LA_Crime_Stats

Mohit Patel

1/20/2017

```
library(RgoogleMaps)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
library(stringr)
library(tm)
```

```
## Loading required package: NLP
```

```
##
## Attaching package: 'NLP'
```

```
## The following object is masked from 'package:ggplot2':
##
##     annotate
```

```
library(RColorBrewer)
library(wordcloud)
library(sp)
getwd()
```

```
## [1] "/Users/Mohit/RStudio/LA_Crimes_2013_Exploratory"
```

```
# Read the data
crime_stats <- read.csv("LACrimes2013.csv")

# Wordmap for top 15 crimes in LA in 2013
topcrimes <- Corpus(VectorSource(as.String(crime_stats$CrmCd.Desc)))
topcrimes <- tm_map(topcrimes,PlainTextDocument)
topcrimes <- tm_map(topcrimes, stripWhitespace)
topcrimes <- tm_map(topcrimes, tolower)
topcrimes <- tm_map(topcrimes, removeWords, stopwords("english"))
topcrimes <- tm_map(topcrimes,removeNumbers)
topcrimes <- tm_map(topcrimes,removePunctuation)
topcrimes <- tm_map(topcrimes,PlainTextDocument)
wordcloud(topcrimes,max.words=30,rot.per=0,use.r.layout=FALSE,colors=brewer.pal(8,"Ac
cent"))
```


```
# Wordmap for top areas in crime
crime_areas <- Corpus(VectorSource(as.String(crime_stats$AREA.NAME)))
levels(crime_stats$AREA.NAME)  # 21 Areas
```

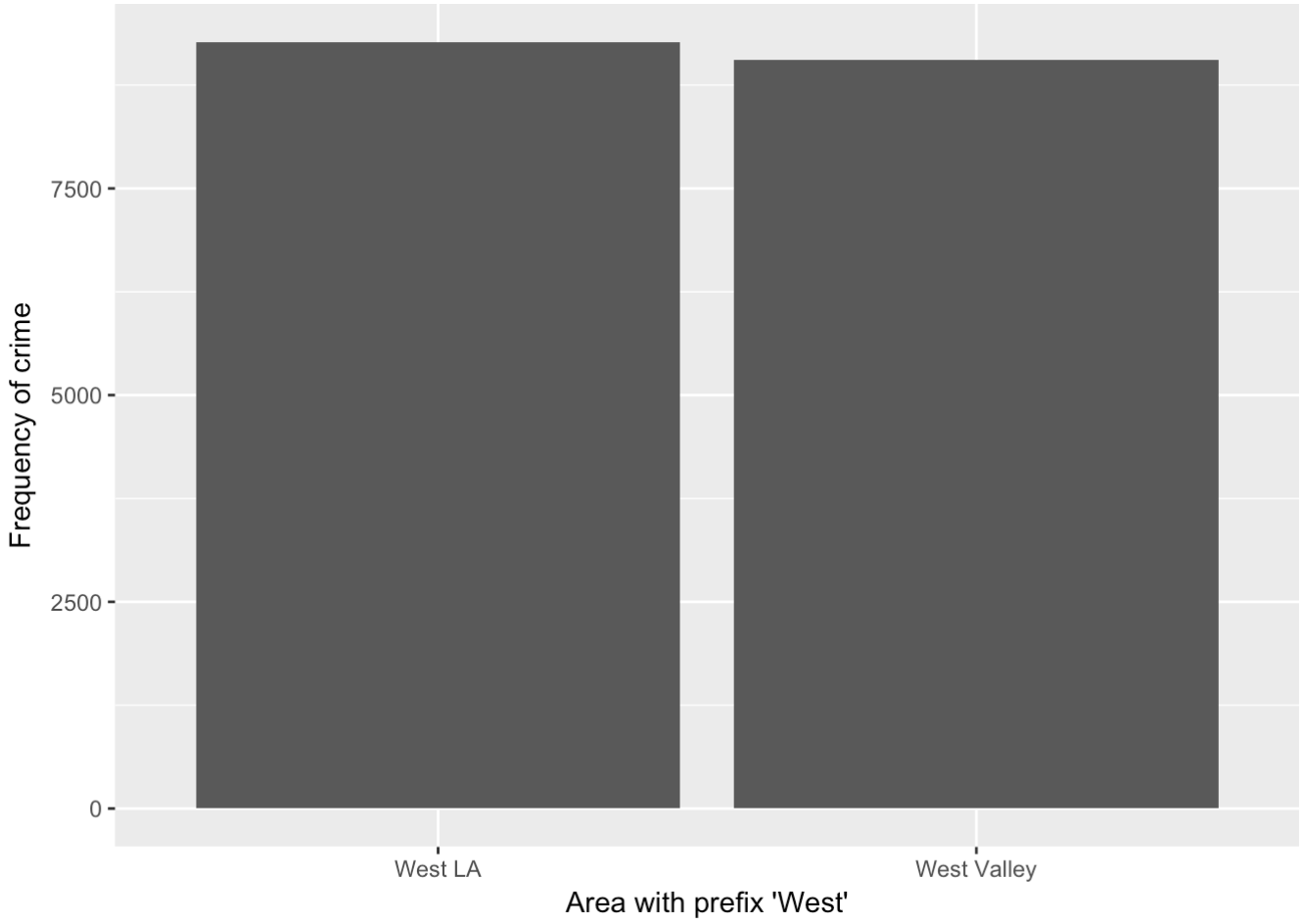
```
##  [1] "77th Street" "Central"     "Devonshire"  "Foothill"    "Harbor"
##  [6] "Hollenbeck"  "Hollywood"   "Mission"     "N Hollywood" "Newton"
## [11] "Northeast"   "Olympic"     "Pacific"     "Rampart"     "Southeast"
## [16] "Southwest"   "Topanga"     "Van Nuys"    "West LA"     "West Valley"
## [21] "Wilshire"
```

```
crime_areas <- tm_map(crime_areas,PlainTextDocument)
crime_areas <- tm_map(crime_areas, stripWhitespace)
crime_areas <- tm_map(crime_areas, tolower)
crime_areas <- tm_map(crime_areas,PlainTextDocument)
wordcloud(crime_areas,max.words=10, random.order=FALSE, rot.per=0,
use.r.layout=FALSE, colors=brewer.pal(5, "Set1"))
```


```
# What does West imply?
west_area_filter <- (crime_stats$AREA.NAME=="West LA") | (crime_stats$AREA.NAME=="Wes
t Valley")
west_area_crime_stats <- crime_stats[west_area_filter,,]
levels(west_area_crime_stats$AREA.NAME)
```

```
##  [1] "77th Street" "Central"     "Devonshire"  "Foothill"    "Harbor"
##  [6] "Hollenbeck"  "Hollywood"   "Mission"     "N Hollywood" "Newton"
## [11] "Northeast"   "Olympic"     "Pacific"     "Rampart"     "Southeast"
## [16] "Southwest"   "Topanga"     "Van Nuys"    "West LA"     "West Valley"
## [21] "Wilshire"
```

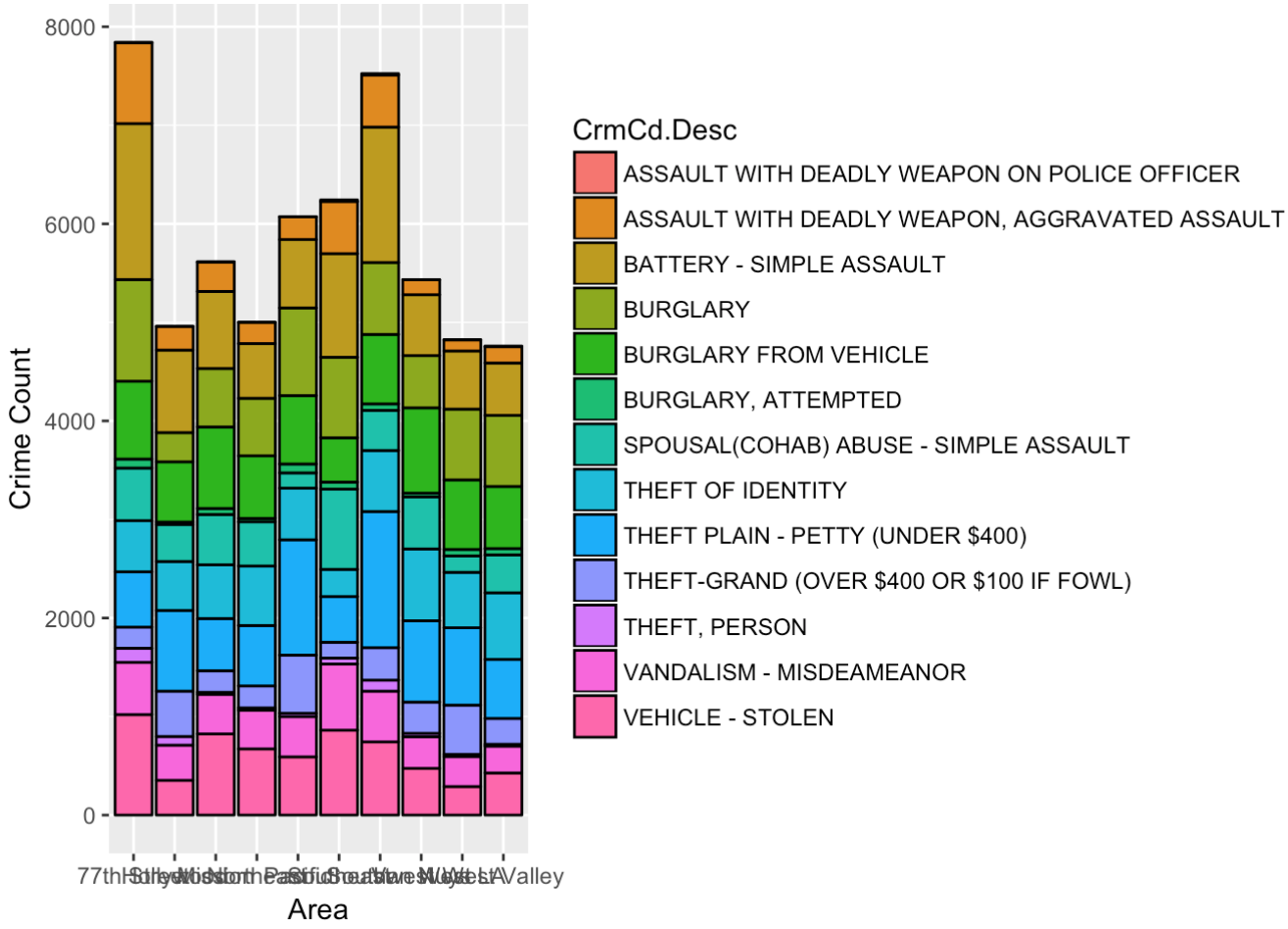
```
ggplot(data=west_area_crime_stats,aes(AREA.NAME))+geom_bar()+xlab("Area with prefix
 'West'")+ylab("Frequency of crime")
```



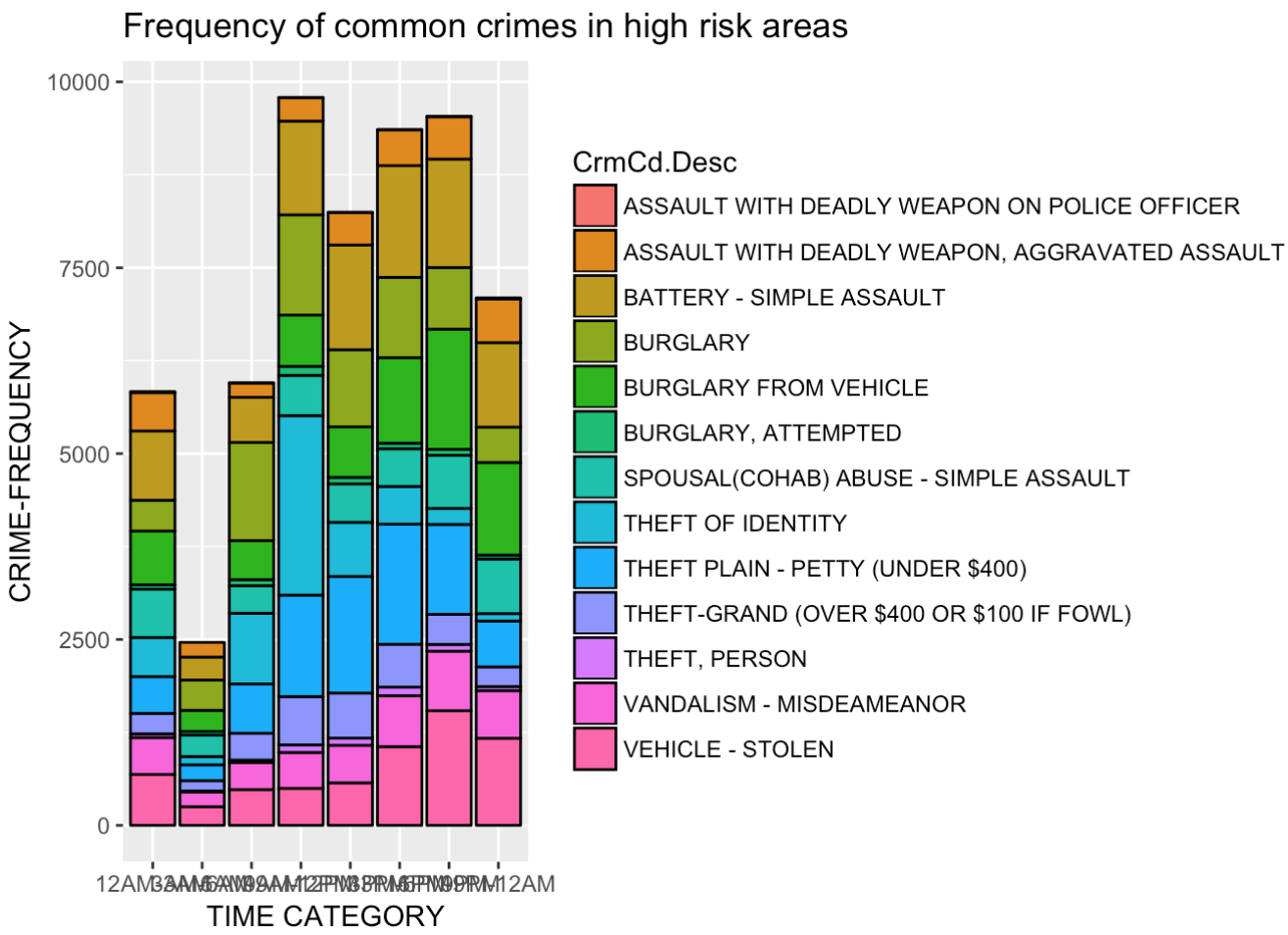

```
# Converting time as category to aggregate crimes
head(crime_stats$TIME.OCC)
```

```
## [1] 2015  445  745 1730 2000 1145
```

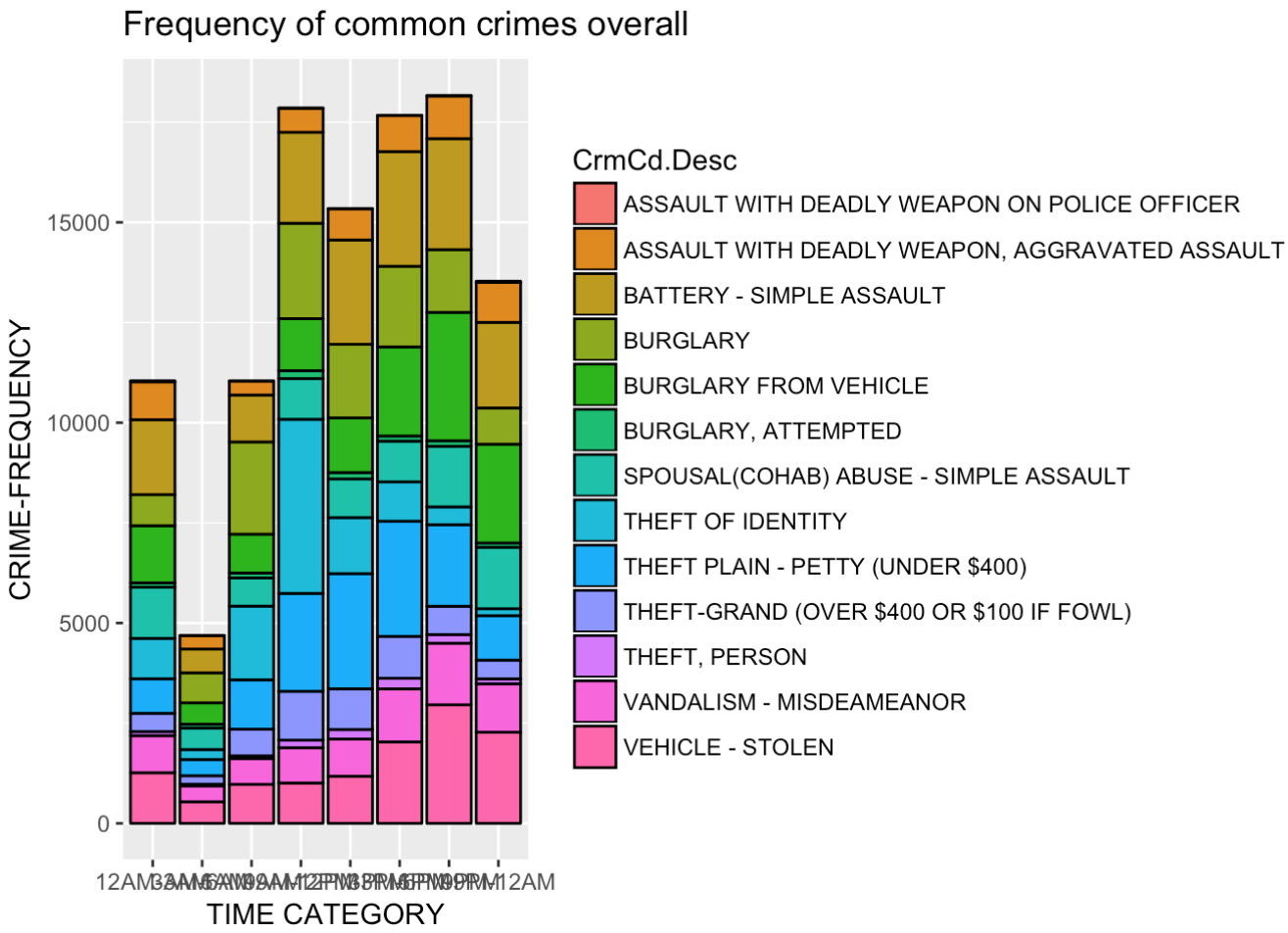
```
crime_stats$TIMECat <- NA

# Convert time of occurance into a Categorical Variable
crime_stats$TIMECat <- cut(crime_stats$TIME.OCC,breaks=c(0000,300,600,900,1200,1500,1
800,2100,2400),labels=c("12AM-3AM","3AM-6AM","6AM-9AM","9AM-12PM","12PM-3PM","3PM-6P
M","6PM-9PM","9PM-12AM"))




# Top crimes in the high risk areas
Area_Filter <- (crime_stats$AREA.NAME=="Northeast")|
(crime_stats$AREA.NAME=="Mission")|(crime_stats$AREA.NAME=="Southwest")|
(crime_stats$AREA.NAME=="Hollywood")|(crime_stats$AREA.NAME=="West LA")|
(crime_stats$AREA.NAME=="West Valley")|(crime_stats$AREA.NAME=="Pacific")|(crime_stat
s$AREA.NAME=="Southeast")|(crime_stats$AREA.NAME=="Van Nuys")|(crime_stats$AREA.NAME=
="77th Street")

Area_df <- crime_stats[Area_Filter,,]

Top_Crime_Filter <- (Area_df$CrmCd.Desc=="ASSAULT WITH DEADLY WEAPON ON POLICE OFFICE
R") | (Area_df$CrmCd.Desc=="ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT")|
(Area_df$CrmCd.Desc=="BURGLARY")|(Area_df$CrmCd.Desc=="BURGLARY FROM VEHICLE")|(Area_
df$CrmCd.Desc=="BURGLARY, ATTEMPTED")|(Area_df$CrmCd.Desc=="SPOUSAL(COHAB) ABUSE - SI
MPLE ASSAULT")|(Area_df$CrmCd.Desc=="THEFT, PERSON")|(Area_df$CrmCd.Desc=="VEHICLE -
 STOLEN")|(Area_df$CrmCd.Desc=="VANDALISM - MISDEAMEANOR")|(Area_df$CrmCd.Desc=="THEF
T PLAIN - PETTY (UNDER $400)")|(Area_df$CrmCd.Desc=="VEHICLE - STOLEN")|(Area_df$CrmC
d.Desc=="BATTERY - SIMPLE ASSAULT")|(Area_df$CrmCd.Desc=="BATTERY - SIMPLE ASSAULT")|
(Area_df$CrmCd.Desc=="SPOUSAL(COHAB) ABUSE - SIMPLE ASSAULT")|(Area_df$CrmCd.Desc=="T
HEFT OF IDENTITY")|(Area_df$CrmCd.Desc=="THEFT-GRAND (OVER $400 OR $100 IF FOWL)")|(A
rea_df$CrmCd.Desc=="THEFT PLAIN - PETTY (UNDER $400)")

Top_Crime_df <- Area_df[Top_Crime_Filter,,]

Top_Crime_Overall_Filter <- (crime_stats$CrmCd.Desc=="ASSAULT WITH DEADLY WEAPON ON P
OLICE OFFICER") | (crime_stats$CrmCd.Desc=="ASSAULT WITH DEADLY WEAPON, AGGRAVATED AS
SAULT")|(crime_stats$CrmCd.Desc=="BURGLARY")|(crime_stats$CrmCd.Desc=="BURGLARY FROM
 VEHICLE")|(crime_stats$CrmCd.Desc=="BURGLARY, ATTEMPTED")|
(crime_stats$CrmCd.Desc=="SPOUSAL(COHAB) ABUSE - SIMPLE ASSAULT")|(crime_stats$CrmCd.
Desc=="THEFT, PERSON")|(crime_stats$CrmCd.Desc=="VEHICLE - STOLEN")|(crime_stats$CrmC
d.Desc=="VANDALISM - MISDEAMEANOR")|(crime_stats$CrmCd.Desc=="THEFT PLAIN - PETTY (UN
DER $400)")|(crime_stats$CrmCd.Desc=="VEHICLE - STOLEN")|(crime_stats$CrmCd.Desc=="BA
TTERY - SIMPLE ASSAULT")|(crime_stats$CrmCd.Desc=="BATTERY - SIMPLE ASSAULT")|(crime_
stats$CrmCd.Desc=="SPOUSAL(COHAB) ABUSE - SIMPLE ASSAULT")|
(crime_stats$CrmCd.Desc=="THEFT OF IDENTITY")|(crime_stats$CrmCd.Desc=="THEFT-GRAND
 (OVER $400 OR $100 IF FOWL)")|(crime_stats$CrmCd.Desc=="THEFT PLAIN - PETTY (UNDER
 $400)")
Top_Crime_overall_df <- crime_stats[Top_Crime_Overall_Filter,,]

ggplot(data=Top_Crime_df,aes(AREA.NAME,fill=CrmCd.Desc))+geom_bar(color="Black")+xlab("
rea")+ylab("Crime Count")
```




```
#ggplot(data=Top_Crime_df,aes(AREA.NAME,fill=CrmCd.Desc))+geom_bar(color="Black")+xla
b("Area")+ylab("Crime Count")

# Frequency of occurance vs Time of Occurance

# Frequency vs Time based occurance of most common crime in high risk areas identifie
d in the word cloud
ggplot(data=Top_Crime_df,aes(x=TIMECat,fill=CrmCd.Desc))+geom_bar(colour="Black")+xla
b("TIME CATEGORY")+ylab("CRIME-FREQUENCY")+ggtitle("Frequency of common crimes in hig
h risk areas")
```




```
# Frequency vs Time based occurance of most common crimes in all areas according to T
IME
ggplot(data=Top_Crime_overall_df,aes(TIMECat,fill=CrmCd.Desc))+geom_bar(colour="Blac
k")+xlab("TIME CATEGORY")+ylab("CRIME-FREQUENCY")+ggtitle("Frequency of common crimes
 overall")
```




```
# Plot occurances of crime on the Map

# Get Map
la_get_map <- GetMap(center=c(lat=34.052,lon=-118.24),size=c(640,640),zoom=9,maptype
= "terrain",destfile="map1.png")

# Separate location data
location_data <- crime_stats$Location.1
typeof(location_data)
```

```
## [1] "integer"
```

```
head(location_data)
```

```
## [1] (34.0776, -118.308)  (34.1113, -118.3336) (33.9406, -118.2338)
## [4] (33.9449, -118.2332) (33.8135, -118.2992) (33.9931, -118.3308)
## 37387 Levels: (33.706, -118.2898) ... (34.7907, -118.317)
```

```
Latitude <- str_match(as.character(location_data), "(\\w+.\\w+),\\s(-\\w+.\\w+)")


head(Latitude)
```

```
##      [,1]                 [,2]      [,3]
## [1,] "34.0776, -118.308"  "34.0776" "-118.308"
## [2,] "34.1113, -118.3336" "34.1113" "-118.3336"
## [3,] "33.9406, -118.2338" "33.9406" "-118.2338"
## [4,] "33.9449, -118.2332" "33.9449" "-118.2332"
## [5,] "33.8135, -118.2992" "33.8135" "-118.2992"
## [6,] "33.9931, -118.3308" "33.9931" "-118.3308"
```

```
LLmatrix <- cbind(Latitude[,2],Latitude[,3])
head(LLmatrix)
```

```
##      [,1]      [,2]
## [1,] "34.0776" "-118.308"
## [2,] "34.1113" "-118.3336"
## [3,] "33.9406" "-118.2338"
## [4,] "33.9449" "-118.2332"
## [5,] "33.8135" "-118.2992"
## [6,] "33.9931" "-118.3308"
```

```
numeric_lat <- as.numeric(LLmatrix[,1])
head(numeric_lat)
```

```
## [1] 34.0776 34.1113 33.9406 33.9449 33.8135 33.9931
```

```
numeric_lon <- as.numeric(LLmatrix[,2])
head(numeric_lon)
```

```
## [1] -118.3080 -118.3336 -118.2338 -118.2332 -118.2992 -118.3308
```

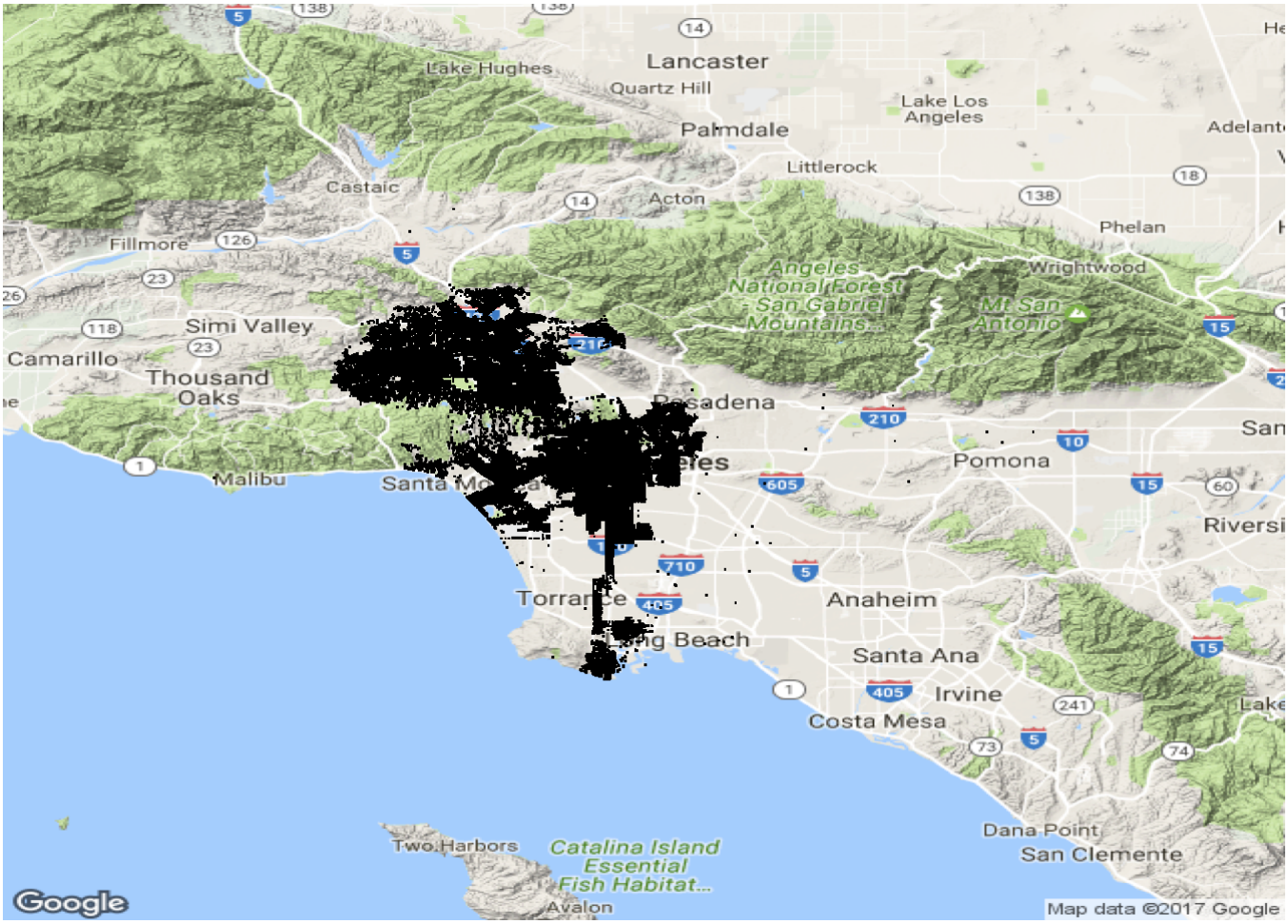
```
numericll <- cbind(numeric_lat,numeric_lon)
head(numericll)
```

```
##      numeric_lat numeric_lon
## [1,]     34.0776   -118.3080
## [2,]     34.1113   -118.3336
## [3,]     33.9406   -118.2338
## [4,]     33.9449   -118.2332
## [5,]     33.8135   -118.2992
## [6,]     33.9931   -118.3308
```

```
#Add location information back to crime_stats dataset
crime_stats$Latitude <- numericll[,1,drop=F]
crime_stats$Longitude <- numericll[,2,drop=F]
head(crime_stats)
```

```
##    Date.Rptd     DR.NO   DATE.OCC TIME.OCC AREA   AREA.NAME   RD Crm.Cd
## 1 03/20/2013 132007717 03/20/2013     2015   20     Olympic 2004    997
## 2 03/10/2013 130608787 03/10/2013      445    6   Hollywood  635    997
## 3 12/18/2013 131820260 12/18/2013      745   18   Southeast 1839    997
## 4 10/18/2013 131817514 10/18/2013     1730   18   Southeast 1827    997
## 5 05/26/2013 130510483 05/25/2013     2000    5      Harbor  507    440
## 6 05/24/2013 131213618 05/22/2013     1145   12 77th Street 1211    997
##                        CrmCd.Desc Status Status.Desc
## 1                    TRAFFIC DR #    UNK     Unknown
## 2                    TRAFFIC DR #    UNK     Unknown
## 3                    TRAFFIC DR #    UNK     Unknown
## 4                    TRAFFIC DR #    UNK     Unknown
## 5 THEFT PLAIN - PETTY (UNDER $400)    UNK     Unknown
## 6                    TRAFFIC DR #    UNK     Unknown
##                                      LOCATION
## 1                                      OXFORD
## 2              ODIN                        ST
## 3              105TH                       ST
## 4              101ST                       ST
## 5  1300      W  SEPULVEDA                   BL
## 6                                      54TH
##                        Cross.Street           Location.1
## 1                          OAKWOOD  (34.0776, -118.308)
## 2    CAHUENGA                   BL (34.1113, -118.3336)
## 3    CROESUS                    AV (33.9406, -118.2338)
## 4    JUNIPER                    ST (33.9449, -118.2332)
## 5                                  (33.8135, -118.2992)
## 6                         CRENSHAW (33.9931, -118.3308)
##                                Formatted.Address Zipcode  TIMECat
## 1  4650-4652 Oakwood Ave, Los Angeles, CA 90004, USA   90004  6PM-9PM
## 2   2314 N Cahuenga Blvd, Los Angeles, CA 90068, USA   90068  3AM-6AM
## 3        2148 E 105th St, Los Angeles, CA 90002, USA   90002  6AM-9AM
## 4 10100-10198 Juniper St, Los Angeles, CA 90002, USA   90002  3PM-6PM
## 5       1303 Sepulveda Blvd, Torrance, CA 90501, USA   90501  6PM-9PM
## 6         3400 W 54th St, Los Angeles, CA 90043, USA   90043 9AM-12PM
##   numeric_lat numeric_lon
## 1     34.0776   -118.3080
## 2     34.1113   -118.3336
## 3     33.9406   -118.2338
## 4     33.9449   -118.2332
## 5     33.8135   -118.2992
## 6     33.9931   -118.3308
```

```
pmb_lat <- c(crime_stats$Latitude)
pmb_lon <- c(crime_stats$Longitude)

PlotOnStaticMap(la_get_map,lat=c(pmb_lat),lon=c(pmb_lon),destfile="map2.png",zoom=10,ce
=1,pch='.')
```




```
?PlotOnStaticMap
```