Descriptive Analysis Document

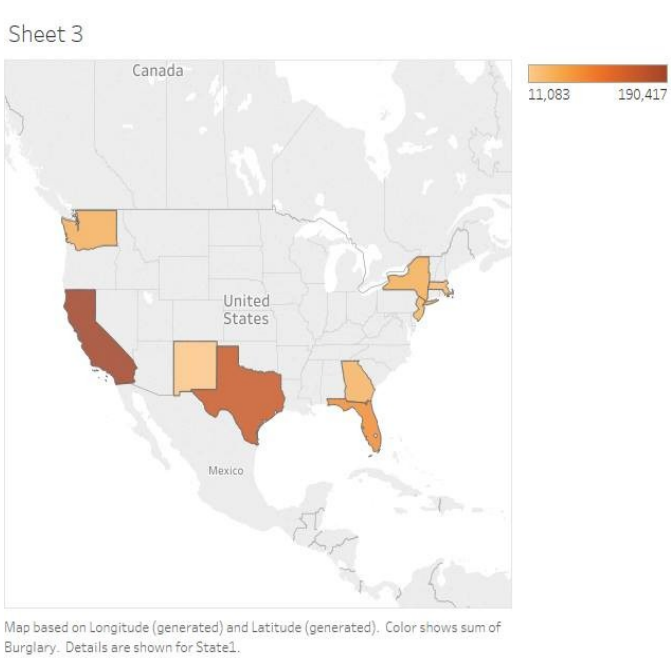
Vivek I(IMT2013017)
Aravind K(IMT2013019)
Nikhilesh P(IMT2013031)
Arun Prasad(IMT2013037)

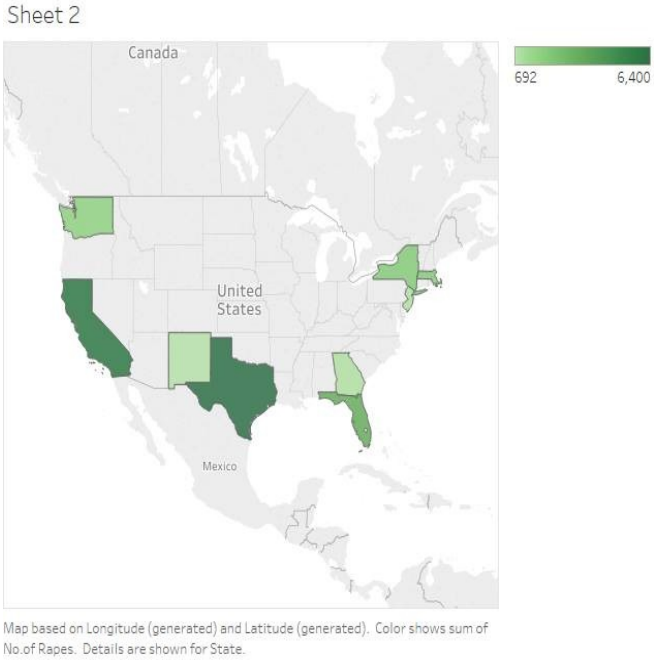
Before we start analysing our data, we need to do the required pre-processing over it inorder to avoid any kind of faulty inferences. So getting specific to our FBI crime data set, the pre-processing we done is the following:

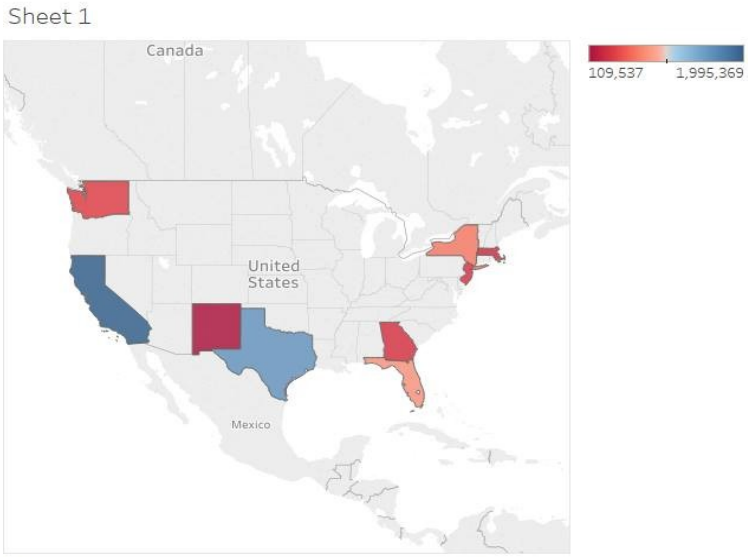
- Few of the states have an extra column regarding the number of rapes given by an agency which is empty and so we removed it and rearranged the sheet.
- Some of the quantites that we used for the descriptive analysis like the CityCrime Index, need the data regarding the average of a particular attribute values. So we created few extra columns/sheet and recorded the data obtained by averaging them so as to make our work clear while plotting the graphs.
    - Before calculating the average we excluded the outliers so that we get a more reasonable value, yet we did not remove them from the data set so as to avoid any kind of data loss.
    - During this process, we identified few of our data points to be incorrect(i.e. they were not matching with the actual values, for example the total population should be greater than the number of people effected by a particular crime and in few cases it was not) and so we removed them from our data set to avoid getting into wrong comclusions.

After doing this necessary pre-processing, we have summarized some useful information about our data in the following reports. All the reports have been generated using tableau.

<u>Crime-State variations</u> :







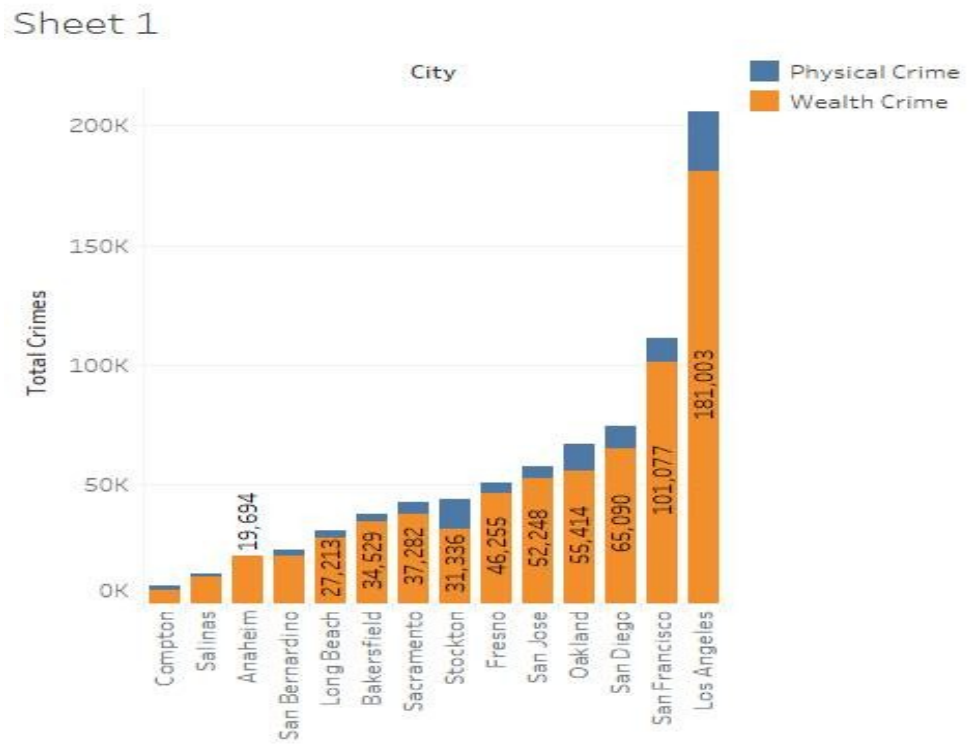
Map based on Longitude (generated) and Latitude (generated). Color shows sum of Total Crimes. Details are shown for State.

The above representation shows the level of burglary(One of the Property crimes),rapes(One of the Violent crimes) and total number of crimes that are being commited in these states(the highlited areas in the map). With the help of this representation we can get into conclusions as in which states are being badly effected and which are not very prone to this particular crime.

Using our data set we calculated the total number of rape victims in that state and plotted it onto the map using a colour scale. Similarly method has been followed for the burglary crime and total number of crimes.

From this graph,we can easily observe that California is more in the number of crimes either in Rape (or) Burglary (or) Total number of crimes and New Mexico is least in the number of crimes either in Rape (or) Burglary (or) Total number of crimes

<u>Wealth & Physical crimes</u> :



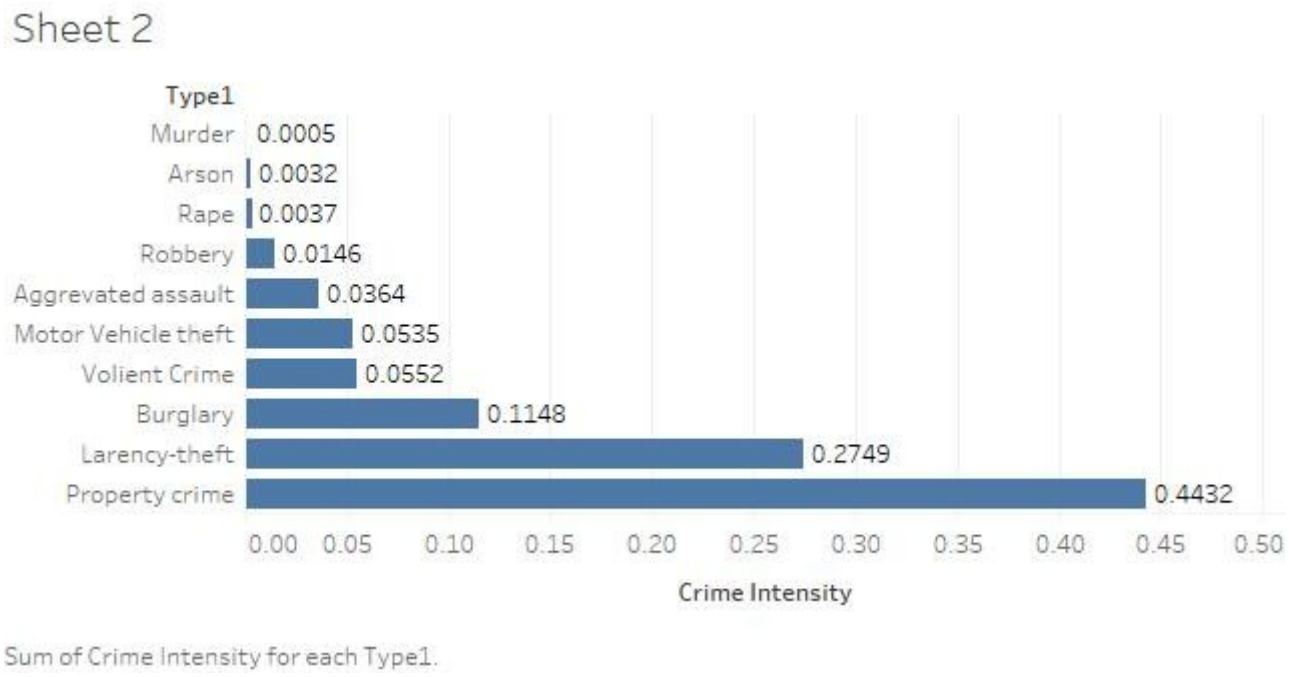
Sum of Total Crimes for each City. Color shows details about Type.

We have categorized the crimes into Wealth Crimes(where harm is done to wealth of the people) and Physical Crimes(where harm is done to people). Wealth crimes include Robbery, Burglary, Larency Theft, Motor Vehicle Theft, Arson and Physical crimes include Murder, Rape, Aggrevated Assault.

The above stacked bar chart shows the portion of wealth and physical crimes in the overall crimes commited for few of the places in California. As the description suggests, each bar shows what fraction of the total crimes that are being commited in that place involves physical harrasment and what fraction involves loss of property.

From this graph, we can conclude that Wealth Crimes happen more as compared to Physical Crimes.

<u>Crime Intensity</u>:



The above bar chart between the type of crime and their corresponding crime intensities for a state is generated in order to get a clear picture of what kind of crime is being commited widely across the state and to what extent are they being commited, so that the authorities can focus primarily on those crimes by taking precautionary measures.

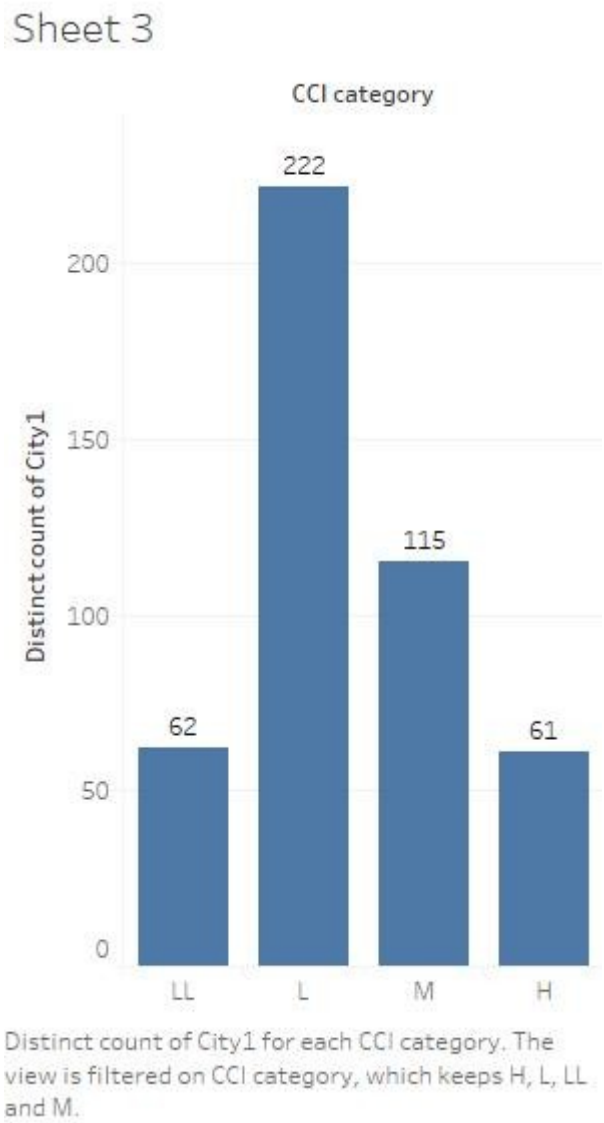
The method followed to determine this crime intensity is as follows: for a place we calculate the ratio of the number of people effected by a particular crime and total number of crimes commited. For that particular crime we calculate this ratio for all the places in the state considered and determine an average of these ratios resulting in the crime intensity of that particular crime.

From the above graph, we can conclude that Larency-theft(Property Crime) which has more crime intensity and Murder(Violent Crime) has less crime intensity.

<u>City Crime Index(CCI) to Number Of Cities</u> :

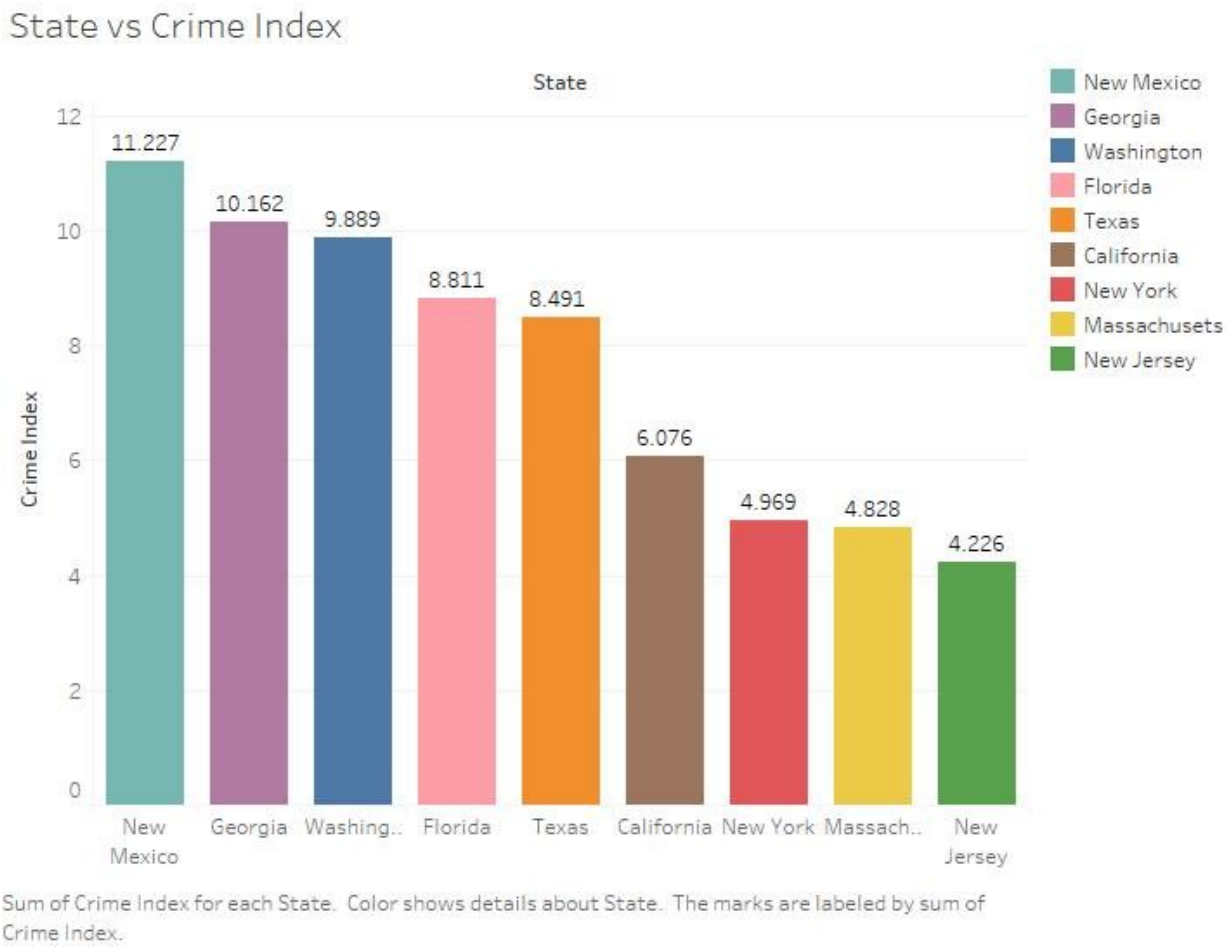
For each city(in California), the ratio of number of crimes in that city to the total polulation of that city is calculated and termed as CCI(City Crime Index). Based in these CCI valuesof the cities they have been divided among various categories namely LL, L, M and H. The below graph shows the number of cities in each category.(Categories along X-axis are in the order of LL,L,M and H i.e, lowest range of CCI to highest range if CCI).

LL----CCI index of the city is >=0 and <3
L------CCI index of the city is >=3 and <6
M-----CCI index of the city is >=6 and <9
H-----CCI index of the city is >=9
These ranges are decided based on average of the CCI(of all cities).
This can be used to know the crime situation in the state as a whole. If number of cities in LL,L category are more compared to number of cities in M,H category, we can say that crime situtation in the state is little controlled where as if number of cities in M,H category are more compared to number of cities in LL,L category, we can say that crime situtation in the state is out of control and proper measures need to be taken.



From this graph, we can conclude that more cities fall in the category L which has crime index in the range of 3 to 6.

State Crime Index :
The graph is similar to the above graph but Crime index(Total crimes in the state/Population of the State) for 9 different states is calculated and plotted. This can be used to compare across the states.



From this graph, we can conclude that among the 9 states considered New Mexico has more crime Index and New Jersey has least crime index.

<u>Crime across States</u> :

A particular crimet(rapes) is chosen and the following quantity is calculated(for that crime) for the 9 different states considered and plotted.

$$\frac{Number\ of\ Crimes\ of\ that\ particular\ crime\ \text{in}\ that\ State}{Total\ Population\ of\ the\ 9\ states\ considered} \times Population\ of\ that\ state$$

This can be used to compare the situation of a particular crime across different states. This graph is little different from the first graph of crime-state variations since here the number of crimes is normalized with total poulation of that state.

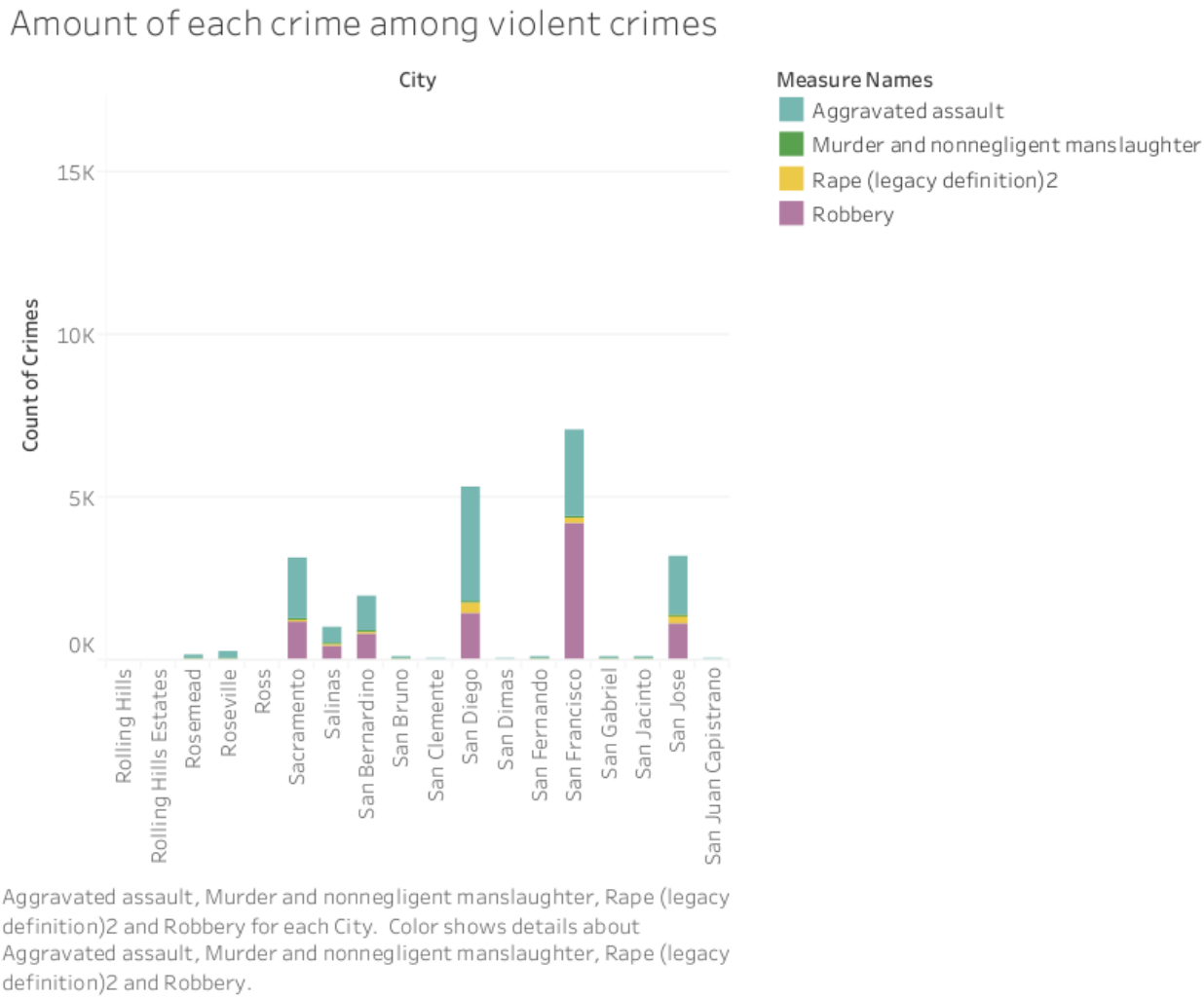
Say a woman wants to live in one of the 9 places in a state and so wants to know which place to choose based on the rapes situation in different places. Also, she wants to take the poluation of that place(may be because the facilities in a city is effected by the population) into consideration, then the first graph will be more useful than this.

The shown graph is for Rapes. Similar thing can be done for each crime.

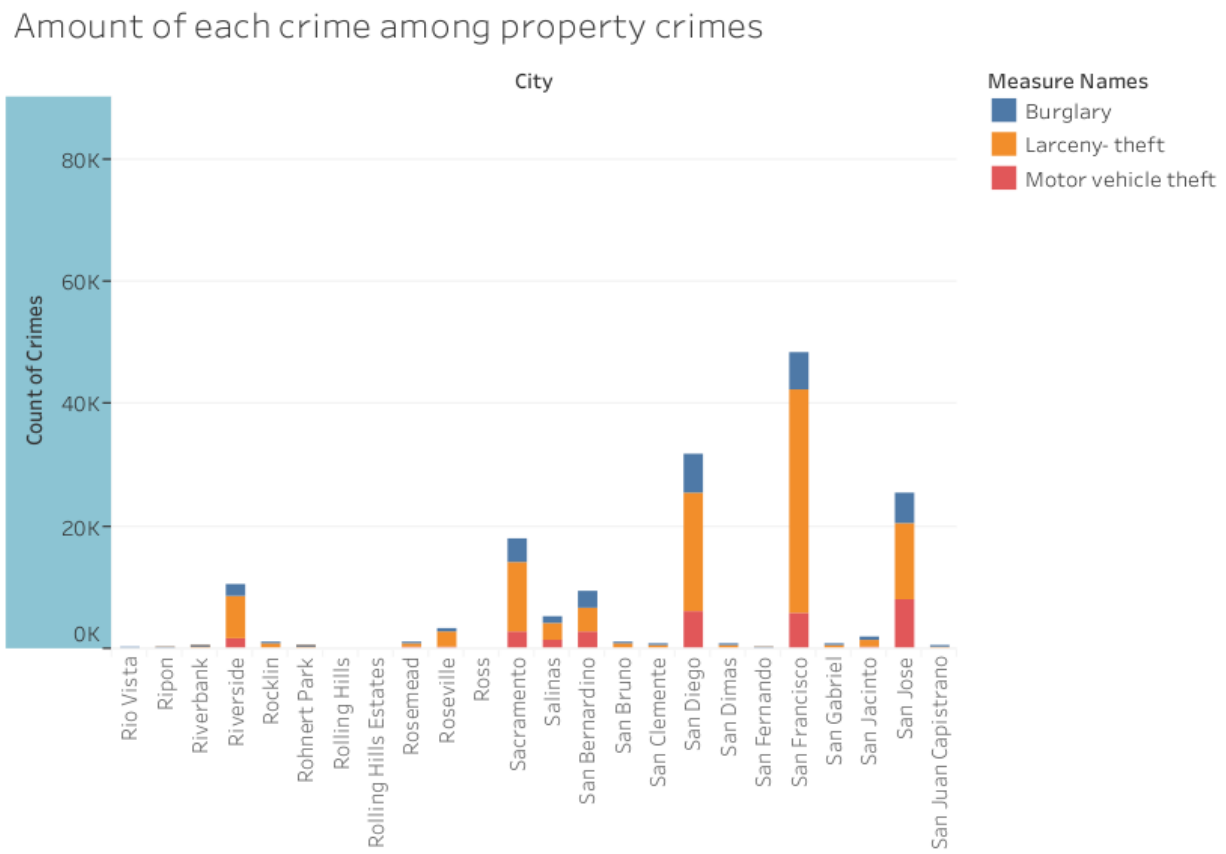
From this Pie chart, it's clear that Crime intensity of rapes is more in Claifornia.

<u>Property Crimes and Volent Crimes</u> :

Different types of crimes in our data are categorized into Property Crimes and Violent crimes. The following two graphs show different proportions of each of the property and voilent crimes. Due to space constraints, only few cities have been considered.



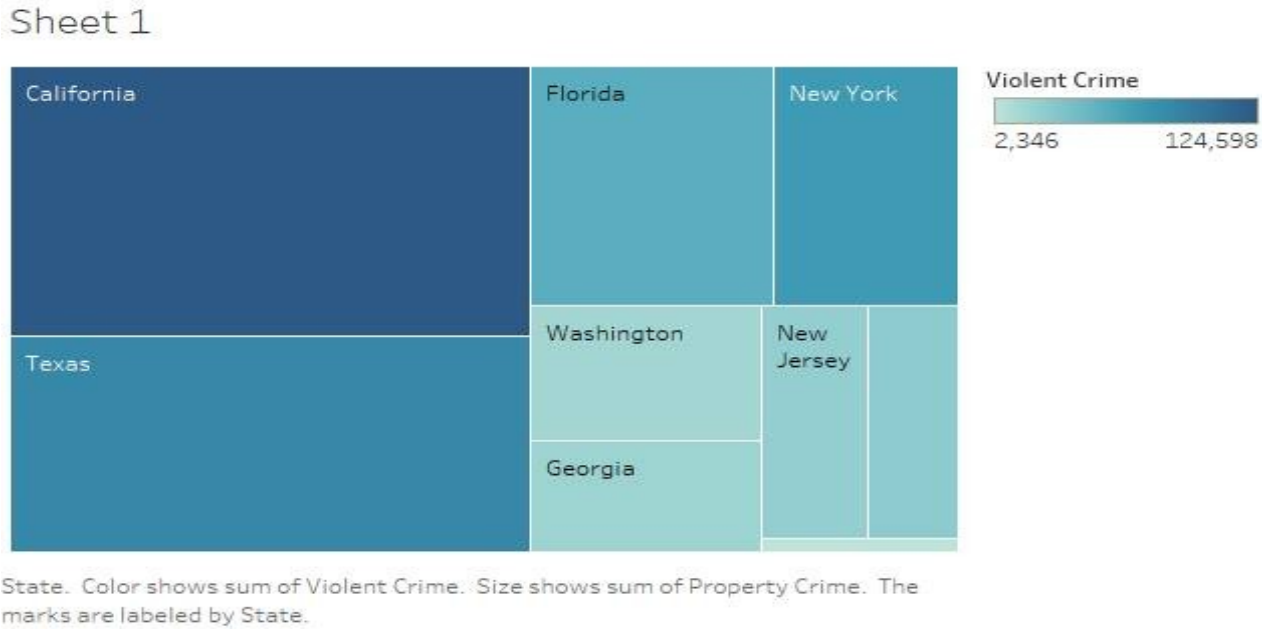
From this graph it's clear that among Violent crimes, Robbery accounts to more percentage in some places where as in some other places Aggrevated assault accounts to more percentage.



From this graph it's clear that among Property Crimes, Larency-theft accounts to more percentage.

<u>Property and Violent Crimes across States</u> :

The following tree map shows the number of Violent and Property crimes accross 9 different states considered.



These kind of tree maps are useful to compare the change along two numerical quantities. In the above tree map, the shade of NewYork is little more compared to the shade of Florida which implies more Voilent Crimes in NewYork compared to Florida where as the area of rectangle is more for Florida compared to NewYork which implies there are more property crimes in in Florida compared to NewYork.

From this tree map, we can conclude that number of Violent Crimes(represented by shade of the rectangle) and the number of Property Crimes(represented by area of the rectangle) are more in California state and least in NewMexico(right bottom small rectangle)