

CSUEB Statistics 6610, Winter 2017, Prof. E. Suess  
Gui Larangeira, March 14th, 2017

## PROJECT 1 – PLOTTING DENSITY MAPS (HEATMAPS) FOR SAN FRANCISCO CRIME DATA USING TABLEAU, GGPLOT AND SHAPE FILES

---

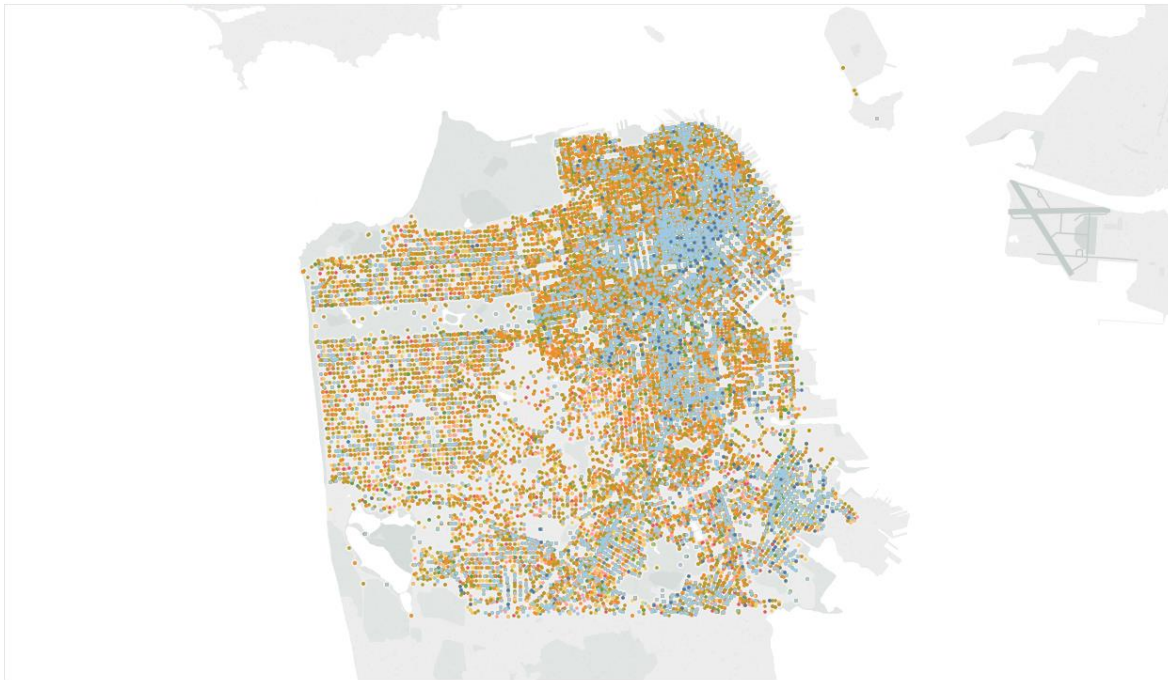
We are interested in public safety and crime rates in San Francisco. So we downloaded the public safety crime rate database for 2016 from SF Open Data data base (source: <https://data.sfgov.org/Public-Safety/SFPD-Incidents-Current-Year-2016-/9v2m-8wqu>). The data fields which especially interest us are Crime Category and location (by district and geographical coordinates):

IncidentNum	Category	Descript	DayOfWee	Date	Time	PdDistrict	Resolution	Address	X	Y	
160089099	ASSAULT	AGGRAVATED /	Sunday	1/31/2016	0:00	0:14	PARK	NONE	CLAYTON ST / P	-122.44877	37.77073127
160089102	WARRANTS	WARRANT ARR	Saturday	1/30/2016	0:00	18:30	SOUTHERN	ARREST, BO	800 Block of BR	-122.4034	37.77542071
120350636	WARRANTS	WARRANT ARR	Monday	7/11/2016	0:00	8:22	TENDERLOIN	ARREST, BO	0 Block of UNIT	-122.41432	37.77994441
160089118	OTHER OFFENSES	DRIVERS LICEN	Saturday	1/30/2016	0:00	12:50	BAYVIEW	ARREST, BO	PALOU AV / JEN	-122.38529	37.73079074
160089124	SUSPICIOUS OCC	INVESTIGATIVE	Sunday	1/31/2016	0:00	0:34	SOUTHERN	ARREST, BO	5TH ST / MISSIC	-122.40648	37.78273179
160089124	OTHER OFFENSES	DRIVERS LICEN	Sunday	1/31/2016	0:00	0:34	SOUTHERN	ARREST, BO	5TH ST / MISSIC	-122.40648	37.78273179
160089196	VANDALISM	MALICIOUS MI	Sunday	1/31/2016	0:00	1:23	NORTHERN	NONE	GREEN ST / VAN	-122.4238	37.79761624
160089522	SUSPICIOUS OCC	INVESTIGATIVE	Sunday	1/31/2016	0:00	7:46	NORTHERN	NONE	600 Block of ED	-122.41838	37.78325838
160089538	OTHER OFFENSES	DRIVERS LICEN	Sunday	1/31/2016	0:00	7:53	SOUTHERN	ARREST, BO	FOLSOM ST / 11	-122.41403	37.77186362

### 1. Tableau with OData Connection

First, we tried using Tableau. We imported the data through an OData connection to <https://data.sfgov.org/OData.svc/ritf-b9ki>. We then scaled the points to an appropriate size and color coded them according to crime category. This is similar to what we did for the parking meter data in class. The points are scaled so that we can see where they are more concentrated, indicating higher density. We could call this a quasi-heatmap. But it is not truly a heatmap because the elements are not adjacent (or continuous).

<(SF OPen Data) San Francisco Crime Incidents, 2016>



Map based on Location Longitude and location\_latitude. Color shows details about Category. The data is filtered on Number of Records as an attribute, which includes values less than or equal to 1. The view is filtered on Category, which keeps 13 of 39 members.

**Category**

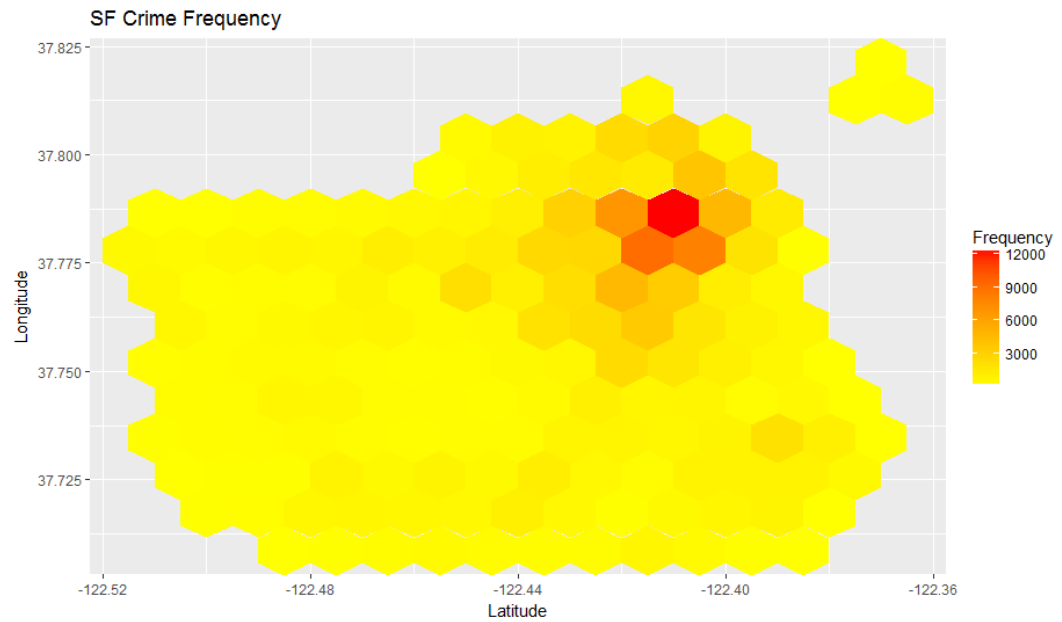
- ARSON
- ASSAULT
- BURGLARY
- DRIVING UNDER THE INFLUENCE
- DRUG/NARCOTIC
- KIDNAPPING
- LARCENY/THEFT
- OTHER OFFENSES
- ROBBERY
- STOLEN PROPERTY
- VANDALISM
- VEHICLE THEFT
- WEAPON LAWS

## 2. Ggplot with hexbins

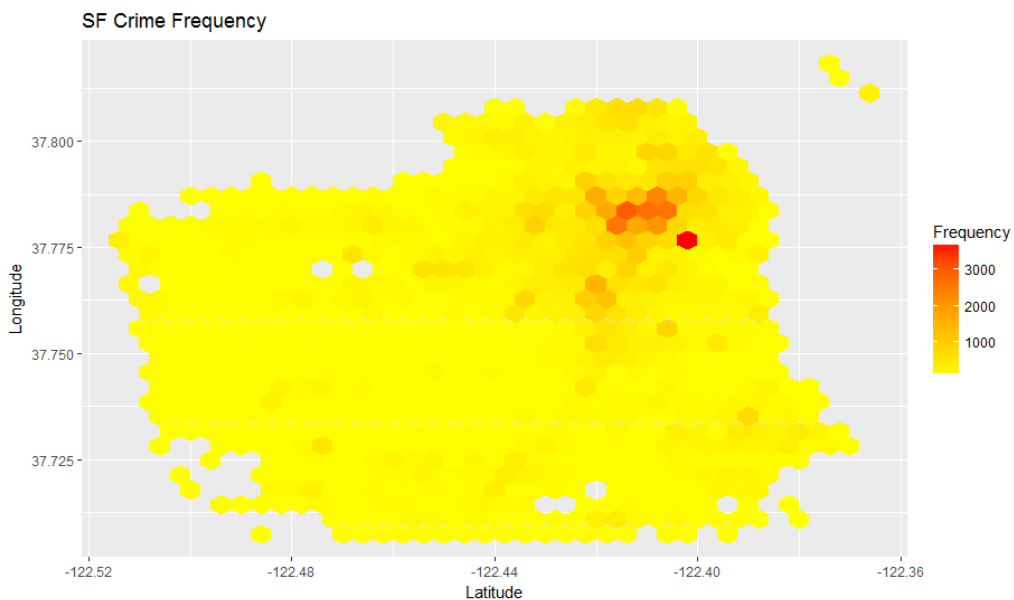
Our next attempt was to use ggplot, which we covered in the last week of class, along with the hexbin parameter, which allows us to bin the data in discrete but contiguous hexagonal bins, thus rendering a map that might be more appropriately called a heat map. What's cool is that we can easily scale the bins through a parameter adjustment:

```
SFPD_Incidents <-  
read_csv("C:/Users/Gui/Dropbox/STATS6610/HW4/SFPD_Incidents_  
_Current_Year__2016_.csv")  
  
ggplot(SFPD_Incidents, aes(X, Y)) + stat_binhex(binwidth=c(.01,.01)) +  
scale_fill_gradientn(colours = c("yellow", "red"),
```

```
name = "Frequency") + ggtitle("SF Crime Frequency")+ labs(x="Latitude",
y="Longitude")
```



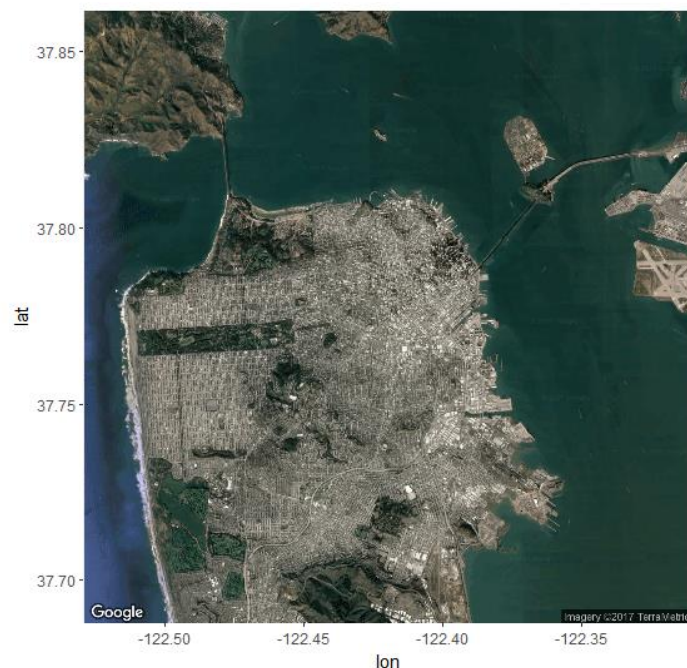
```
ggplot(SFPD_Incidents, aes(X, Y)) + stat_binhex(binwidth=c(.004,.004)) +
scale_fill_gradientn(colours = c("yellow", "red"),
name = "Frequency") + ggtitle("SF Crime Frequency")+ labs(x="Latitude",
y="Longitude")
```



### 3. Ggmap (of ggplot)

Finally, we use the `[ggmap]` package that is part of `[ggplot2]`. `[ggmap]` allows overlaying spatial data over Google Maps and OpenStreetMaps. We start by using `[getmap]` to plot the basemap. Then we overlay the crime data we have retrieved from OpenSF:

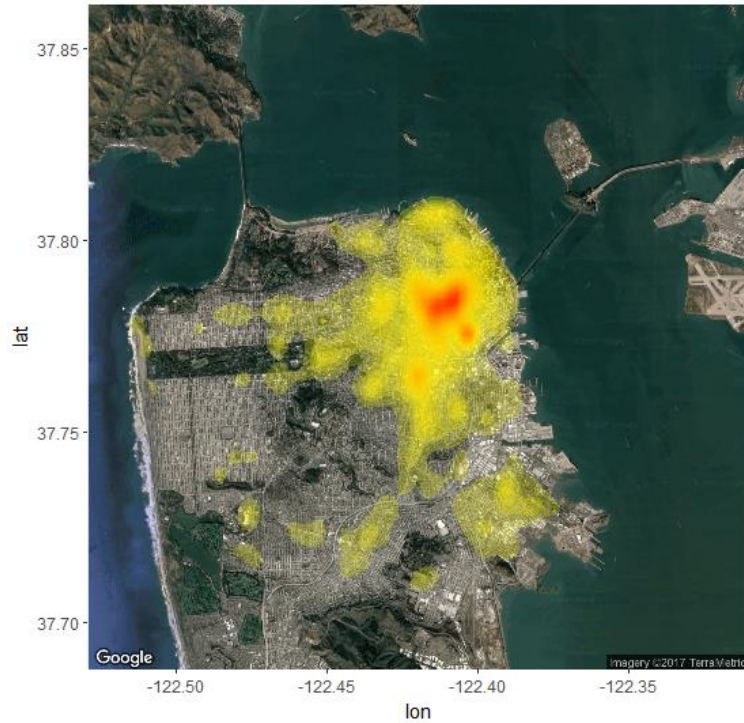
```
library(ggmap)
# Get the basemap
phBasemap <- get_map(location="San Francisco, CA", zoom=12, maptype = 'satellite')
# Take a look
ggmap(phBasemap)
```



```
ggmap(phBasemap) +      # make the heatmap
  stat_density2d(aes(x = X, y = Y,
                    fill = ..level..,      # value corresponding to discretized density estimates
                    alpha = ..level..),
  bins = 25, # number of bands
  data = SFPD_Incidents,
  geom = "polygon") + # creates the bands of different colors
  ## Configure the colors, transparency and panel
```

```
scale_fill_gradient(low = "yellow", high = "red") +  
scale_alpha(range = c(.25, .55)) +  
theme(legend.position="none")
```

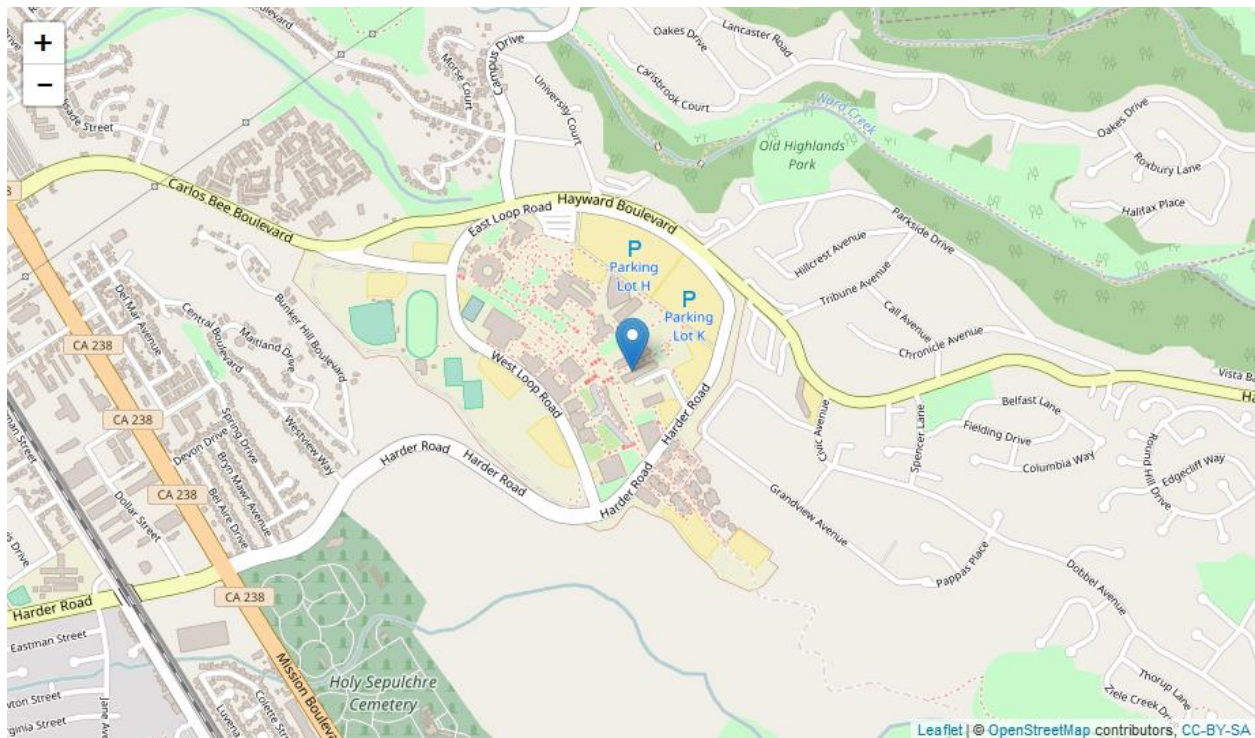
And what you see (below) is what you get!



#### 4. CSUEB The Birthplace of R!

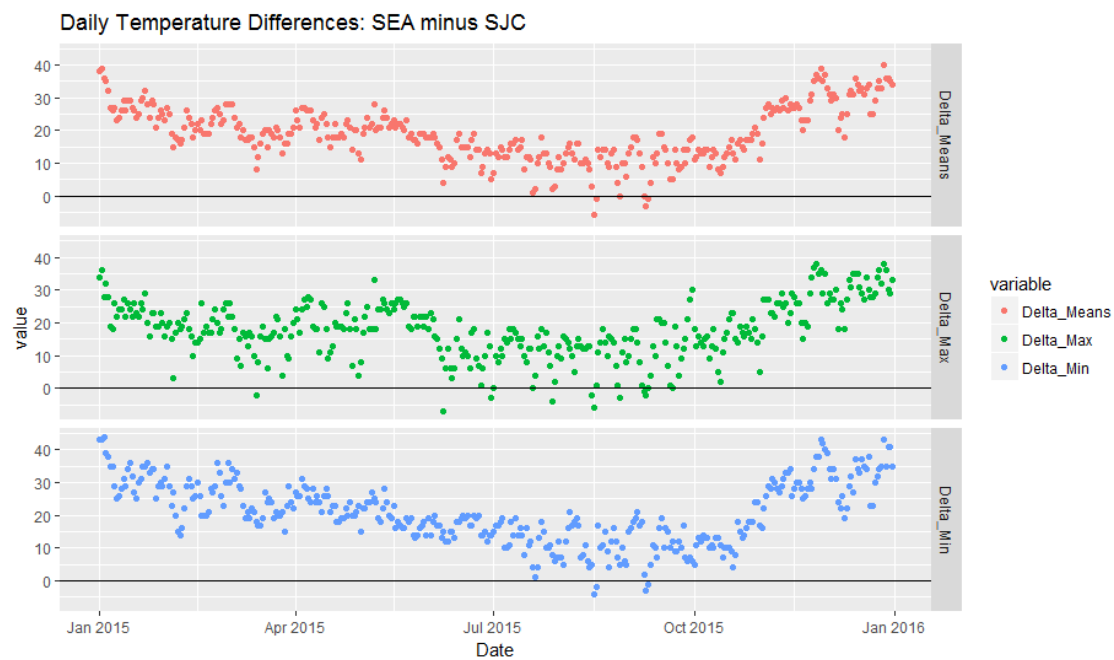
Using the Leaflet package below. (No modifications).





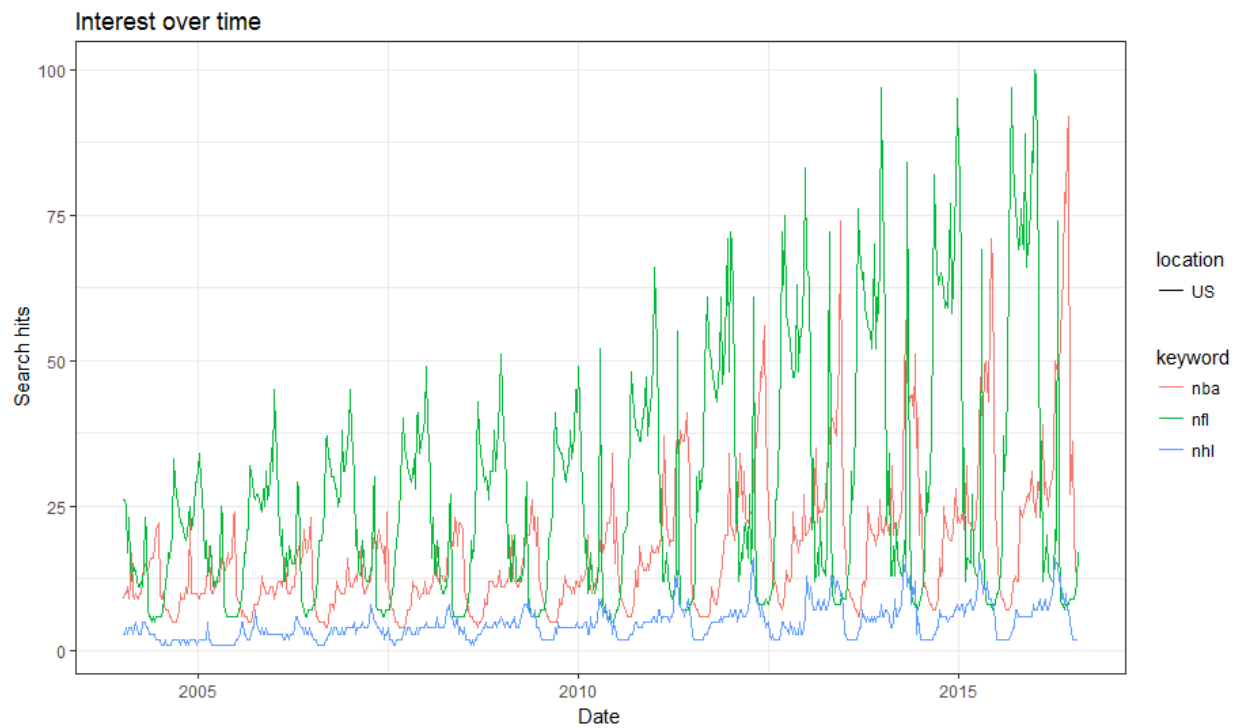
## 5. Weather Report

I replaced SFO with SJC (San Jose, CA) and was very surprised to find the shape went from convex to concave. SJC has warmer Summer but colder Winter than SEA.



## 6. Google Trends

Very interesting to observe the TS components growth and seasonal patterns, with NFL and NBA clearly alternating.



## 7. Behold the 3-d t-distribution

Fat tail and sharp teeth!

