Los Angeles City Crime Analysis

Naga Jyothi Kunaparaju

Data

Dataset for this project is collected from:

https://catalog.data.gov/dataset/crimes-2012-2015.

Dataset consists of 760,000 rows and 15 Attributes, including:

Timeocc, Area, CrmCd, CrmCdDesc, Status, Location, CrossStreet, Lat, Long, DateRptd, Dateocc

Method

- Pandas and Pyplot Libraries are used to create plots
- In raw dataset, Lat & long combined into single column. For Geographical representation, it has been split into two columns.
- Data set is also loaded into Excel, Tableau for cleansing and formatting.
- Crime Dataset is maintained by LAPD.

```
(763739, 15)
```

[113]:
```
crimeData.head()
```

[113]:

	Timeocc	Area	Areaname	RD	CrmCd	CrmCdDesc	Status	StatusDesc	Location	CrossStreet	Lat	Long	DateRptd	Dateocc
5728	5	16	Foothill	1612	997	TRAFFIC DR #	IC	Invest Cont	GLENOAKS BL	FILMORE ST	34.2800	-118.4183	11/13/14	11/6/14
5729	15	16	Foothill	1681	997	TRAFFIC DR #	IC	Invest Cont	ROSCOE BL	LAUREL CANYON BL	34.2217	-118.4100	11/13/14	11/13/14

```
from pandas import *
#crimeData['Dates']=pd.to_datetime(crimeData['CRIME_DATE'])
#crimedata_date = pd.to_datetime(crimeData['CRIME_DATE'], format='%Y%m%d', errors='ignore')
#dat = crimeData[1]

#print(crimena)

crimeData['Dateocc']=pd.to_datetime(crimeData["Dateocc"])
crimeData['Yearocc']=crimeData['Dateocc'].dt.year
crimeData['Monthocc']=crimeData['Dateocc'].dt.month

month_map={1:'Jan', 2:'Feb', 3:'Mar', 4:'Apr', 5:'May', 6:'Jun', 7:'Jul', 8:'Aug', 9:'Sep', 10:'Oct', 11:'Nov', 12:'Dec
crimeData['Monthocc'].replace(month_map, inplace=True)
crimeData['Yearocc']=crimeData['Yearocc'].astype(str)
#crimeData['HourRptd']=crimeData['HourRptd'].astype(str)


YearlyData = pd.DataFrame(crimeData["Yearocc"].value_counts())
HourlyData = pd.DataFrame(crimeData["Timeocc"].value_counts())
MonthlyData = pd.DataFrame(crimeData["Monthocc"].value_counts())

AreaData = pd.DataFrame(crimeData["Area"].value_counts())
AreanameData = pd.DataFrame(crimeData["Areaname"].value_counts())
crimeData.head()

#crimeData.head()
```

Method contd...

This data is divided into multiple datasets, for example:

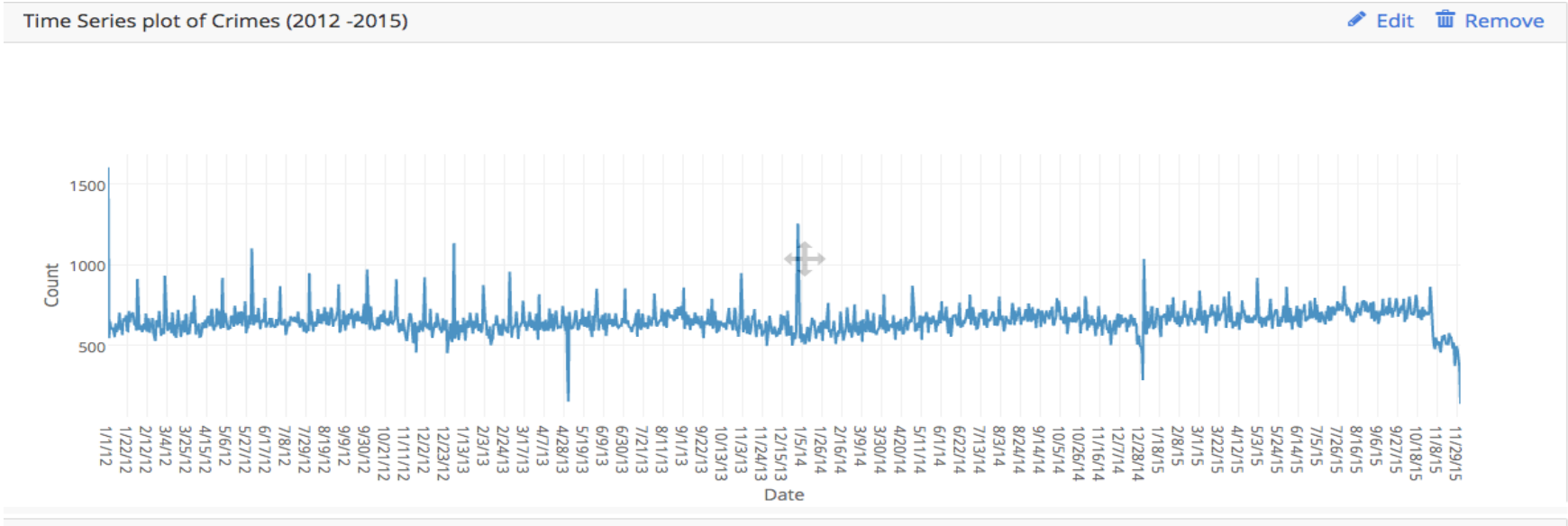
1. Time variant such as Daily, Monthly, Hourly

2. Geographical information split into area and location.

3. List of Violent crimes with crime codes 1 to 8 are separated into a different dataset for analysis.

Plots are drawn with this data using Pyplots and plotly

Timeseries Plot 2012-2015 https://plot.ly/~nkunaparaju/13/

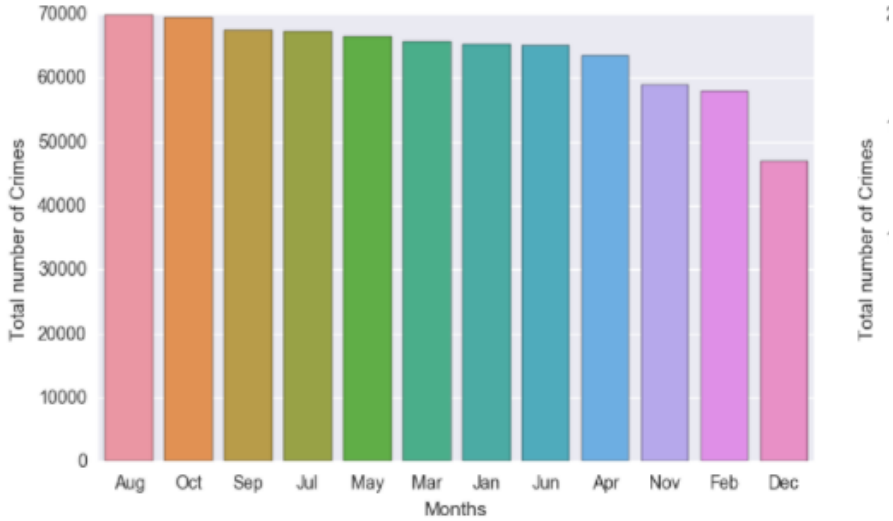


Hourly total 2012 -2015 - https://plot.ly/~nkunaparaju/11/

Total number of crimes: Monthly

In [166]:
```
plt.figure(figsize=(16,10))
ax3 =  plt.subplot2grid((2,2),(1,0))
ax3.set_ylabel('dxzvxvxzv')
sns.barplot(x=MonthlyData.index, y="Monthocc", data=MonthlyDat
ax3.set(xlabel='Months', ylabel='Total number of Crimes')
ax4 =  plt.subplot2grid((2,2),(1,1))
ax4.set_title('Yearly')
sns.barplot(x=YearlyData.index, y="Yearocc", data=YearlyData)
ax4.set(xlabel='Year', ylabel='Total number of Crimes')
```

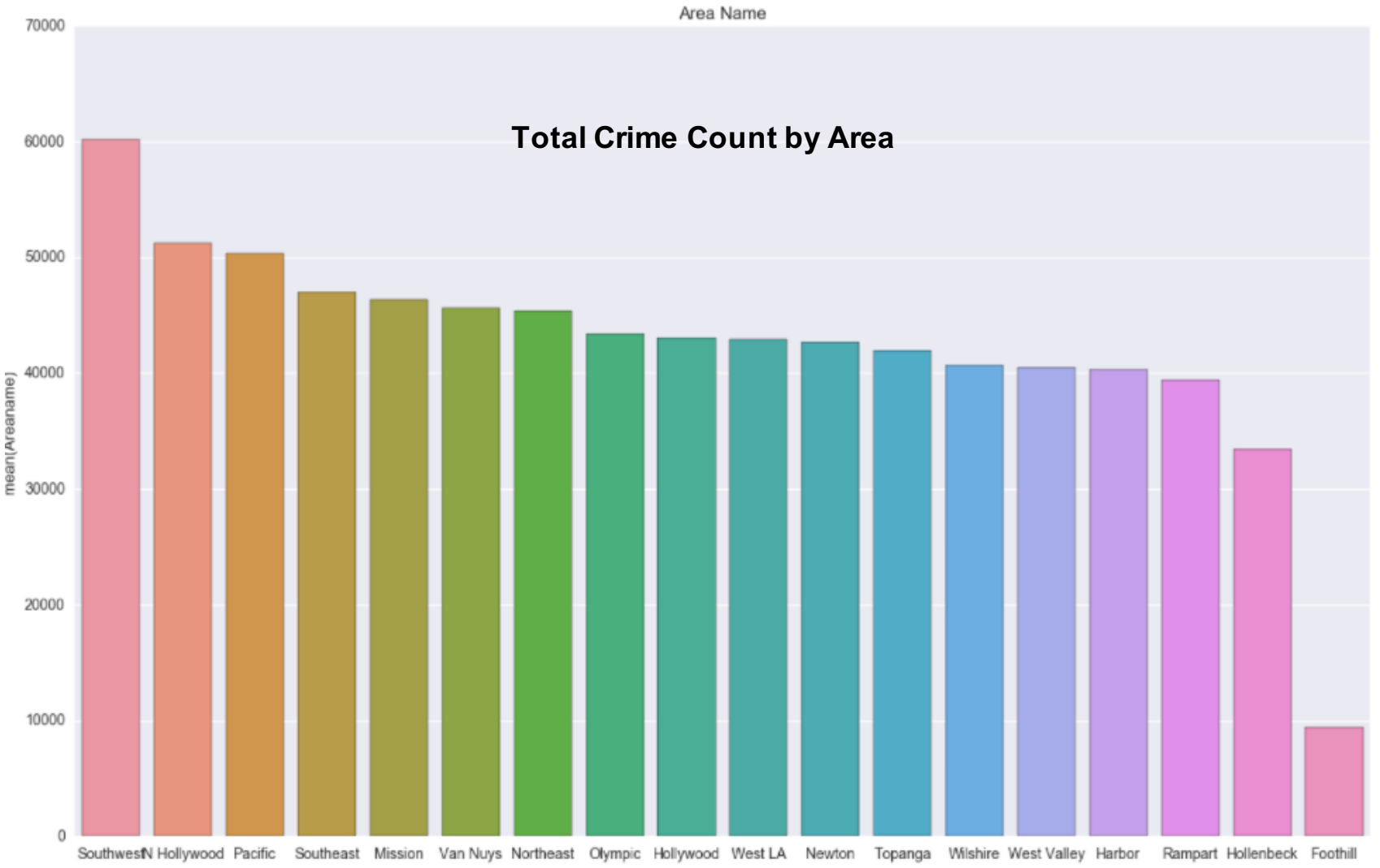
Out[166]:
```
[<matplotlib.text.Text at 0x138228710>, <matplotlib.text.Text
```

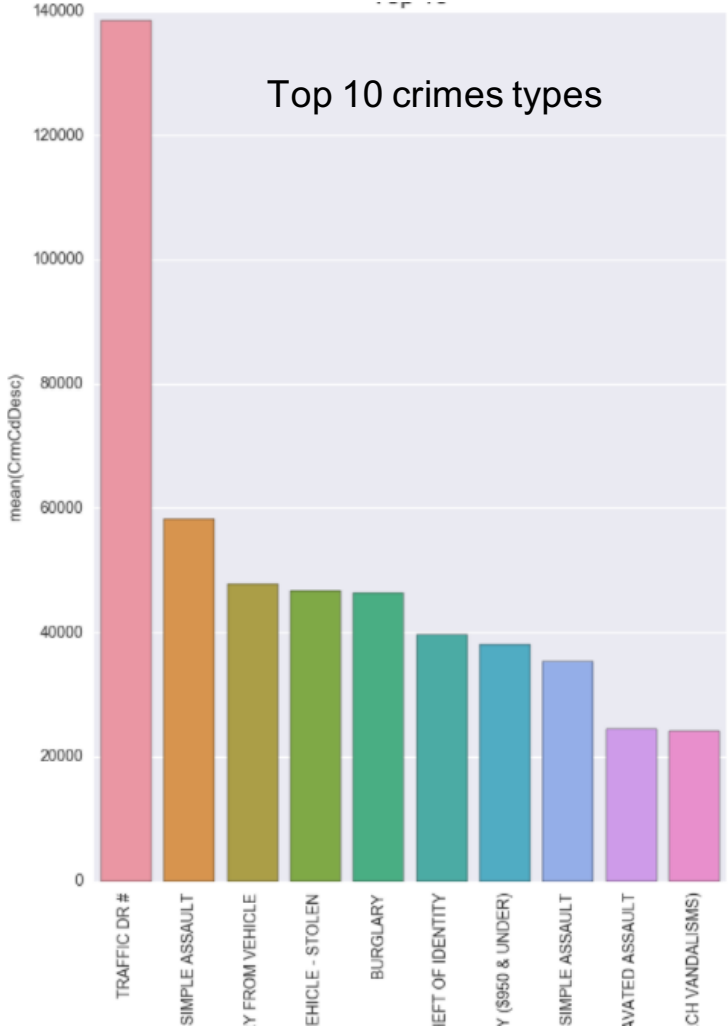


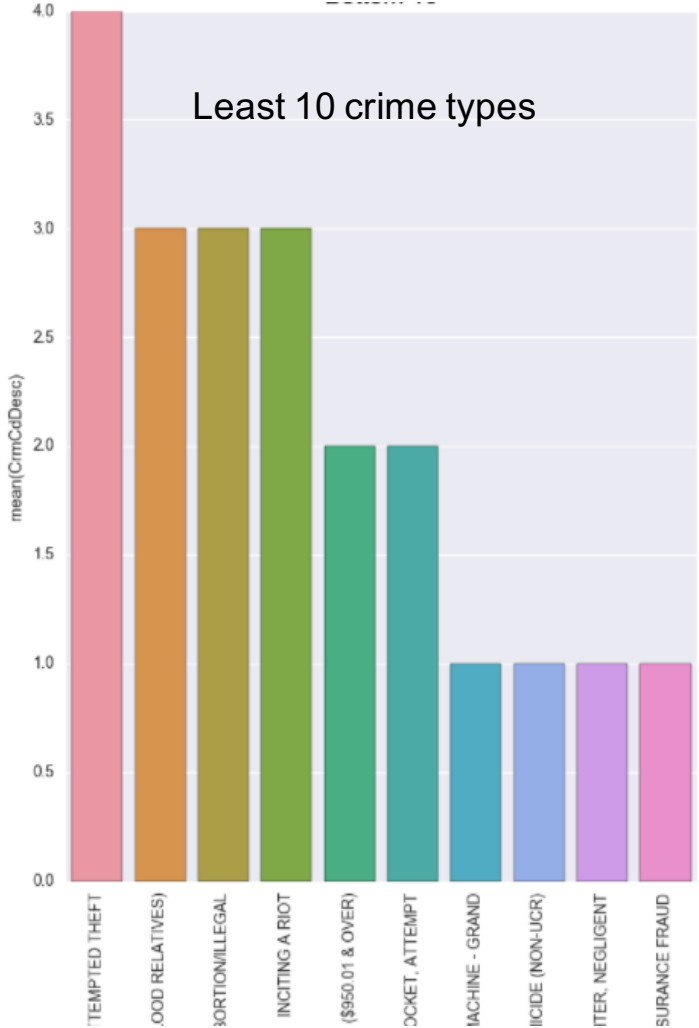
Total number of crimes for each year

```
g= sns.factorplot(x=YearlyData.index, y="Yearocc", data=Yearly
g.set_axis_labels('Years','Count by year ')
```

```
<seaborn.axisgrid.FacetGrid at 0x1240a9dd8>
```

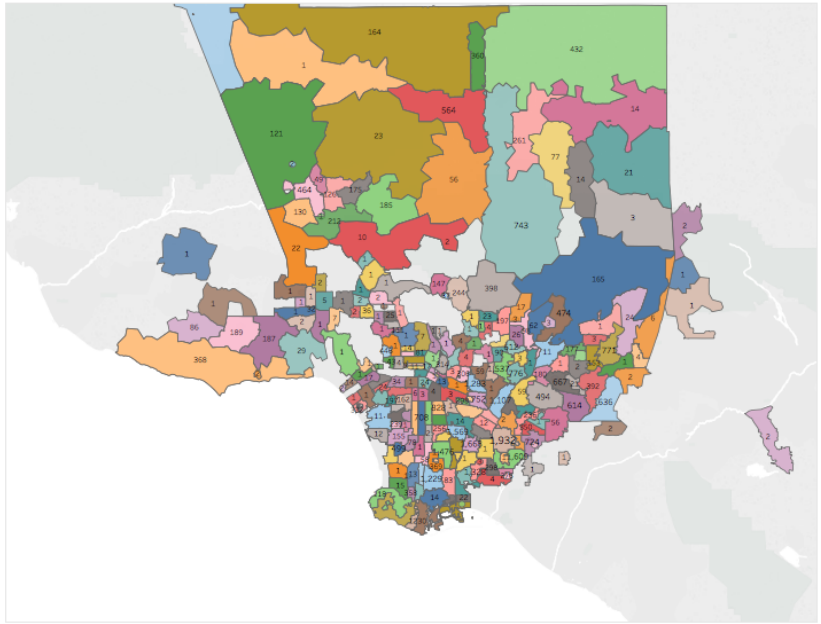

Area Name
Total Crime Count by Area
mean(Areaname)
70000
60000
50000
40000
30000
20000
10000
0
Southwest
N Hollywood
Pacific
Southeast
Mission
Van Nuys
Northeast
Olympic
Hollywood
West LA
Newton
Topanga
Wilshire
West Valley
Harbor
Rampart
Hollenbeck
Foothill


Top 10 crimes types
mean(CrmCdDesc)
140000
120000
100000
80000
60000
40000
20000
0
TRAFFIC DR #
SIMPLE ASSAULT
Y FROM VEHICLE
EHICLE - STOLEN
BURGLARY
IEFT OF IDENTITY
Y ($950 & UNDER)
SIMPLE ASSAULT
RAVATED ASSAULT
CH VANDALISMS)


Least 10 crime types
mean(CrmCdDesc)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
TTEMPTED THEFT
OOD RELATIVES)
ORTION/ILLEGAL
INCITING A RIOT
($950.01 & OVER)
OCKET, ATTEMPT
MACHINE - GRAND
ICIDE (NON-UCR)
TER, NEGLIGENT
SURANCE FRAUD

Violent crimes by Zipcode

Sheet 1



Number of Records
Abc 1
Abc 1,000
Abc 2,000
Abc 3,000
Abc 4,033

Null
89019
90000
90001
90002
90003
90004
90005
90006
90007
90008
90010
90011
90012
90013
90015
90016
90017
90018
90020
90021
90022
90023
90024
90025
90026
90027
90028
90029
90031
90032
90033
90034
90036
90037
90038
90039
90040
90042
90043
90044
90045

Map based on Longitude (generated) and Latitude (generated). Color shows details about ZIP. Size shows sum of Number of Records. Details are shown for Station Name.

Most violent crimes are

- Criminal Homicide
- Forcible Rape
- Robbery
- Aggravated Assault
- Burglary
- Larceny Theft
- Grand Theft Auto
- Arson

Conclusions By each Plot

Daily : Every day on average around 100 to150 crimes happen in LA.

Hourly : By observing Hourly plot, Peak is around 12 pm. Afternoon, there are more number crimes compared to early hours of the day

Monthly : Monthly plot shows crime rate higher in August

Yearly : There is reduction in crimes in 2013 relative to rest of the years.

Area : Southwest area, where most criminal activity happening

Crime Type : Among all crime types Traffic related issues are higher

We see that Southwest area of L.A. has the highest crime in terms of the number of crimes committed. However, a bulk of the crimes committed are minor offences. Further analysis is required to draw final conclusions on which area is the more violence affected area.