

Note 1-Density map

Tiancheng hou

November 22, 2018

Hello!!! :) This is my first Note. I just drew a density map using the ggmap package, I want to record what I learned. And I hope my notes can help those in need. Ok, first, load all the packages we need.

```
library(ggplot2)
library(tidyverse)
library(ggmap)
library(dplyr)
library(forcats)
```

Load my data.

```
crime <- read.table("crime.csv", sep = ",", header = T)
mean(crime$Latitude)
```

```
## [1] 34.06228
```

```
mean(crime$Longitude)
```

```
## [1] -118.3396
```

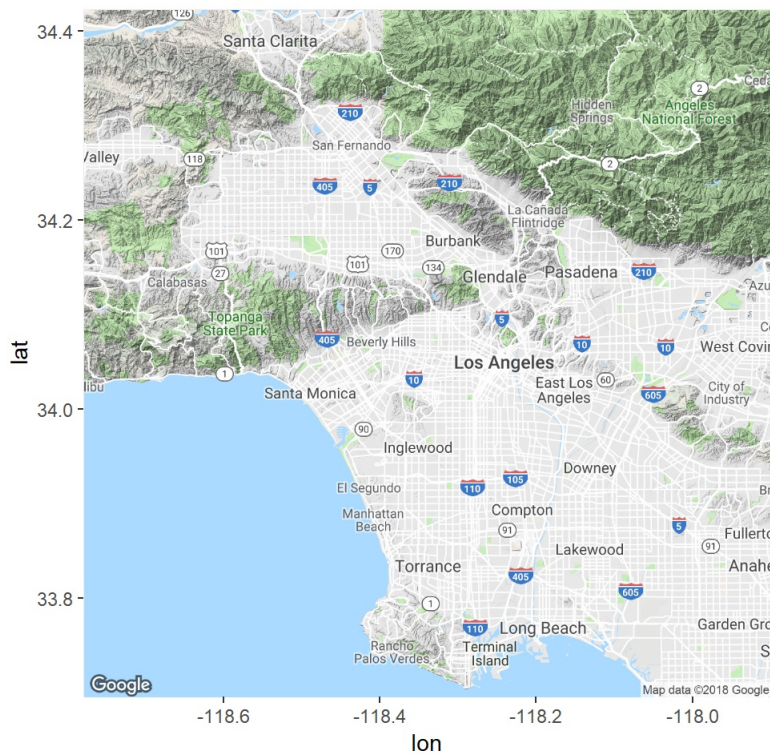
The reason for calculating the average of latitude and longitude is because I want to find out a rough position where all the crimes occur. Once we know it, we can set the center of the map by simply setting the argument "location" equal to c(lon = -118.34, lat = 34.06). There is another argument called "zoom" that I need to adjust. If we are drawing a city map, then 10 is the appropriate value. However, if we want to map a country, we could set the value smaller.

```
#register_google("API-key")##Here is my API-key, replace it with your new API-key. To get one from "https://console.cloud.google.com/google/maps-apis". Select"Maps Static API". Remember to set billing account to activate.
```

```
googlemap <- get_map(location = c(lon = -118.34, lat = 34.06), zoom = 10,
  scale = "auto", maptype = c("terrain"))
```

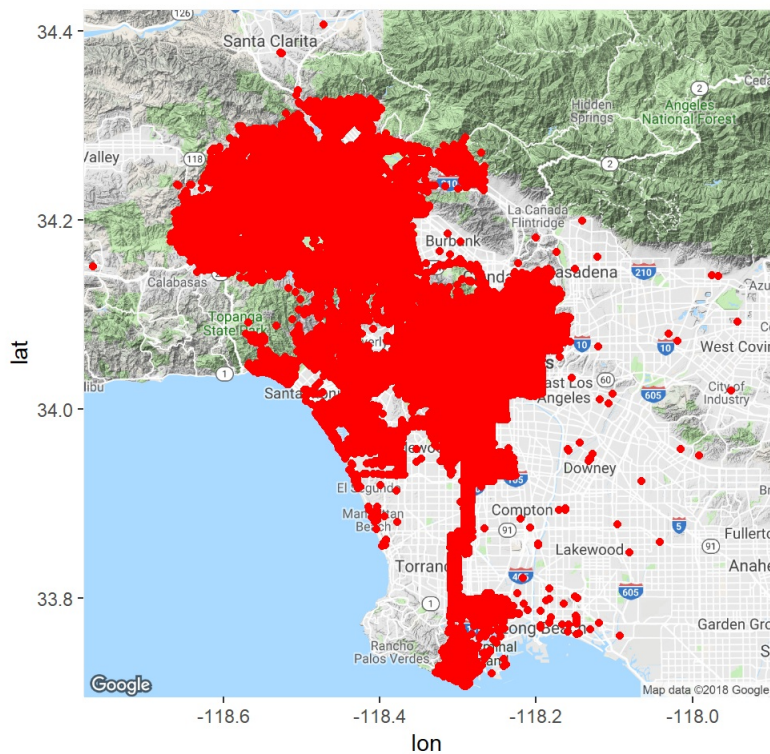
```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=34.06,-118.34&zoom=10&size=640x640&scale=2&maptype=terrain&language=en-EN&key=AIzaSyB1H982W0V1XFTFSbcMP-7AOVsNxR12py8
```

```
a <- ggmap(googlemap)
a
```



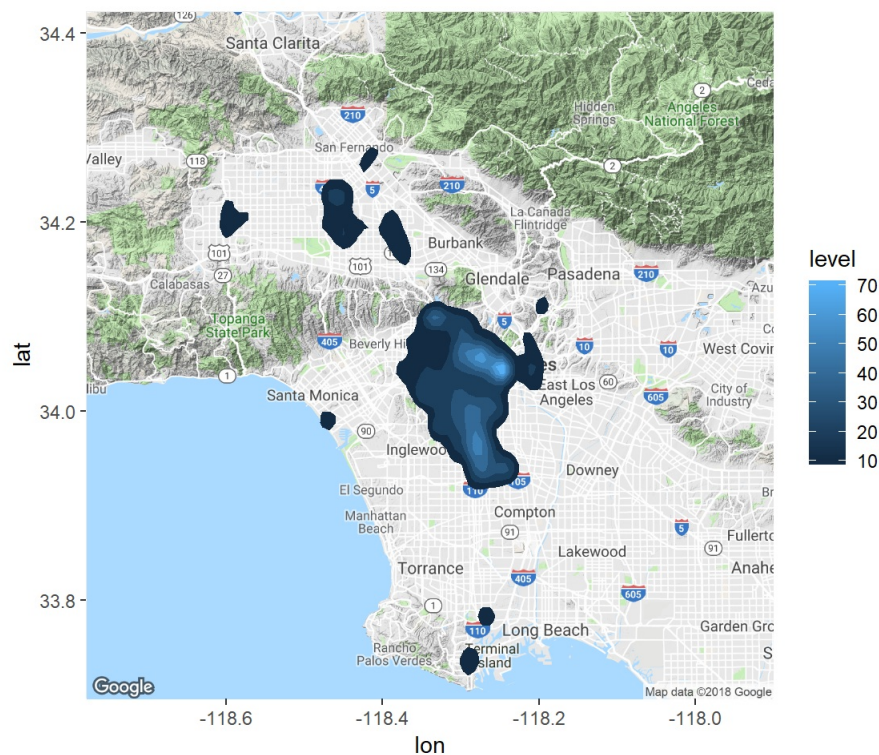
Then, we got our map! Great! Now, We try to plot all the crime location on the map as points.

```
a+geom_point(data = crime, aes(Longitude, Latitude), col="red")
```



Obviously, this plot doesn't look pretty. But it does give us some information. For example, in the northwestern part of the city, we have more crimes than the southeast. But how can we get more information and make the map look more cute?

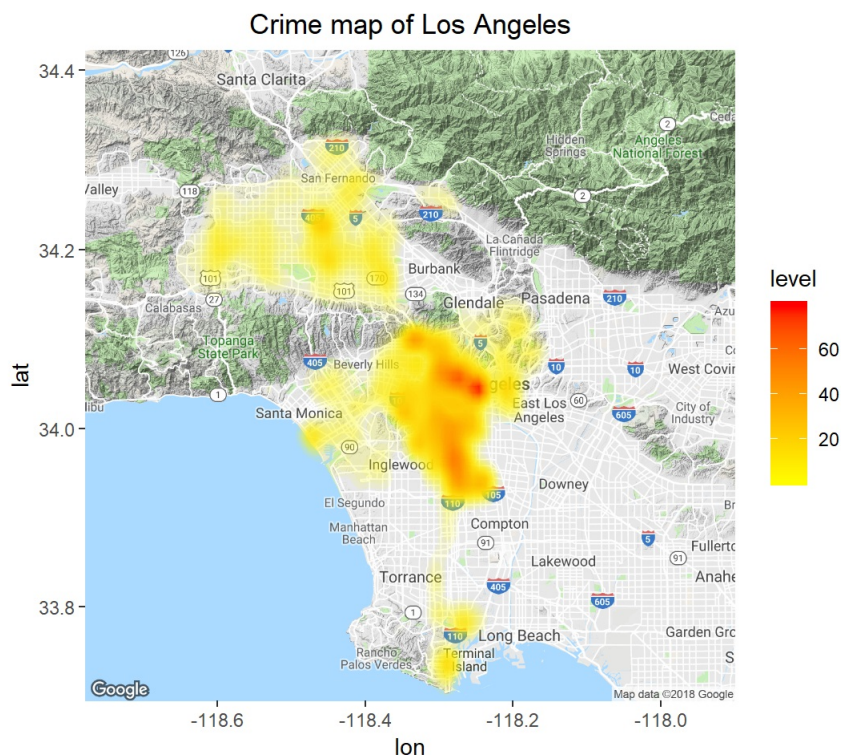
```
b<- a+stat_density_2d(data = crime, aes(Longitude, Latitude, fill = ..level..), geom = "polygon")
b
```



Better Perhaps more adjustment? Remember to adjust the argument “bins”, because it determines how much area the density plot would cover.

```
b <- a+stat_density_2d(data = crime, aes(Longitude, Latitude, fill = ..level.., alpha = ..level..)
                        , size = 0.05, bins = 70, geom = "polygon") +
  scale_fill_gradient2(low = "white", mid = "yellow", high = "red") +
  scale_alpha(range = c(0.1, 0.4), guide = FALSE)

# Add a title
c <- b+labs(title = "Crime map of Los Angeles") +
  theme(plot.title = element_text(hjust = 0.5))
c
```



Finally, this is a more satisfactory picture. Let's see what else we can find using our data

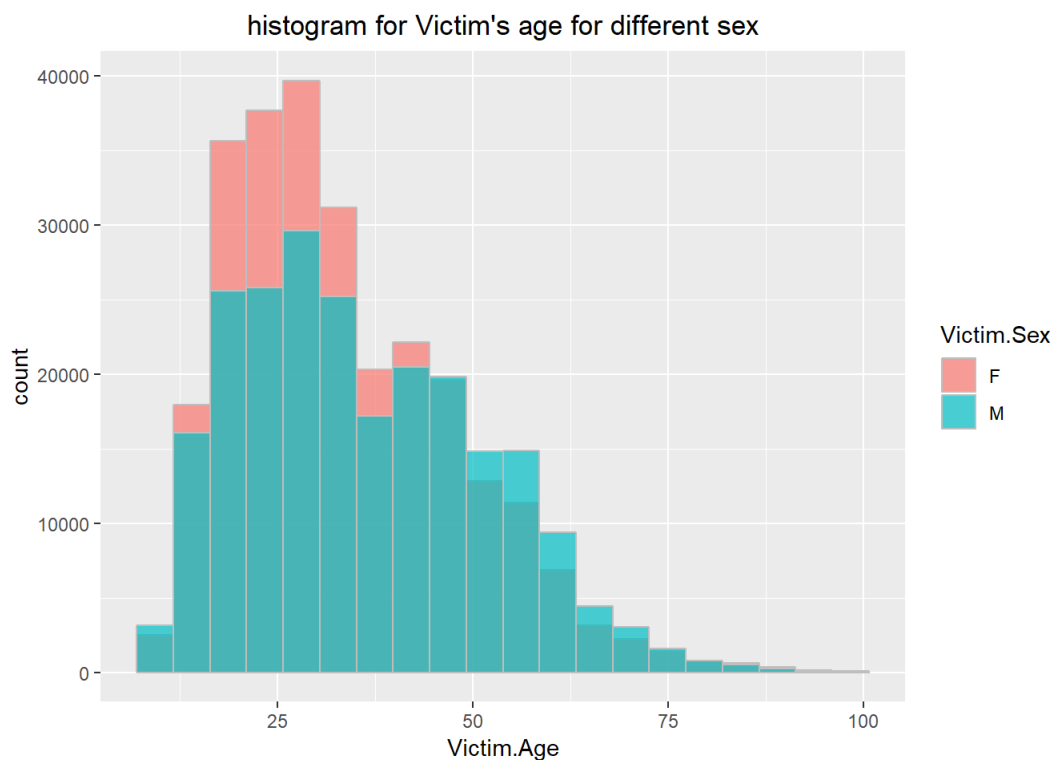
```
variable.names(crime)
```

```
## [1] "DR.Number"      "Date.Reported"   "Date.Occurred"
## [4] "Time.Occurred"  "Area.ID"         "Reporting.District"
## [7] "Crime.Code"     "Victim.Age"      "Victim.Sex"
## [10] "Victim.Descent" "Premise.Code"    "Weapon.Used.Code"
## [13] "Status.Code"    "Latitude"        "Longitude"
```

Okey, we find there are also some information about the age and gender of the victim. We can draw a histogram

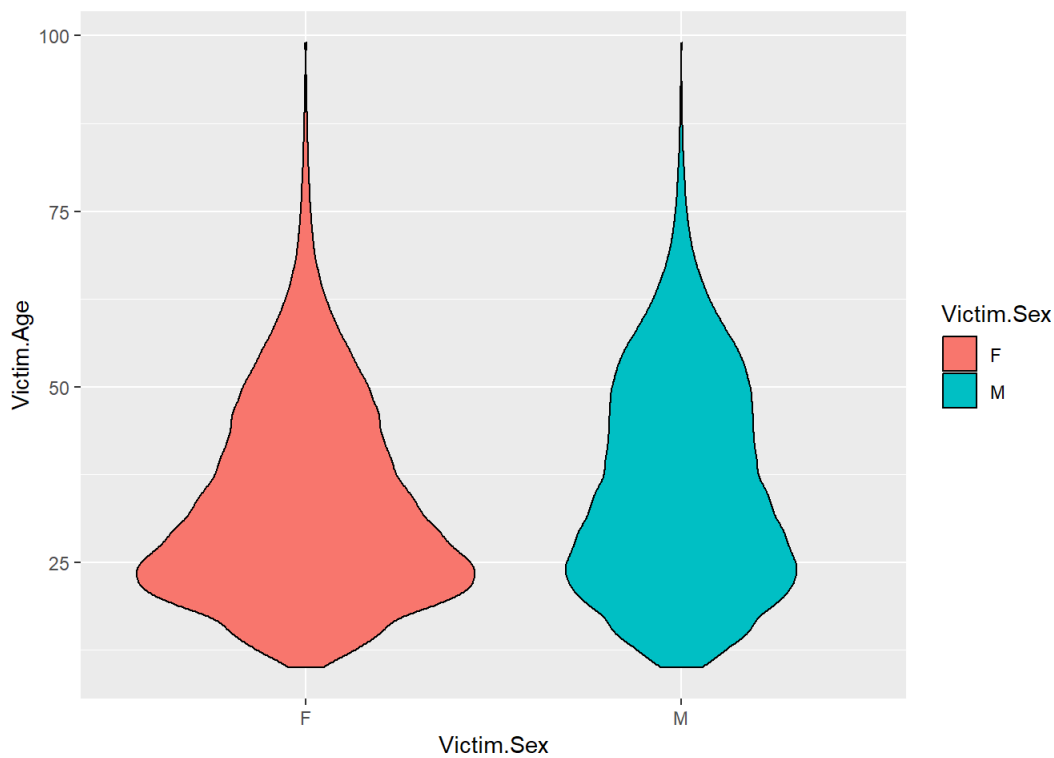
```
filtered_crimes <-select(crime,c(Victim.Sex,Victim.Age)) %>%
  filter(Victim.Sex=="M"|Victim.Sex=="F")

ggplot(data = filtered_crimes,aes(x=Victim.Age,fill=Victim.Sex))+
  geom_histogram(position = "identity",bins = 20,alpha=0.7,col="grey")+
  labs(title="histogram for Victim's age for different sex")+
  theme(plot.title = element_text(hjust=0.5))
```

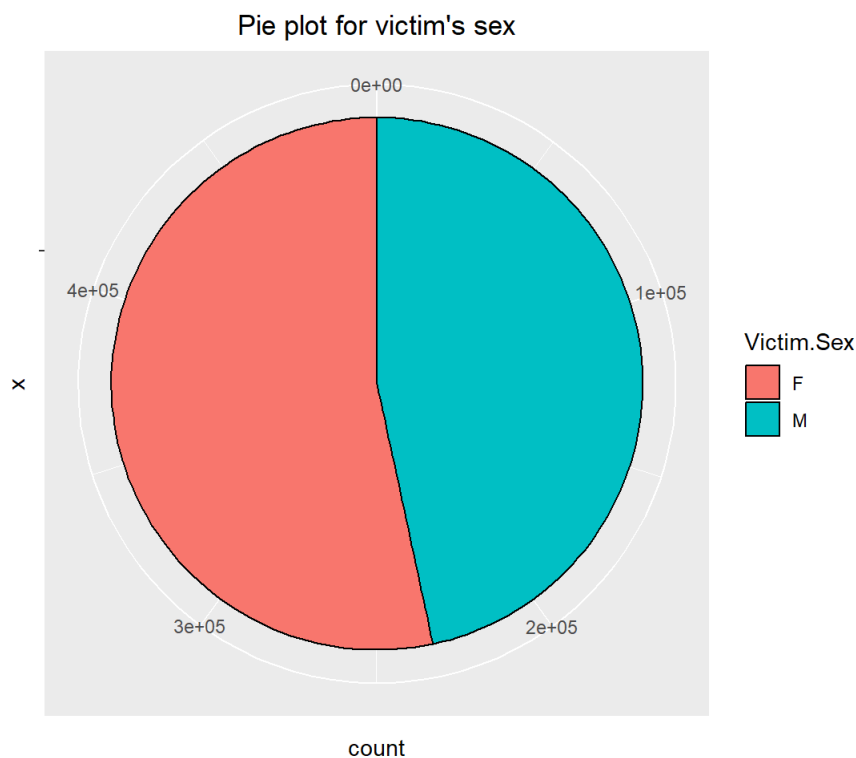


It seems that girls around 25 years old are more likely to be victims. Hope everyone could be careful. We can also draw violin-plot and pie-plot.

```
ggplot(data = filtered_crimes,aes(y=Victim.Age,x=Victim.Sex,fill=Victim.Sex))+
  geom_violin(scale = "count",col="black")
```



```
ggplot(data =filtered_crimes,aes(x="", y=stat(count),fill=Victim.Sex))+
  geom_histogram(stat = "count", width=1,col="black") +
  coord_polar(theta = "y",start = 0)+
  labs(title="Pie plot for victim's sex")+theme(plot.title=element_text(hjust=0.5))
```



In addition to gender, will the number of crimes differ in time? To figure that, we need to manage our data at first. If we want to arrange the data in chronological order. We should convert the date to a timeseries format. I used "xts" packages to do that. Then, we can plot out how many crimes happened in each day.

```
library(xts)
```



```

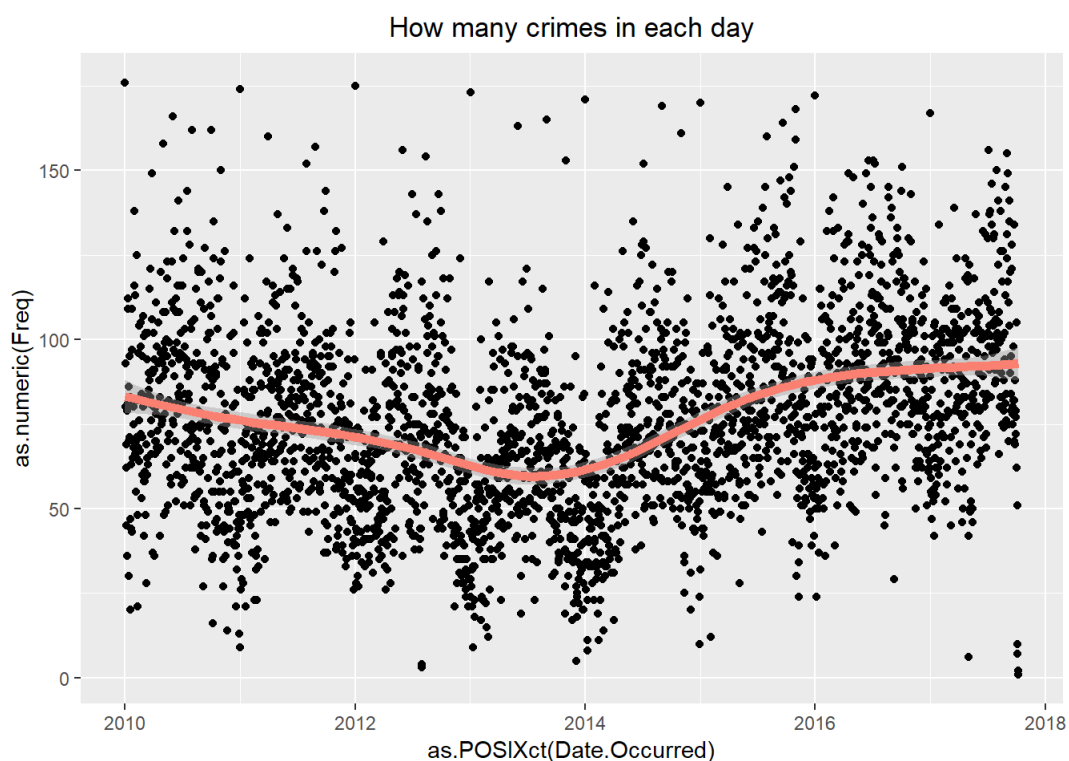
Timeseries <-select(crime,Date.Occurred)%>%
  table(dnn = "Date.Occurred")%>%
  as.data.frame()

Timeseries$Date.Occurred <- as.Date(Timeseries$Date.Occurred,format = "%m/%d/%Y")
Timeseries <- xts(x = Timeseries, order.by = Timeseries$Date.Occurred)

ggplot(Timeseries, aes(x=as.POSIXct(Date.Occurred), as.numeric(Freq)))+
  geom_point()+
  stat_smooth(colour="salmon",size=2)+
  labs(title="How many crimes in each day")+theme(plot.title=element_text(hjust=0.5))

```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



We noticed that there might be a one-year cycle change in the number of crimes. Because we can see from the plot that at the end of each year (i.e. at the beginning of one year), the crime rate will be less. On the contrary, the crime rate is relatively high in the mid of a year. We can further use the time series method for prediction and analysis. For example, we can use Fourier decomposition to further find the period of sequence change.