Data Mining Practice - {ggmap}

Google Maps, OpenStreetMap 을 가지고 spatial visualization 을 하게 해주는 패키지

   간략히 말해서 ggmap 은 map image 를 다운로드 받아 ggplot2 를 이용해 context l ayer 처럼 그려진다. 그리고 추가적으로 data, statistics, or models 의 layer 를 m ap image 위에 그릴 수 있게 해주는 패키지이다.

   Intro.

```
install.packages("ggmap")

## Installing package into 'C:/Users/dox/Documents/R/win-library/3.0'
## (as 'lib' is unspecified)

## Error: trying to use CRAN without setting a mirror

library(ggmap)

## Loading required package: ggplot2
```
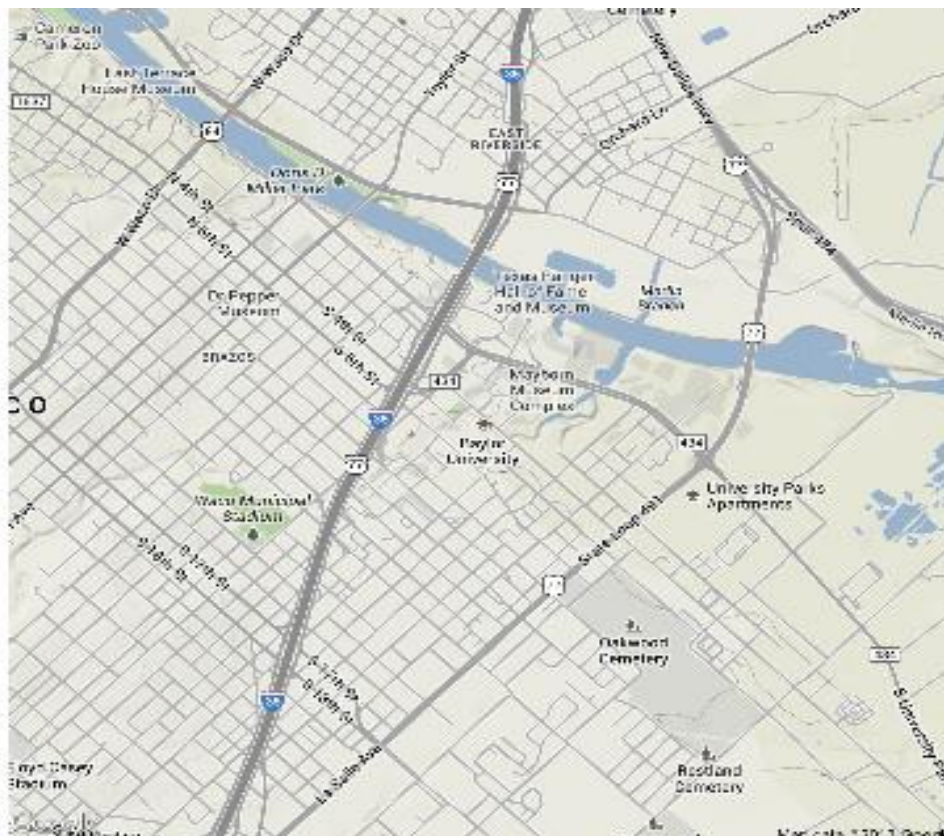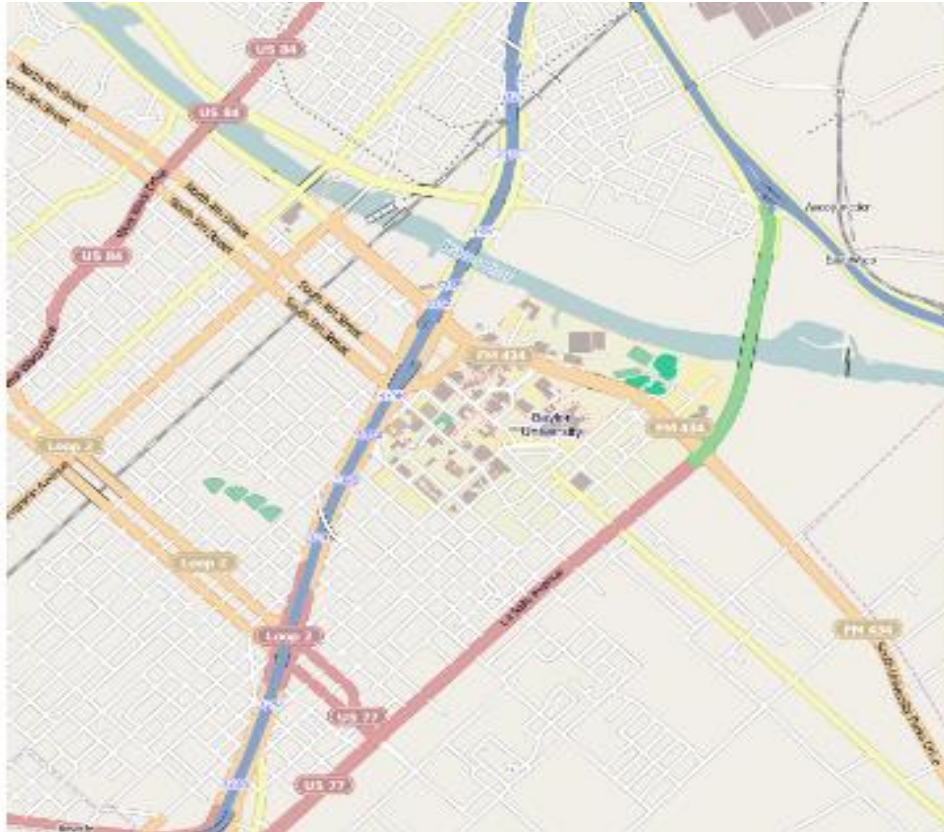
qmap()은 ggplot2 에서 qplot()과 같은 의미로 보면 된다.

```
baylor <- "baylor university"
qmap(baylor, zoom = 14)
```

plot of chunk unnamed-chunk-2
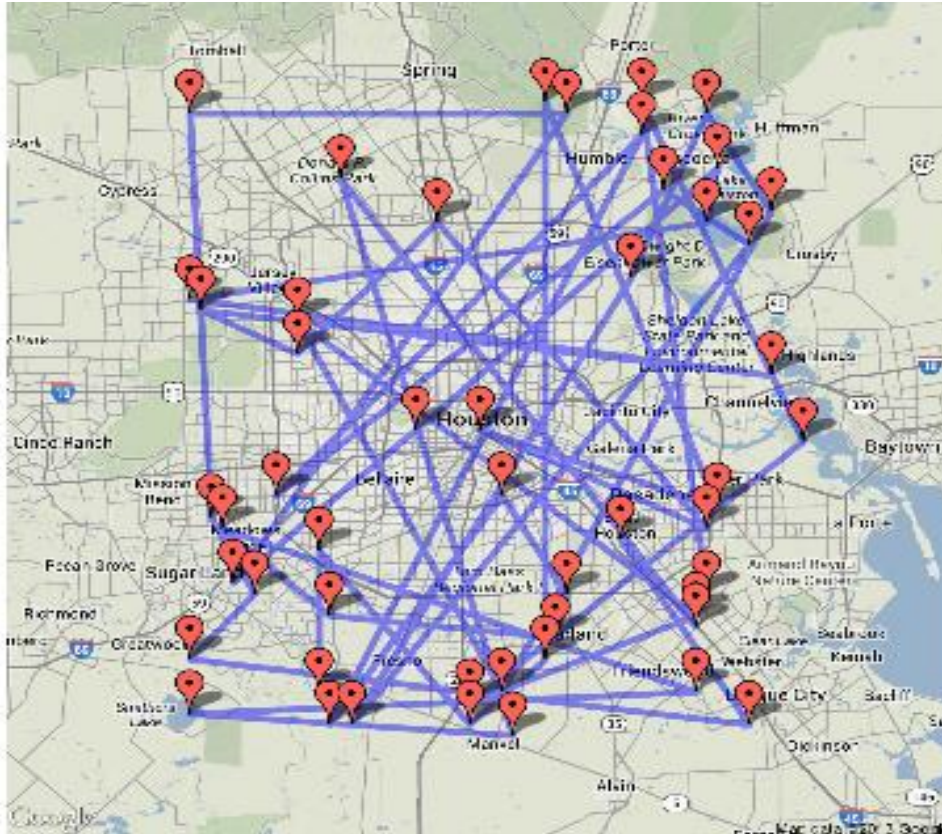
```
qmap(baylor, zoom = 14, source = "osm")
```


plot of chunk unnamed-chunk-2

```
set.seed(500)
df <- round(data.frame(x = jitter(rep(-95.36, 50), amount = 0.3), y = jitt
er(rep(29.76,
    50), amount = 0.3)), digits = 2)

map <- get_googlemap("houston", markers = df, path = df, scale = 2)

ggmap(map, extent = "device")
```

plot of chunk unnamed-chunk-3

```
qmap(baylor, zoom = 14, source = "stamen", maptype = "watercolor")
```
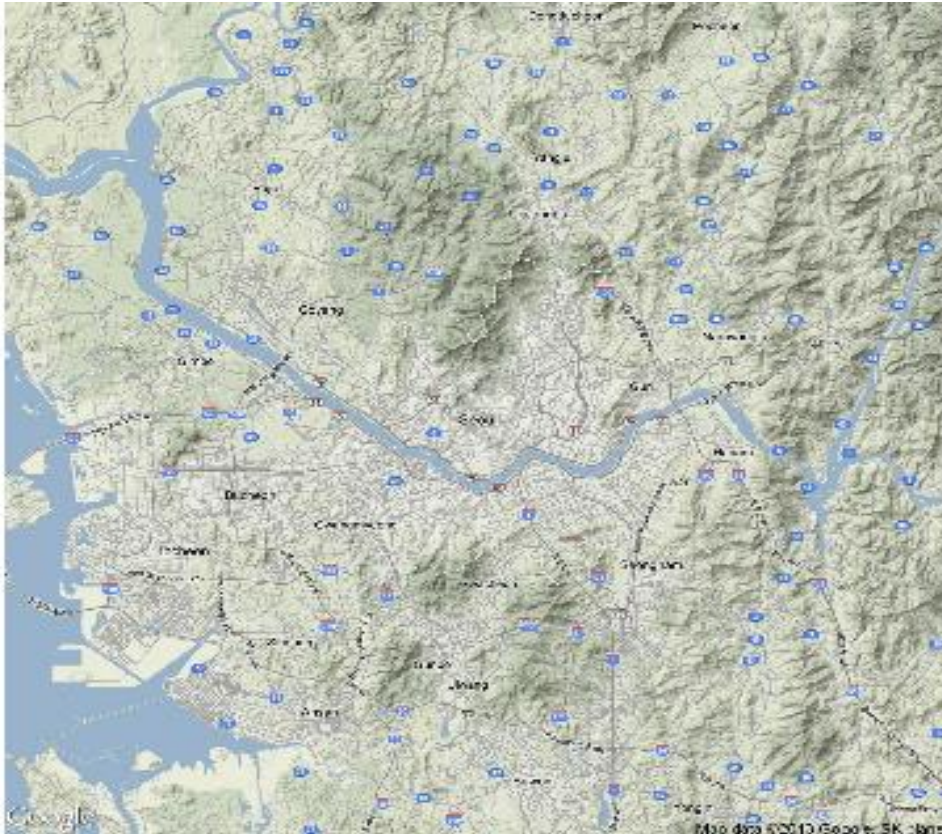
plot of chunk unnamed-chunk-4

```
qmap(baylor, zoom = 14, source = "stamen", maptype = "toner")
```


plot of chunk unnamed-chunk-4

```
seoul <- get_map(location = "seoul")
ggmap(seoul, extent = "device")
```

plot of chunk unnamed-chunk-5

Example

Data

Crime data 는 2010 년 1 월부터 2010 년 8 월까지 축적된 Houston Police Department's website 자료이다. 이 데이터는 ggmap 패키지 안에 내장되어 있는 데이터이다.

```
data(crime)
str(crime)
```

```
## 'data.frame':    86314 obs. of  17 variables:
##  $ time    : POSIXt, format: "2010-01-01 15:00:00" "2010-01-01 15:00:00
" ...
##  $ date    : chr  "1/1/2010" "1/1/2010" "1/1/2010" "1/1/2010" ...
##  $ hour    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ premise : chr  "18A" "13R" "20R" "20R" ...
##  $ offense : Factor w/ 7 levels "aggravated assault",..: 4 6 1 1 1 3 3
3 3 3 ...
##  $ beat    : chr  "15E30" "13D10" "16E20" "2A30" ...
##  $ block   : chr  "9600-9699" "4700-4799" "5000-5099" "1000-1099" ...
##  $ street  : chr  "marlive" "telephone" "wickview" "ashland" ...
##  $ type    : chr  "ln" "rd" "ln" "st" ...
##  $ suffix  : chr  "-" "-" "-" "-" ...
##  $ number  : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ month   : Ord.factor w/ 8 levels "january"<"february"<..: 1 1 1 1 1
```

```
1 1 1 1 1 ...
## $ day      : Ord.factor w/ 7 levels "monday"<"tuesday"<..: 5 5 5 5 5 5
5 5 5 5 ...
## $ location: chr  "apartment parking lot" "road / street / sidewalk" "r
esidence / house" "residence / house" ...
## $ address : chr  "9650 marlive ln" "4750 telephone rd" "5050 wickview
ln" "1050 ashland st" ...
## $ lon     : num  -95.4 -95.3 -95.5 -95.4 -95.4 ...
## $ lat     : num  29.7 29.7 29.6 29.8 29.7 ...
```

도심에서 발생한 강력범죄에 대해서만 살펴보겠다.

```
# find a reasonable spatial extent
qmap("houston", zoom = 13)

## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=hou
ston&zoom=13&size=%20640x640&scale=%202&maptype=terrain&sensor=false
## Google Maps API Terms of Service : http://developers.google.com/maps/te
rms
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?
address=houston&sensor=false
## Google Maps API Terms of Service : http://developers.google.com/maps/te
rms

gglocator(2)
```
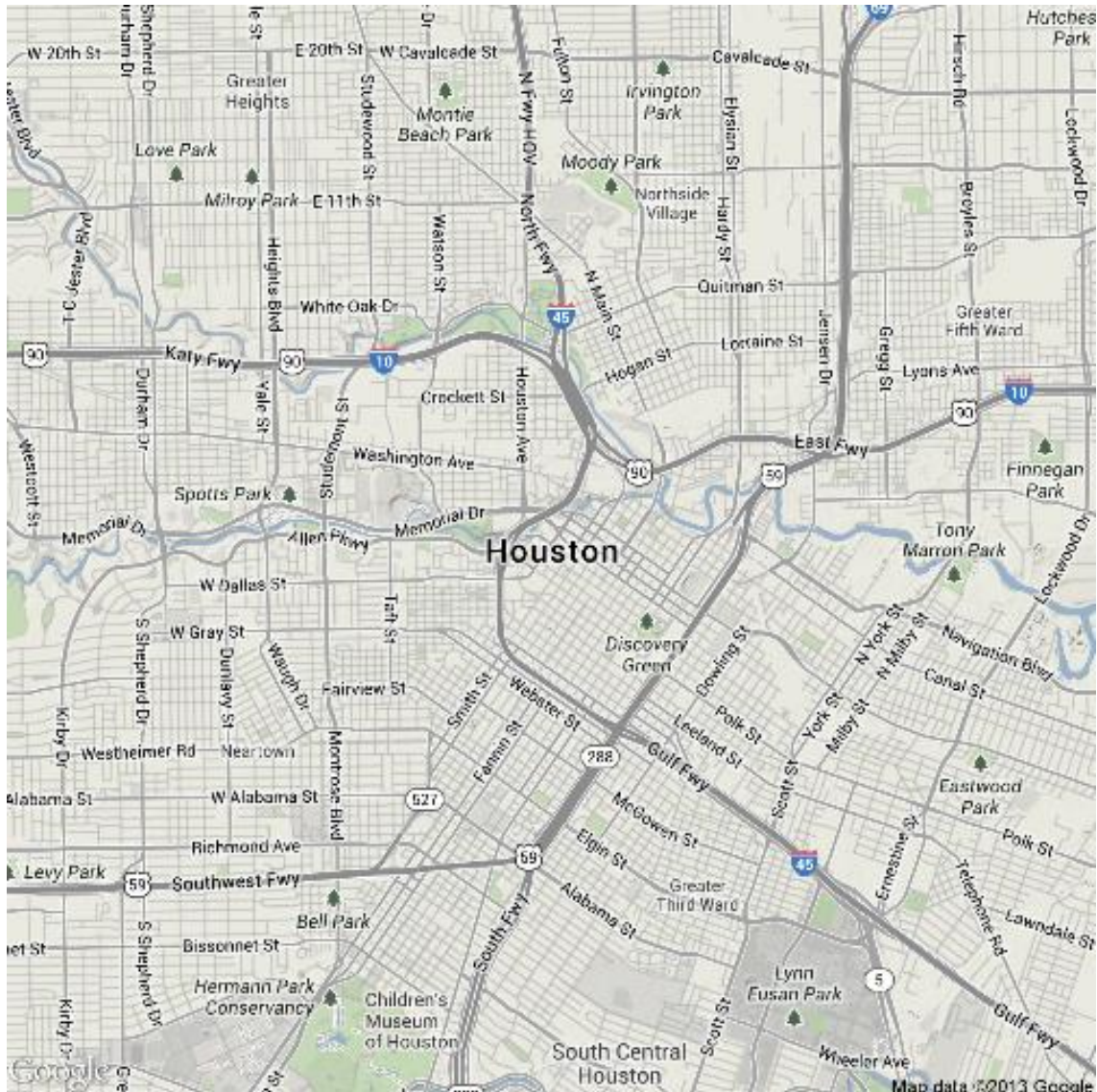
plot of chunk unnamed-chunk-7

```
##    lon lat
## 1  NA  NA
## 2  NA  NA
```

```r
# only violent crimes
violent.crimes <- subset(crime, offense != "auto theft" & offense != "thef
t" &
    offense != "burglary")

# order violent crimes
violent.crimes$offense <- factor(violent.crimes$offense, levels = c("robbe
ry",
    "aggravated assault", "rape", "murder"))
# restrict to downtown
violent.crimes <- subset(violent.crimes, -95.39233 <= lon & lon <= -95.353
```

```
09 &
    29.75384 <= lat & lat <= 29.76909)
```
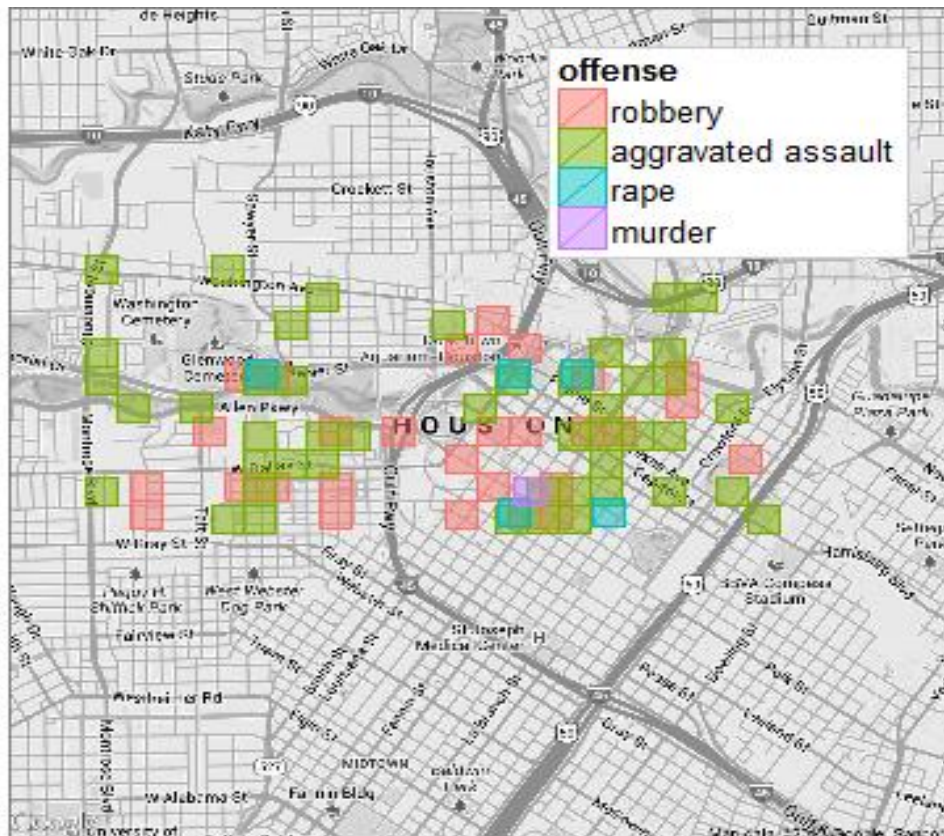
Analysis

```
theme_set(theme_bw(16))
HoustonMap <- qmap("houston", zoom = 14, color = "bw", legend = "topright")

HoustonMap + geom_point(aes(x = lon, y = lat, colour = offense, size = off
ense),
    data = violent.crimes)
```


plot of chunk unnamed-chunk-8

```
HoustonMap + stat_bin2d(aes(x = lon, y = lat, colour = offense, fill = off
ense),
    size = 0.5, bins = 30, alpha = 1/2, data = violent.crimes)
```
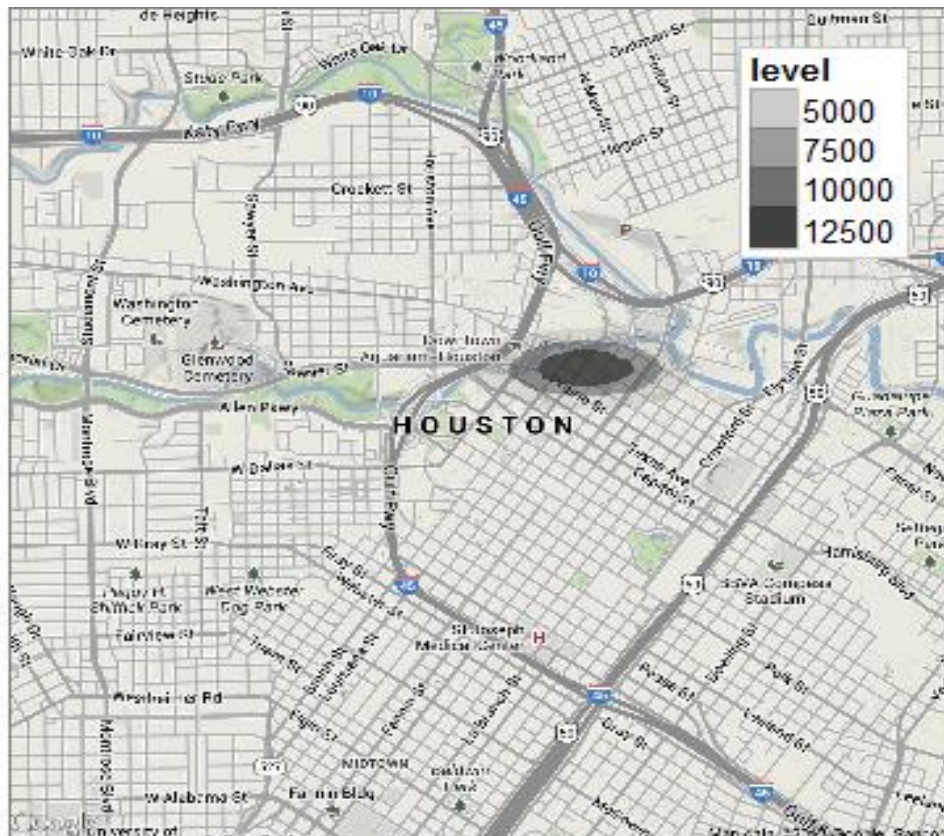
plot of chunk unnamed-chunk-9

범죄의 종류를 배제하면 contour plot 을 이용해 지도상에 강력범죄의 분포를 살펴 볼 수 있다. 즉, 강력범죄의 밀도를 확인 할 수 있다.

```
houston <- get_map(location = "houston", zoom = 14)
HoustonMap <- ggmap(houston, extent = "device", legend = "topright")

HoustonMap + stat_density2d(aes(x = lon, y = lat, fil = ..level.., alpha =
 ..level..),
    size = 2, bins = 4, data = violent.crimes, geom = "polygon")
```
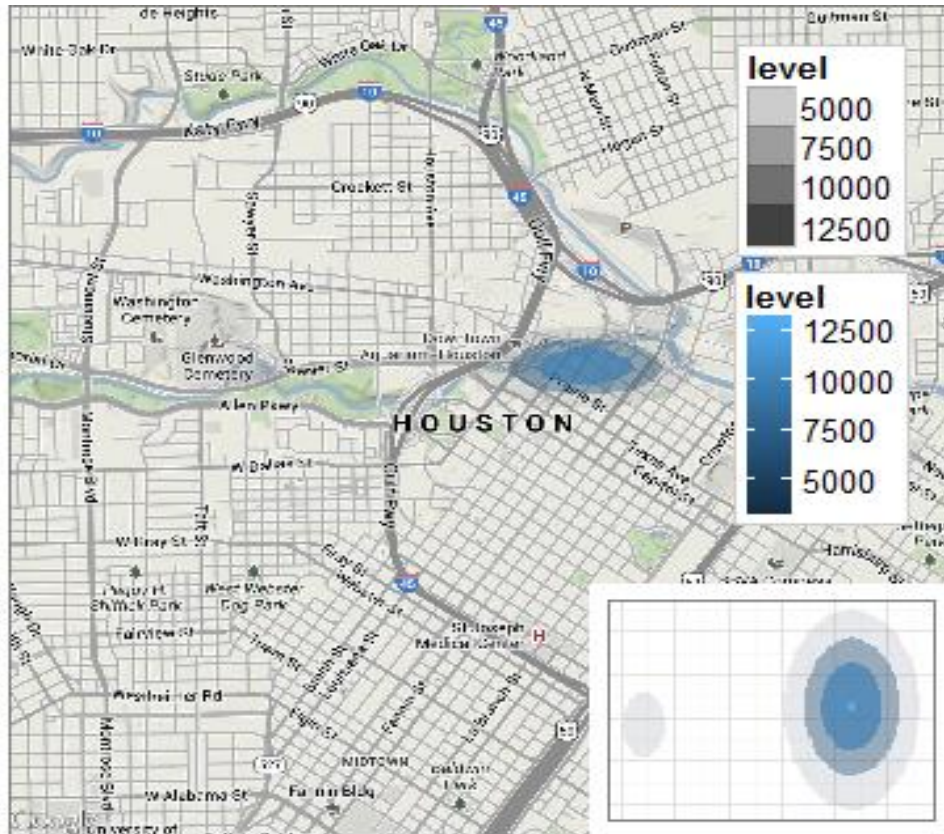
plot of chunk unnamed-chunk-10

```
overlay <- stat_density2d(aes(x = lon, y = lat, fill = ..level.., alpha =
  ..level..),
    bins = 4, geom = "polygon", data = violent.crimes)

HoustonMap + overlay + inset(grob = ggplotGrob(ggplot() + overlay + theme_
inset()),
    xmin = -95.36309, xmax = Inf, ymin = -Inf, ymax = 29.75084)
```
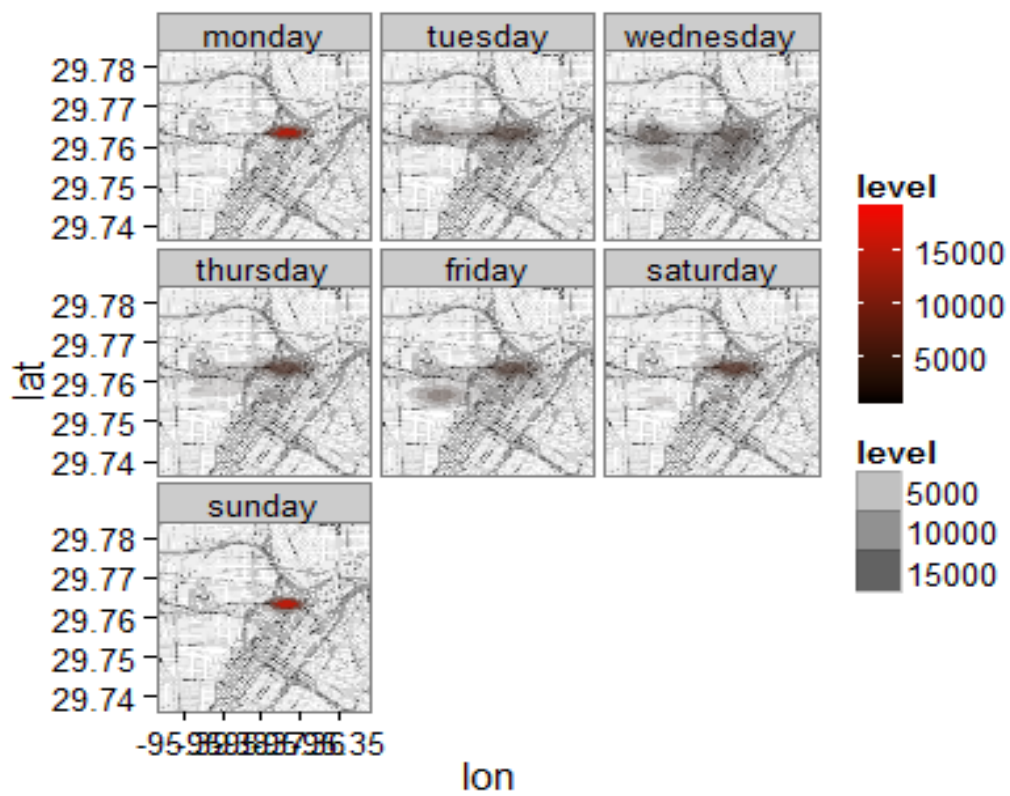
plot of chunk unnamed-chunk-10

다음은 요일별 강력범죄 발생빈도

```
houston <- get_map(location = "houston", zoom = 14, color = "bw", source =
 "osm")
HoustonMap <- ggmap(houston, base_layer = ggplot(aes(x = lon, y = lat), da
ta = violent.crimes))

HoustonMap + stat_density2d(aes(x = lon, y = lat, fill = ..level.., alpha
= ..level..),
    bins = 5, geom = "polygon", data = violent.crimes) + scale_fill_gradie
nt(low = "black",
    high = "red") + facet_wrap(~day)
```

plot of chunk unnamed-chunk-11

Reference : The R Journal