DSO 545 : Statistical Computing and Data Visualisation

**Homelessness in the city of LA**

Census, Crime and 311calls distribution

horizontal line

# Placeholder image

# Francis Chen Louis Yansaud Saksham Grover

# Jae Min Lee Irem Kahyaoglu Abhishek Ashok

# Introduction Homelessness is soaring in Los Angeles and is a severe issue that is a part of everyday life. Our objective for this project:

* Determining the need, demand and potential effects of different strategies for homeless intervention.
* Identifying the areas with high-concentrations of unsheltered homelessness, crime incidents and high danger areas for the homeless community in order to help address the homelessness in full scope, while prioritizing service delivery needs

In order to gain insight on homelessness:

1. We cleaned the data and used geospatial analysis and predictive modeling to create an online mapping application.
2. This online mapping tool helped us evaluate the effectiveness of existing and potential homelessness interventions, functioned as a framework to determine fast delivery alternatives as well as helping us analyze data in full-scope and allow the rapid deployment of resources and services.

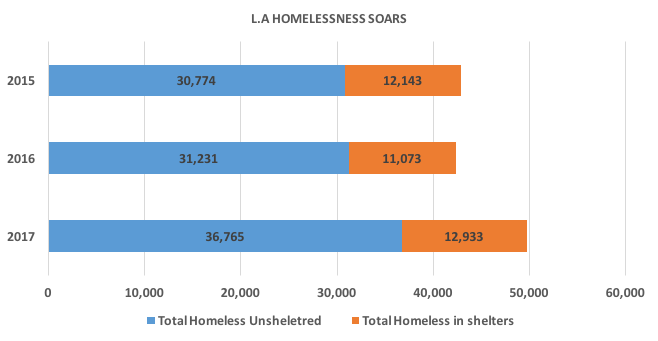
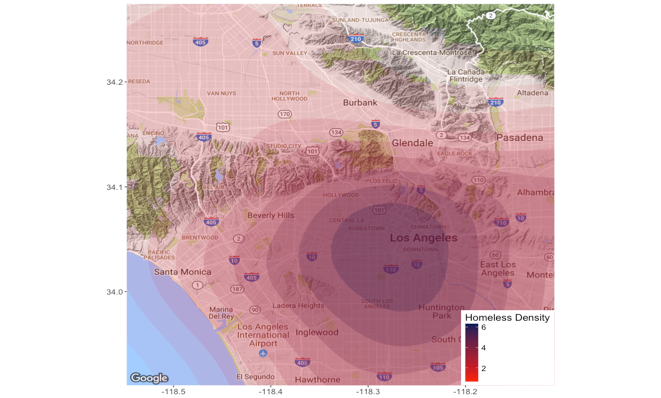
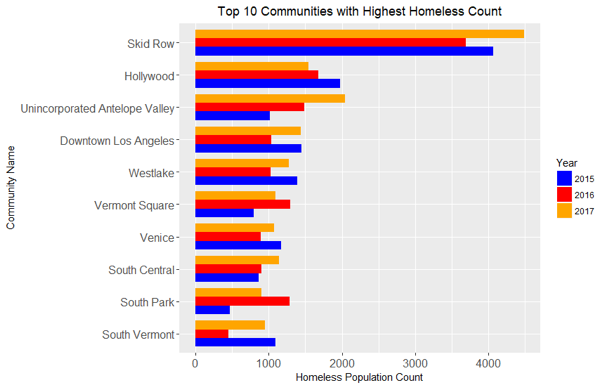
## Data Understanding For the data understanding phase, we started with initial data collection and we proceeded with analytical activities:

* Becoming familiar with the data
* Identifying data quality problems
* Discovering first insights into the data
* Detecting interesting subsets to form hypotheses regarding hidden information.

We mainly focused on the 6 dataset provided by the City of Los Angeles :

* HC 15-17 (Los Angeles Homeless Count 2015, 2016, 2017) – the estimate of homeless population in the Los Angeles geographic area
* Homeless311
* HomelessSheltersService
* CrimeHomelessVictims

We then combined all 6 datasets into one master large data set by their “TractID.” Once we have the master dataset, we then used the library package (ggmap) and the geo-spatial function on R to visualize the homeless population on a geospatial map of Los Angeles with different categories.

Before we got into the Data Preparation phase, we wanted to understand and visualize the big picture of the L.A homeless population from 2015 to 2017.   
  
From 2016 to 2017, we saw the sheltered homeless population increased by 1,860 while Unsheltered increased by 5,534.   
  
Once we had a better understanding of the homeless population, we can then plot on the map to see where the homeless density is mostly concentrated.  
  
  
  
  
And finally for the last phase of the data understanding, we then used the ggplot function on R to analyze the top 10 communities with the highest homeless count to determine where that increase came from. 

## Data Preparation

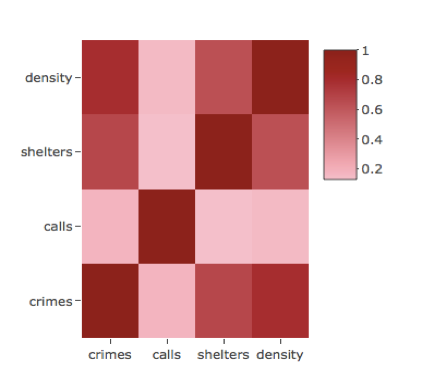
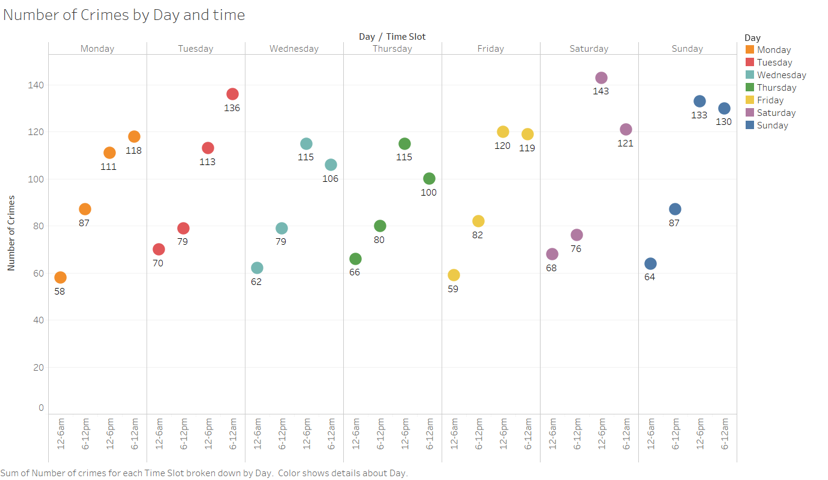
The Census data was given for uniques census tract ID.

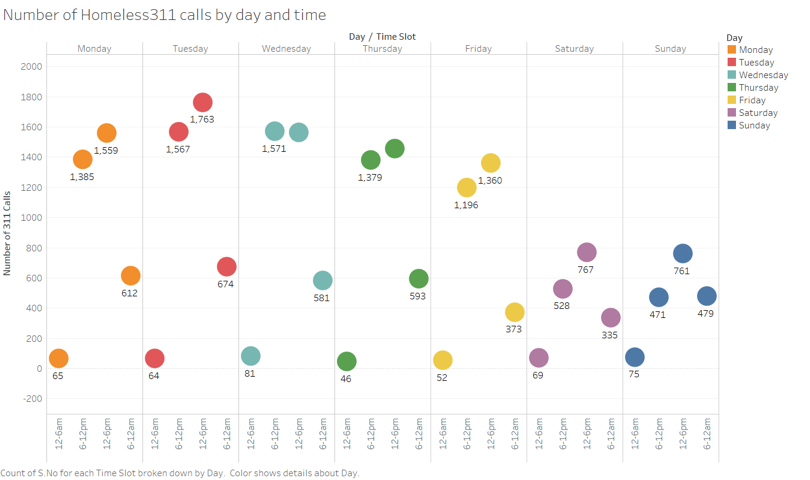
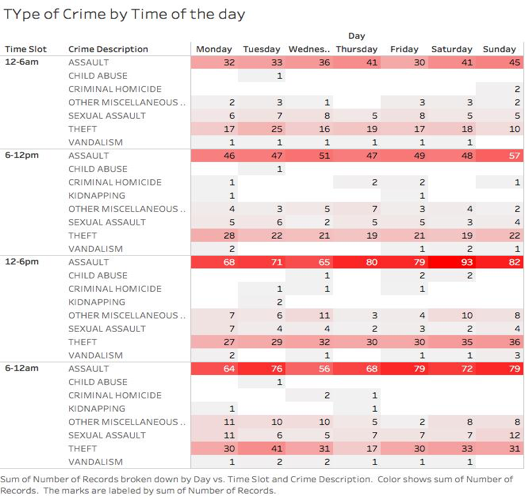
1. We were given the lat long values for crimes, shelters and 311call data and we converted these lat long values to their respective census tract ID with the help of shape file.
2. So then we merged the different data sets from Census, Crimes, No. of shelters and 311call data by Census tract ID to create a master data file.

As the police jurisdiction and administration is community wise, it made more sense to look at these different datasets community wise:

* We then aggregated the data by community name.
* We defined homeless density as a uniform measurement across the communities which was number of homeless people divided by Area of community.

*Homeless Density = (Number of Homeless)/(Area (sq. mile))*

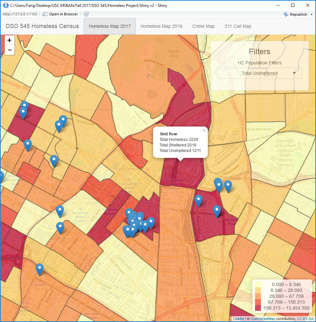
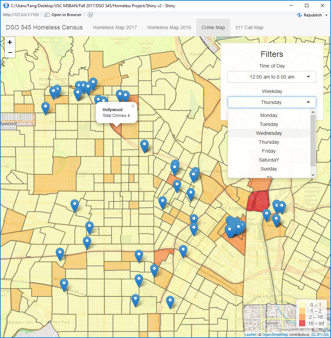
* We plotted a correlation matrix between the homeless density, number of homeless311 calls, crime counts and number of shelters in the matrix.  
    
    
    
    
    
    
    
  To answer some specific questions about the different dataset:
* We studied the distribution of our crimes and 311calls by date time.
* We created a new timeslot variables which categorised the time of activity into 6hr blocks.
* We defined them as morning, afternoon, evening and night time slot.
* We also categorized the data points into different days of week and plotted the distribution of the data point by day and time slots.   
    
  

## Shiny Details

The Interactive Map displays the city of Los Angeles and its homeless population in an interactive map.

The collapsable Map Controls drop down has the following features:

* Change the homeless measure distribution by either Census Tract or Communities
* Filter the homeless population by year and measure.
* Filter the Crime counts and locations by time of day, day of the week, and month of the year.
* Filter the Call counts and locations by time of day, day of the week, and month of the year
* Note that the 24 hour day has been split into 4 buckets:
  + Morning: 12:00 AM to 6:00 AM
  + Afternoon: 6:00 AM to 12:00 PM
  + Evening: 12:00 PM to 6:00 PM
  + Night: 6:00 PM to 12:00 AM
* As well, above the legend in the bottom right corner, the user can select the different layers to display: Homeless Measure, Shelters, Crimes, and/or 311 calls.  
    
    
  

**Crimes and Calls Over Time**

The Crimes and Calls tab displays a plot of the distribution of 311 calls or crime counts across the week and then also buckets them into the time of day following the above mentioned splits. Crimes can be further filtered based on the type of crime from the drop down menu on the left.

**Homeless Comparisons**

The Homeless Comparison tab displays the 2017 and 2016 homeless measures side by side to give a quick visual clue into the year over year homeless distribution change across the city of Los Angeles. Similar to the interactive map, the plot can be changed between viewing census tracts or communities. In addition, the same homeless measure filters can be applied, such as viewing the Total Homeless or Total Unsheltered distributions.

## Conclusion / Recommendation

To conclude this project: we analyzed and visualized the trend of homelessness and associated crimes with the merged dataset by census tract:

* Calculated the correlation between the homeless density and 311 calls.   
  a) We observed a high correlation between density and crimes counts 0.87 – which raised the question that whether high crime rates were a threat to the homeless population or were the homeless causing these crimes.
* Analyzed the distribution of crimes and 311 calls across time. a) Number of 311 calls was the highest on Tuesday and during the afternoon hours

b) Crime rate is highest on Saturday afternoon, followed by Tuesday evening and Sunday afternoon and evening.  
c) The most frequent type of crime is Assault on Saturday afternoon

* Used our Shiny Dashboard to understand which factors could be significant for the unsheltered homeless.

Skid row being at the top even though it has the largest number of shelter counts. Worth noting that calls coming from that area are comparatively quite low and it experiences high crime counts as well.

For recommendations:

* Investigate and access to homeless satisfaction in shelters
* Analyze average age of homeless in shelters to estimate the average age of the total homeless population per census tract. By doing so, we can have then use this data to implement better strategies and procedures to help homelessness.
* Collect more data from successful project to support the decreasing rate of unsheltered homeless.

