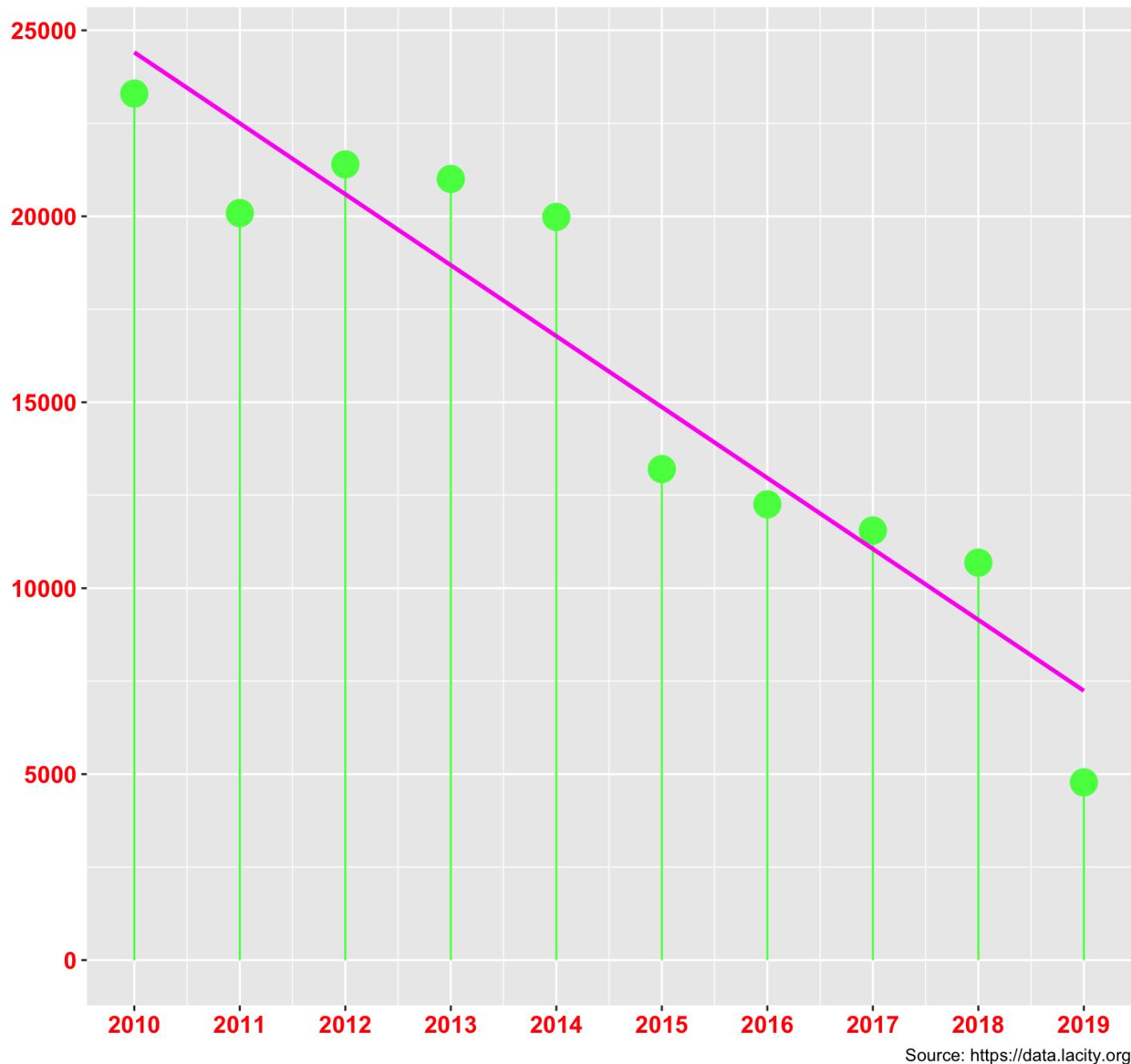
Source: <https://data.lacity.org>

* Plot a bar chart for drug violations by year

```
drug_by_year <- data.frame(Year=integer(), Counts=double())
year_span <- 2010:2019
for (year in year_span) {
  cnts <- nrow(select(filter(df, Charge_Group_Description=='Narcotic Drug Laws',
                           as.numeric(format(Arrest_Date, '%Y'))==year)))
  drug_by_year[nrow(drug_by_year) + 1, ] = list(year, cnts)
}

ggplot(drug_by_year, aes(x=Year, y=Counts)) +
  geom_point(size=6, color='#52FD4D') +
  geom_segment(aes(x=Year, xend=Year,
                    y=0, yend=Counts), color='#52FD4D') +
  scale_x_continuous(breaks=year_span) +
  geom_smooth(method='lm', se=FALSE, color='#FC0BF0') +
  labs(title='Los Angeles Drug Violations By Year\n(January, 2010 - June, 2019)',
       caption='Source: https://data.lacity.org') +
  theme(axis.text.y=element_text(color='red', size=12, face='bold'),
        axis.text.x=element_text(color='red', size=12, face='bold'),
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        plot.title=element_text(color='#52FD4D', face='bold', hjust=0.5))
```

Los Angeles Drug Violations By Year (January, 2010 - June, 2019)

Source: <https://data.lacity.org>

- Function to plot a distribution map for drug violations

```

map_plot <- function(dataset, center) {
  m <- ggmap(get_googlemap(center=c(lon=center[1], lat=center[2]),
                           maptype='roadmap', scale=2, zoom=10))
  m +
    geom_point(aes(x=Longitude, y=Latitude),
                data=dataset, color='#F90359', size=1) +
    stat_density2d(data=dataset,
                   aes(x=Longitude, y=Latitude, alpha=..level..),
                   size=0.01, bins=8, geom='polygon') +
    scale_fill_gradient(low='#18B4B2', high='#F90359') +
    scale_alpha(range=c(0.5, 0.8), guide=FALSE)
}

```

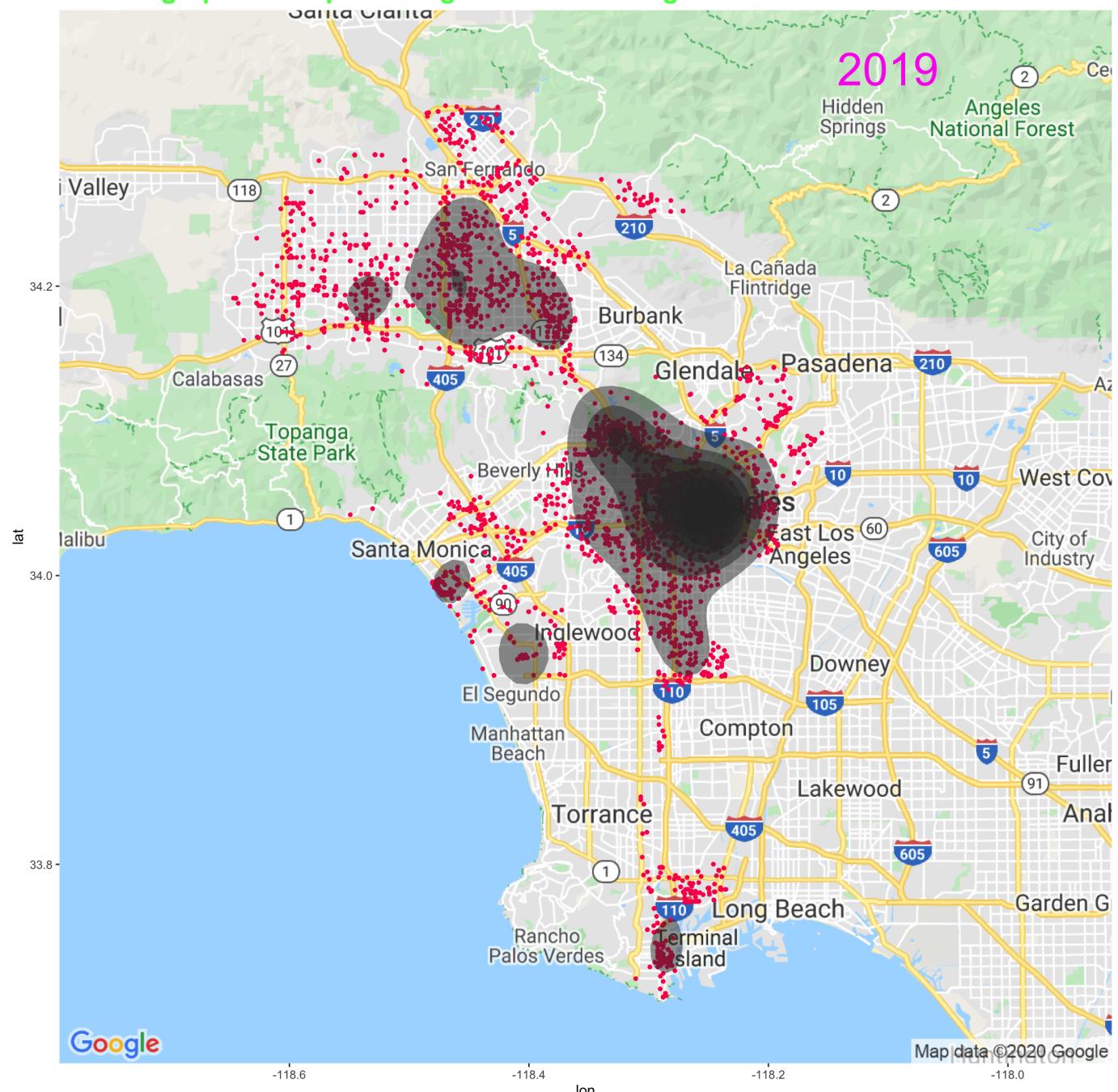
- Plot a geographical map of drug violations:

1. Create a list of data frames which contains the terms ‘Narcotic Drug Laws’ and years of interest
 2. Feed the each of data frames in the list in the plot function to draw its map

```
drugs <- list()
d_plot <- list()
for(year in year_span) {
  drugs[[year]] <- select(filter(df, Charge_Group_Description=='Narcotic Drug Laws',
                                str_sub(Arrest_Date, 1, 4)==as.character(year)),
                         `Latitude`, `Longitude`)
  d_plot[[year]] <- map_plot(drugs[[year]], c(-118.3533783, 34.0274463)) +
    annotate('text', x=-118.1, y=34.35,
             label=year, color='#FC0BF0', size=12)
}
```

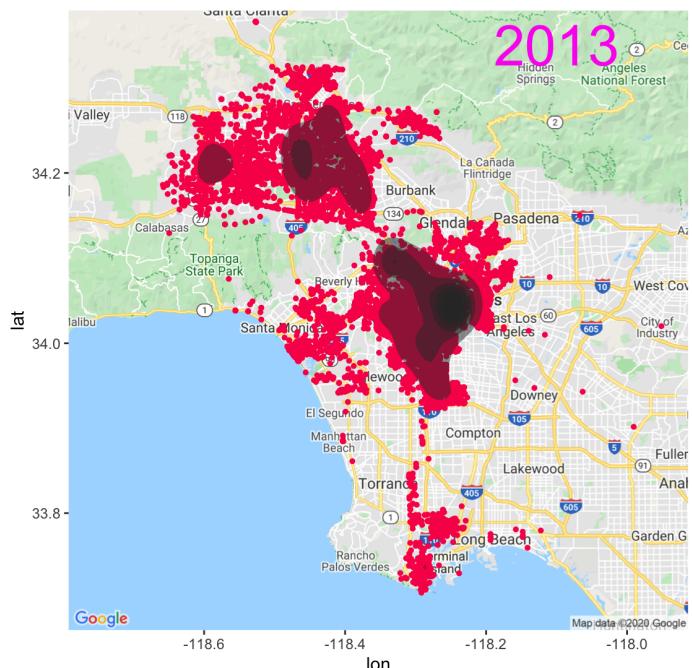
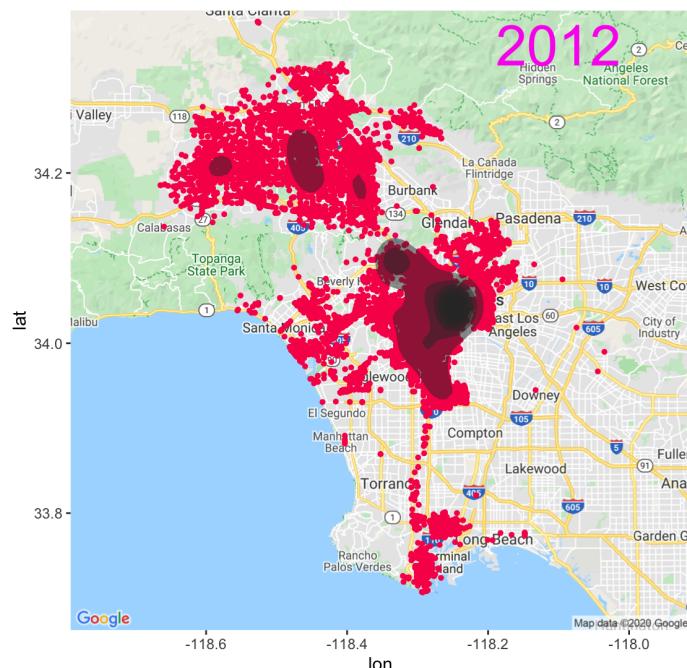
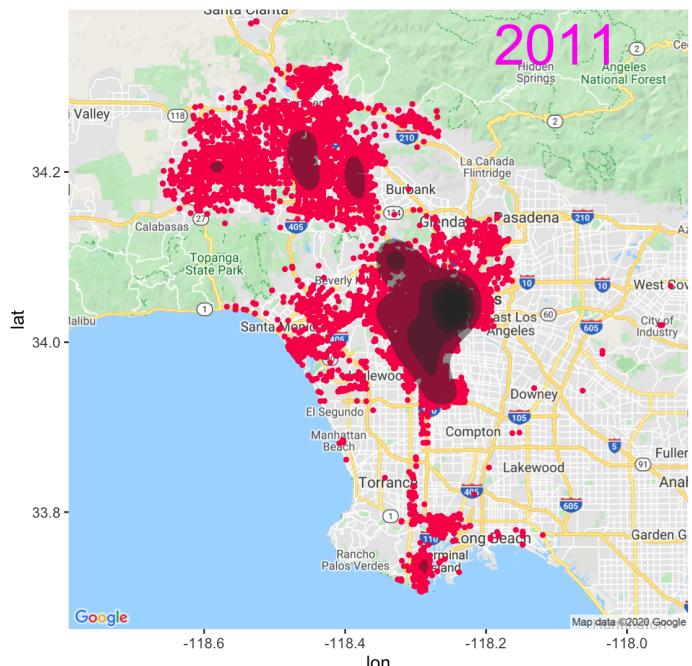
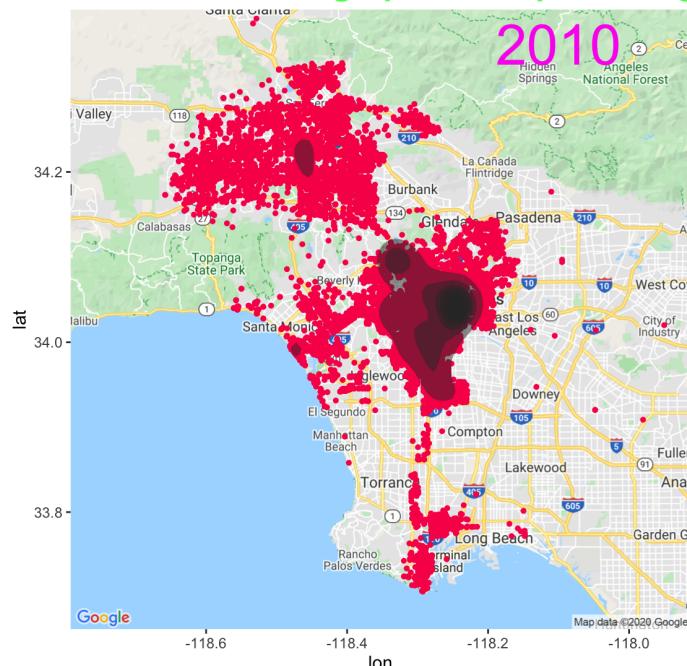
```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=34.027446,-118.353378&zoom=10&size=640x640&scale=2&maptype=roadmap&key=xxx
```

Geographical maps of drug violations during the First 6 Months of 2019



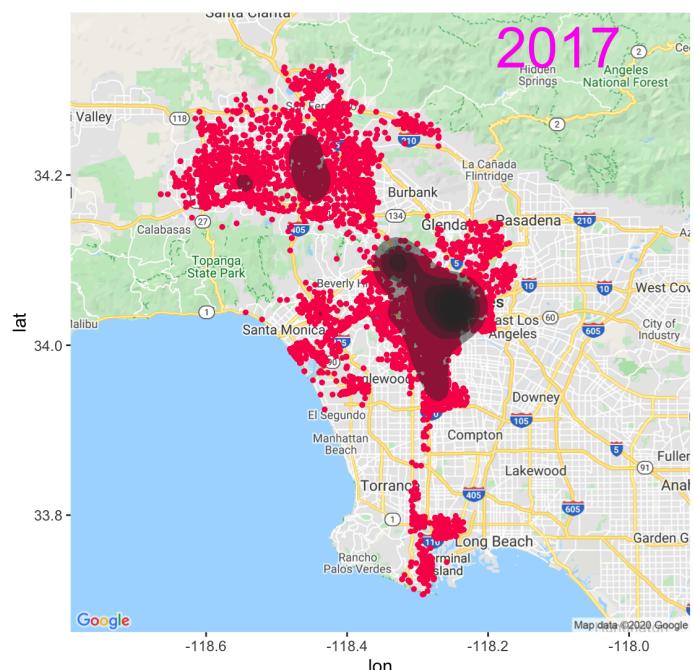
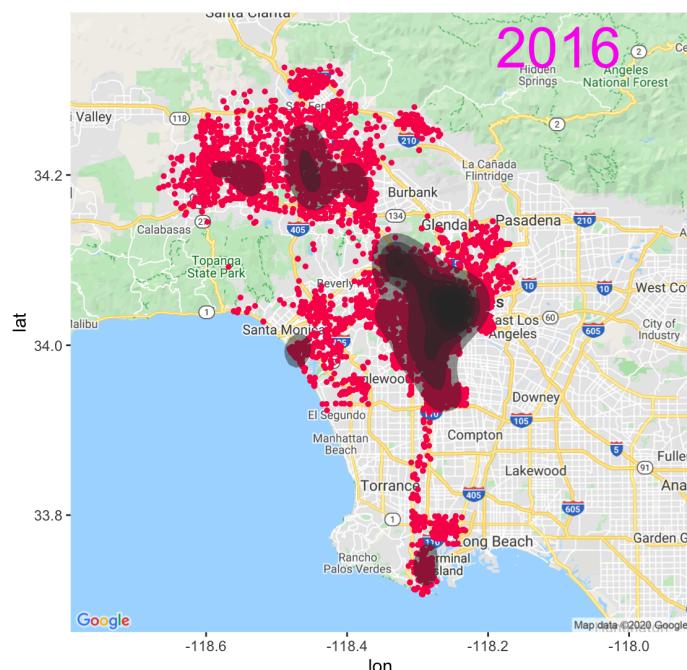
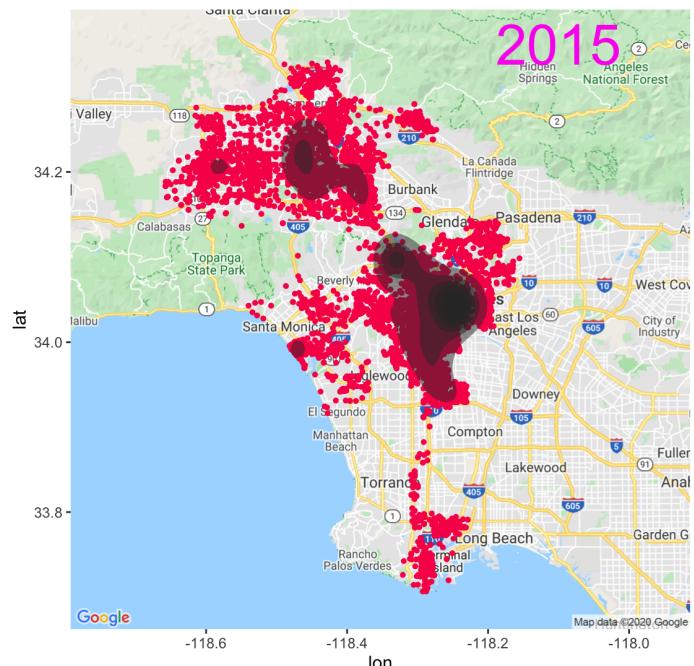
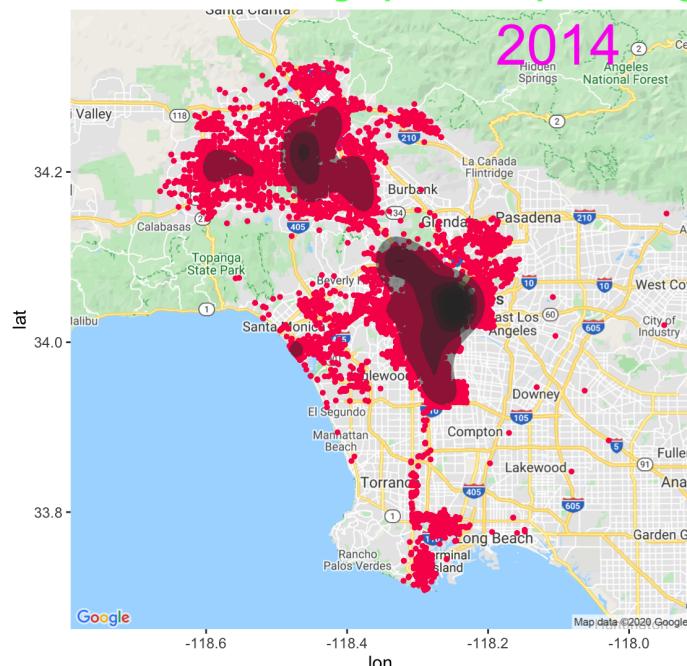
Source: <https://data.lacity.org>

Geographical maps of drug violations from 2010 to 2013



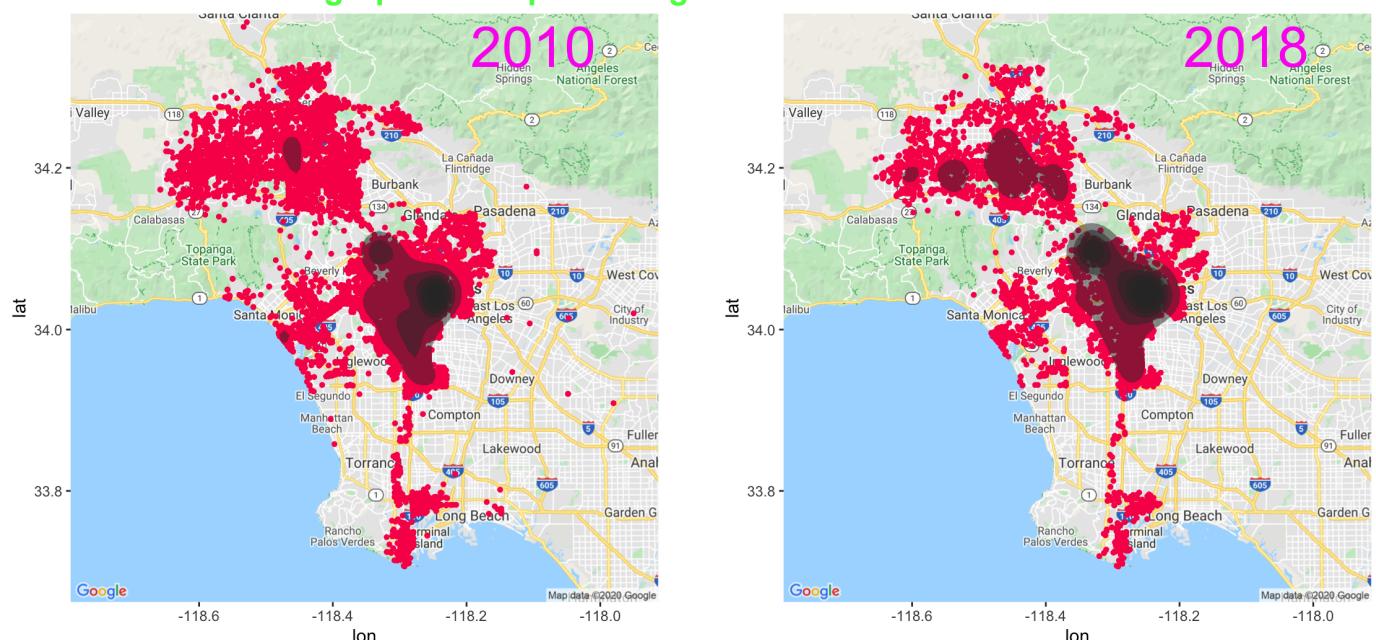
Source: <https://data.lacity.org>

Geographical maps of drug violations from 2014 to 2017



Source: <https://data.lacity.org>

Geographical maps of drug violations in 2010 and 2018



Source: <https://data.lacity.org>