HOUSING PRICE DATA ANALYSIS

MSBA 305 : Business Intelligence and Decision Support

Instructor : Mayank Johri

Group : Anuradha Patel, Bhavin Bhanushali, Navjot Kaur, Veena Chintala

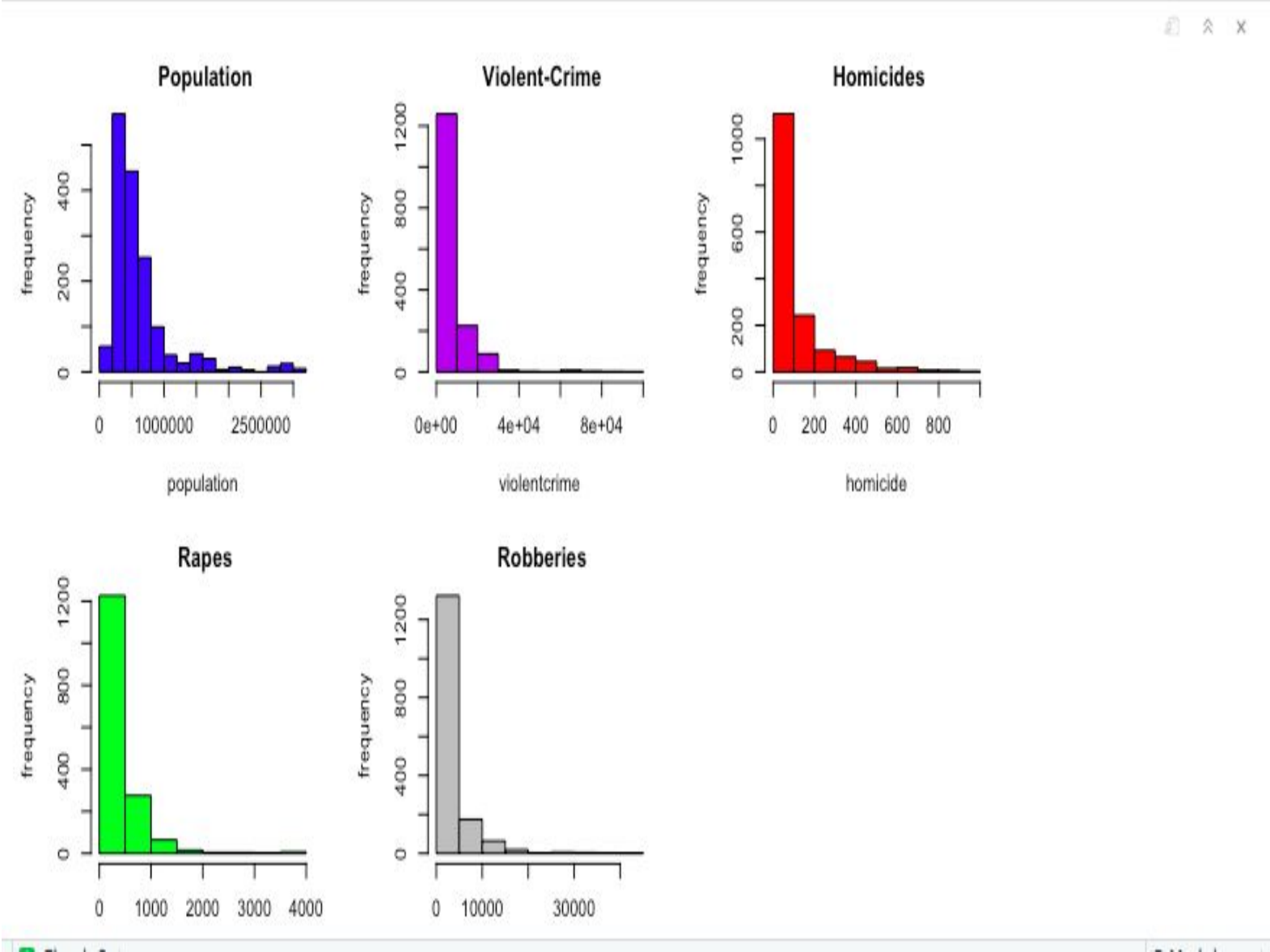
OVERVIEW & INTRODUCTION

- Background
- Data Collection
- Analysis
- Visualization
- Conclusion

DESCRIPTIVE ANALYSIS

The histogram shows the right-skewness.

This signifies that mean is greater than median.



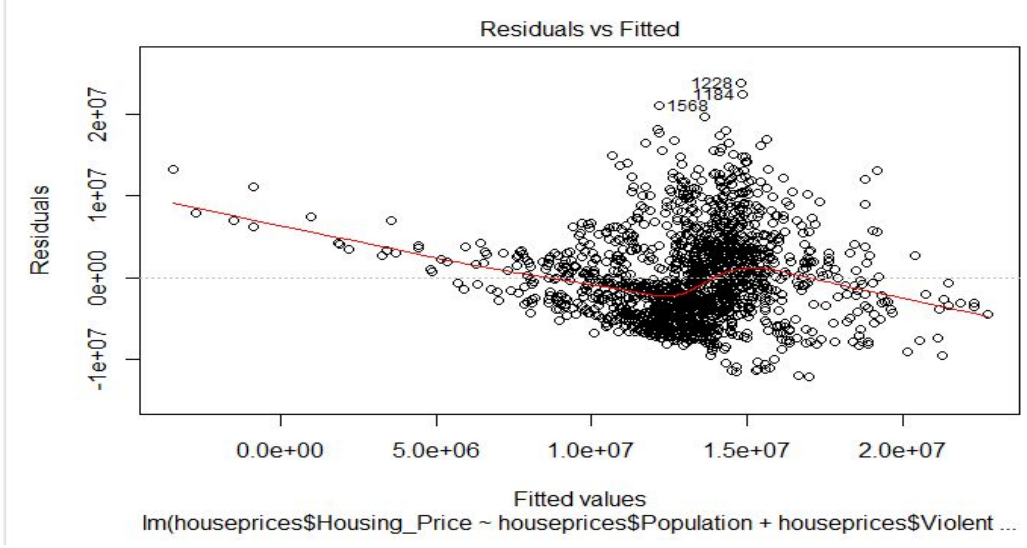
REGRESSION ANALYSIS

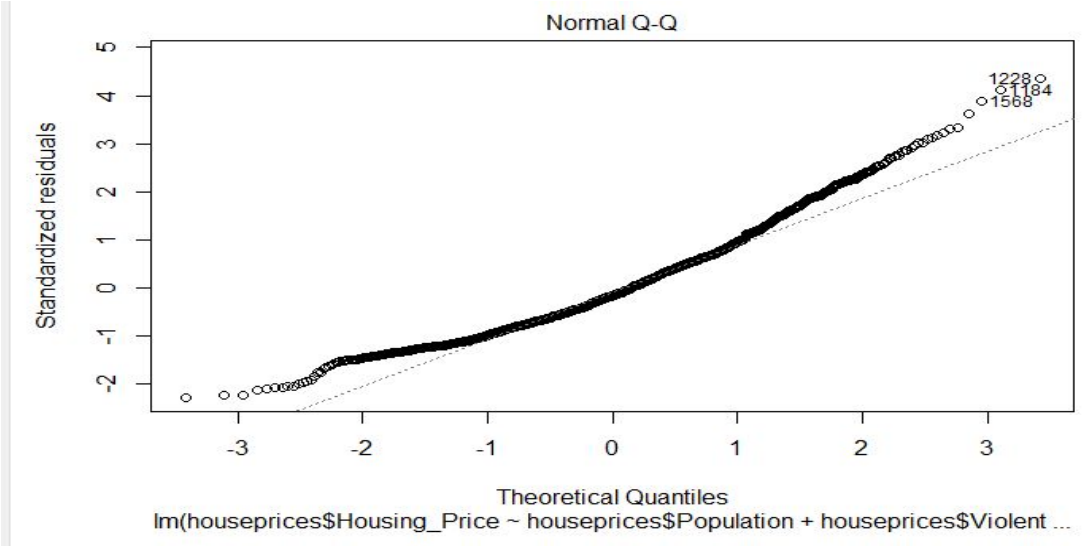
- We have more than one independent variables hence we performed multiple linear regression.
- Model 1 is with all independent variables and found that homicides do not have significant effect on housing prices.
- Model2 is without homicides.
- 0.0000001.21+ 7.0707* Population + 0.06882*Violent Crimes - 0.008359*Rapes - 0.00155*Robberies.

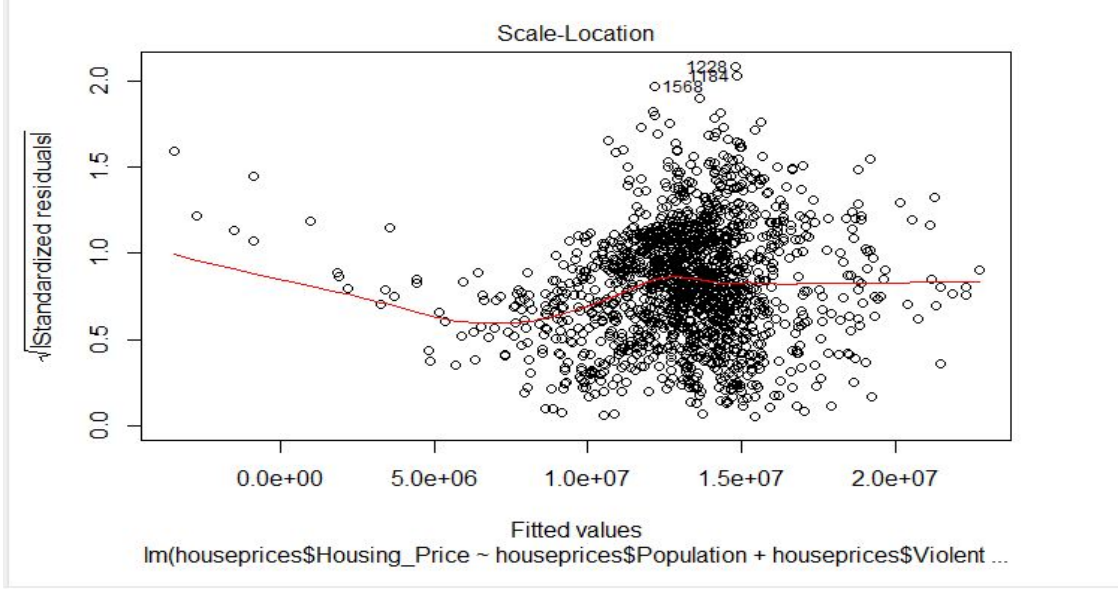
```
Call:
lm(formula = houseprices$Housing_Price ~ houseprices$Population +
    houseprices$Violent.Crimes + houseprices$Homicides + houseprices$Rapes +
    houseprices$Robberies)

Residuals:
      Min        1Q    Median        3Q       Max
-11834051  -4034790   -850761   3161042  23784768

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               1.209e+07  2.317e+05  52.203   <2e-16 ***
houseprices$Population    7.278e+00  5.143e-01  14.151   <2e-16 ***
houseprices$Violent.Crimes 6.845e+02  7.214e+01   9.487   <2e-16 ***
houseprices$Homicides     -4.024e+03  2.503e+03  -1.608    0.108
houseprices$Rapes         -8.408e+03  8.536e+02  -9.850   <2e-16 ***
houseprices$Robberies     -1.434e+03  1.642e+02  -8.730   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5461000 on 1592 degrees of freedom
Multiple R-squared:  0.1895,    Adjusted R-squared:  0.187
F-statistic: 74.45 on 5 and 1592 DF,  p-value: < 2.2e-16
```

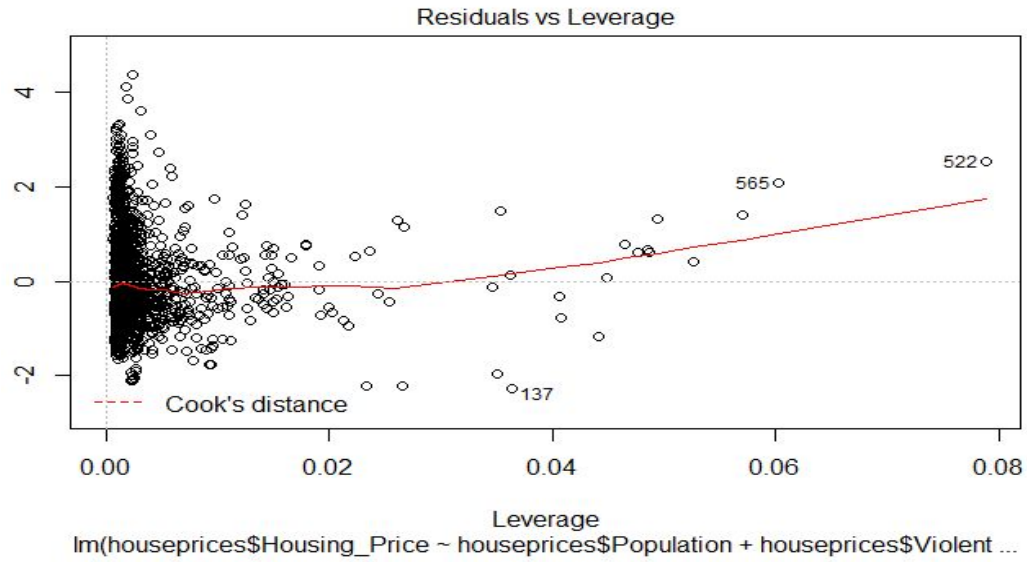
```
Call:
lm(formula = houseprices$Housing_Price ~ houseprices$Population +
    houseprices$Violent.Crimes + houseprices$Rapes + houseprices$Robberies)

Residuals:
      Min        1Q    Median        3Q       Max
-12174921  -4023619   -839569   3154800  23807760

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               1.211e+07  2.315e+05  52.304   <2e-16 ***
houseprices$Population    7.070e+00  4.981e-01  14.196   <2e-16 ***
houseprices$Violent.Crimes 6.882e+02  7.214e+01   9.540   <2e-16 ***
houseprices$Rapes         -8.359e+03  8.535e+02  -9.795   <2e-16 ***
houseprices$Robberies     -1.550e+03  1.476e+02 -10.498   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5463000 on 1593 degrees of freedom
Multiple R-squared:  0.1882,    Adjusted R-squared:  0.1862
F-statistic: 92.33 on 4 and 1593 DF,  p-value: < 2.2e-16
```

REGRESSION DIAGNOSTIC









CORRELATION

- Population has a weak positive correlation with the housing prices
- Violent crimes, rapes and robberies have weak negative correlation with the housing prices

	houseprices.Housing_Price	houseprices.Population	houseprices.Violent.Crimes	houseprices.Rapes	houseprices.Robberies
houseprices.Housing_Price	1.00000000	0.06873349	-0.08678161	-0.1541507	-0.1387294
houseprices.Population	0.06873349	1.00000000	0.81157151	0.8076039	0.8077311
houseprices.Violent.Crimes	-0.08678161	0.81157151	1.00000000	0.9115993	0.9744025
houseprices.Rapes	-0.15415068	0.80760386	0.91159935	1.0000000	0.9050507
houseprices.Robberies	-0.13872936	0.80773115	0.97440250	0.9050507	1.0000000

ANOVA TEST

Population, Violent Crimes, Rapes and Robberies have significant effect on Housing prices

```
                      Df    Sum Sq   Mean Sq F value  Pr(>F)
houseprices$Population   1 2.767e+14 2.767e+14   7.576 0.00598 **
Residuals             1596 5.830e+16 3.653e+13
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
                      Df    Sum Sq   Mean Sq F value   Pr(>F)
houseprices$Rapes     1 1.392e+15 1.392e+15   38.85 5.85e-10 ***
Residuals          1596 5.718e+16 3.583e+13
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
                         Df    Sum Sq   Mean Sq F value   Pr(>F)
houseprices$Violent.Crimes   1 4.411e+14 4.411e+14   12.11 0.000515 ***
Residuals                 1596 5.813e+16 3.642e+13
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
                       Df    Sum Sq   Mean Sq F value   Pr(>F)
houseprices$Robberies   1 1.127e+15 1.127e+15   31.32 2.57e-08 ***
Residuals            1596 5.745e+16 3.599e+13
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

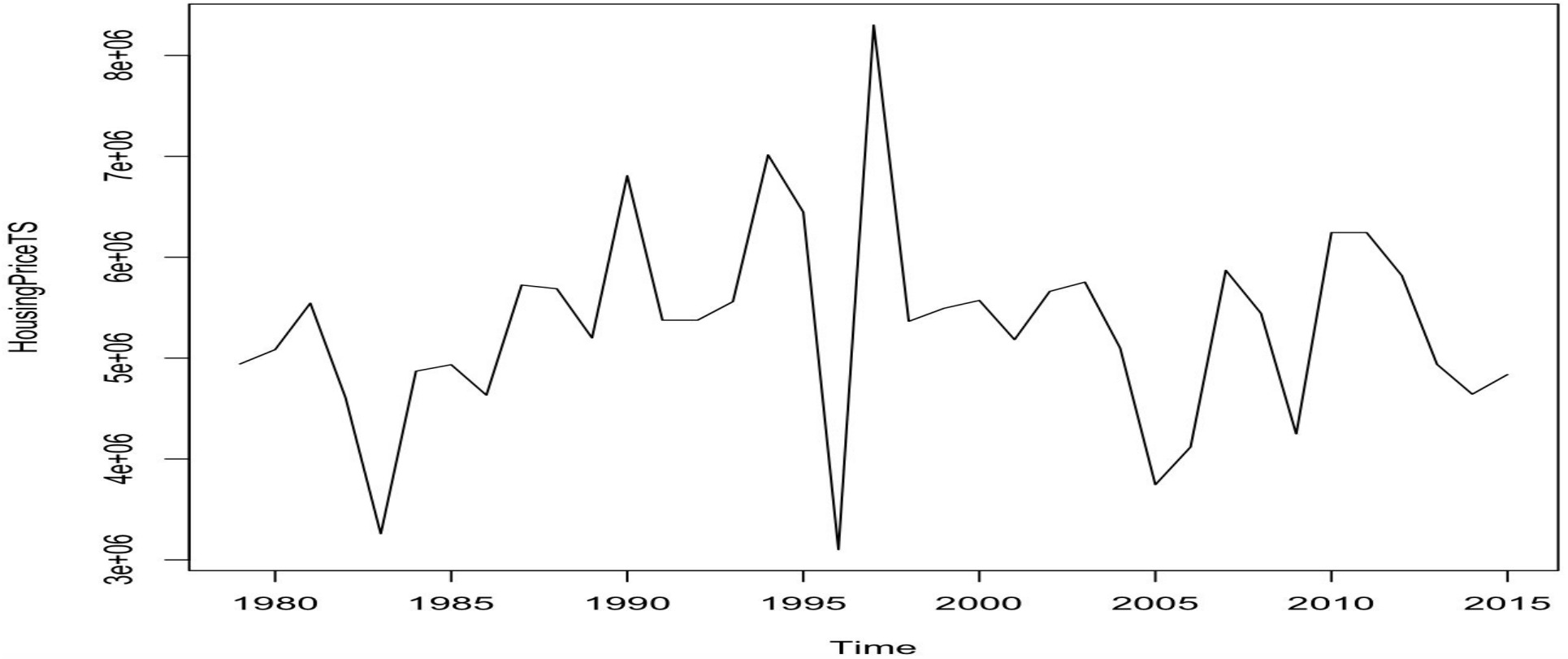
ANOVA TEST

Housing prices for different cities is significantly different.

```
                          Df    Sum Sq   Mean Sq F value   Pr(>F)
houseprices$City..State   43 3.438e+15 7.996e+13   2.254 7.98e-06 ***
Residuals               1554 5.514e+16 3.548e+13
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

TIME SERIES ANALYSIS

- Time series analysis gives the trend of housing price over these years.
- The graph shows that Housing Price has increased and decreased over the years, and it was highest in the year 1997.

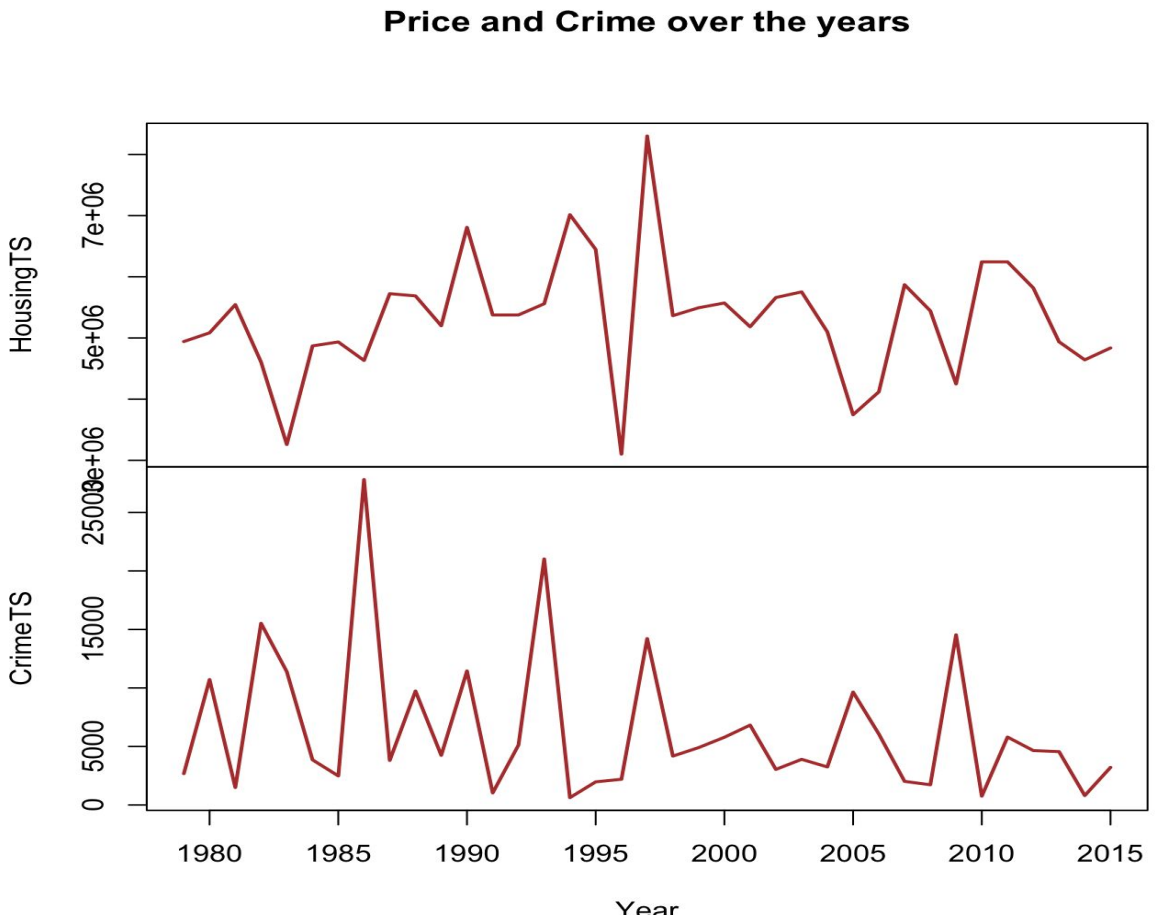


RELATION OF CRIME AND PRICE WITH YEAR

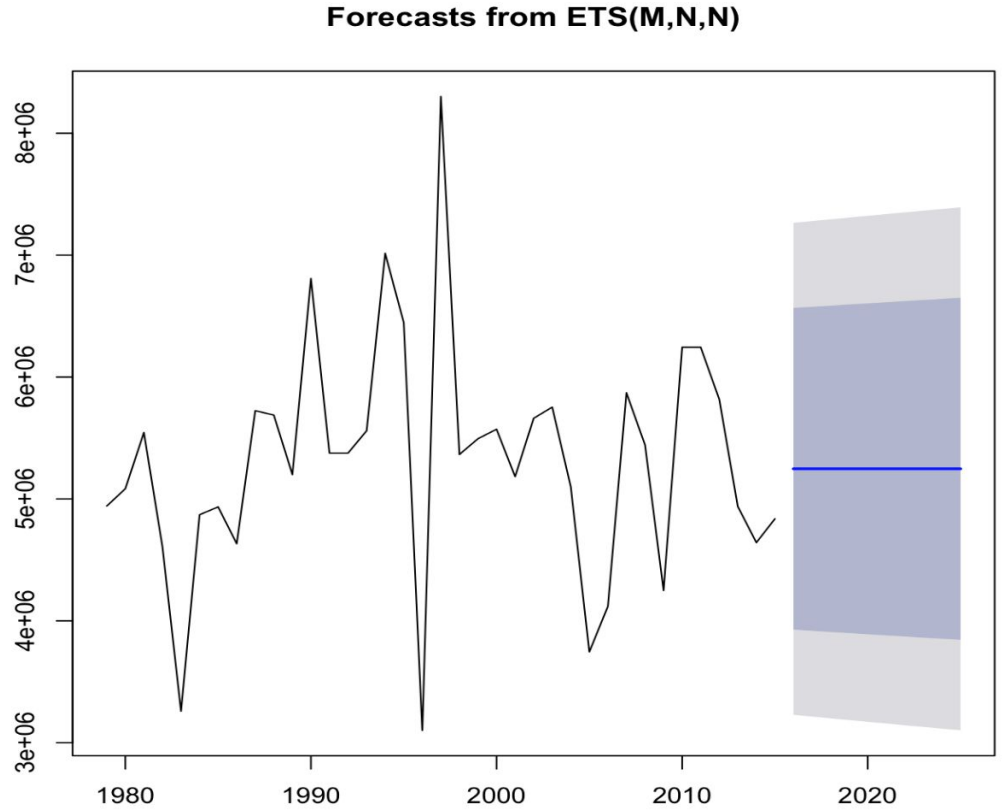
> HousingPriceTS <- ts(HousingData$Violent.Crimes, start=c(1979,1),end = c(2015))
> HousingRapeTS <- ts(HousingData$Rapes, start=c(1979,1),end = c(2015))

Individual Graph: plot.ts(HousingPriceTS) plot.ts(HousingRapeTS)

Combined: > plot.ts(cbind(HousingRapeTS, HousingPriceTS))



It is forecasted that the probability of projected value of house price will be more in darker area



PRESCRIPTIVE ANALYSIS

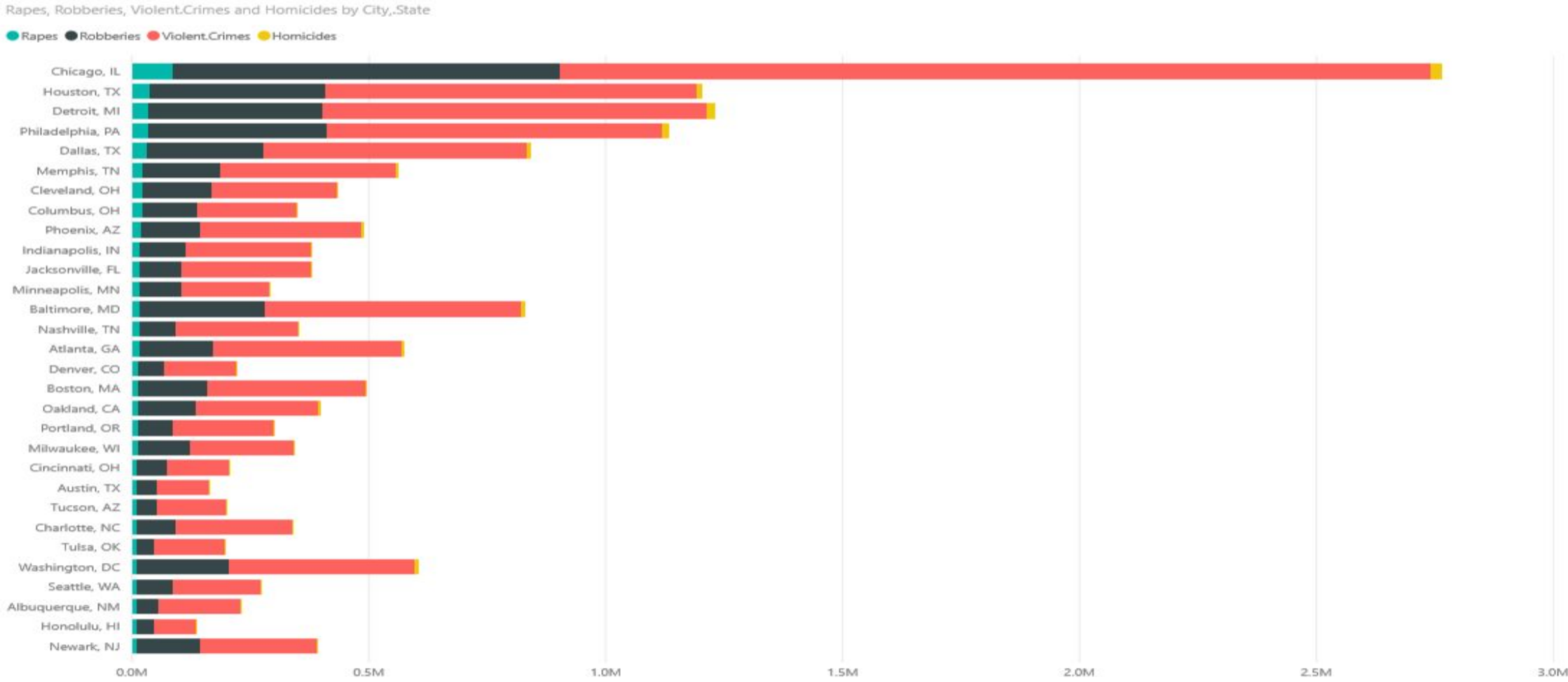
- Prescriptive analysis is the field of business analytics which is mainly to find the best course of action after evaluating any situation.
- After doing our descriptive and predictive analysis on housing price dataset, housing price across the region differs according to the rate of different types of crime occurrence, we have few recommendations to be considered.

1. The real estate should develop affordable homes, so that common man can have a quality of life, which would eventually decrease in the number of crimes.
2. The dealers should have detailed analysis of a particular region's housing price history before investing in. The customers and company, both should take decision wisely after looking through the analysis.

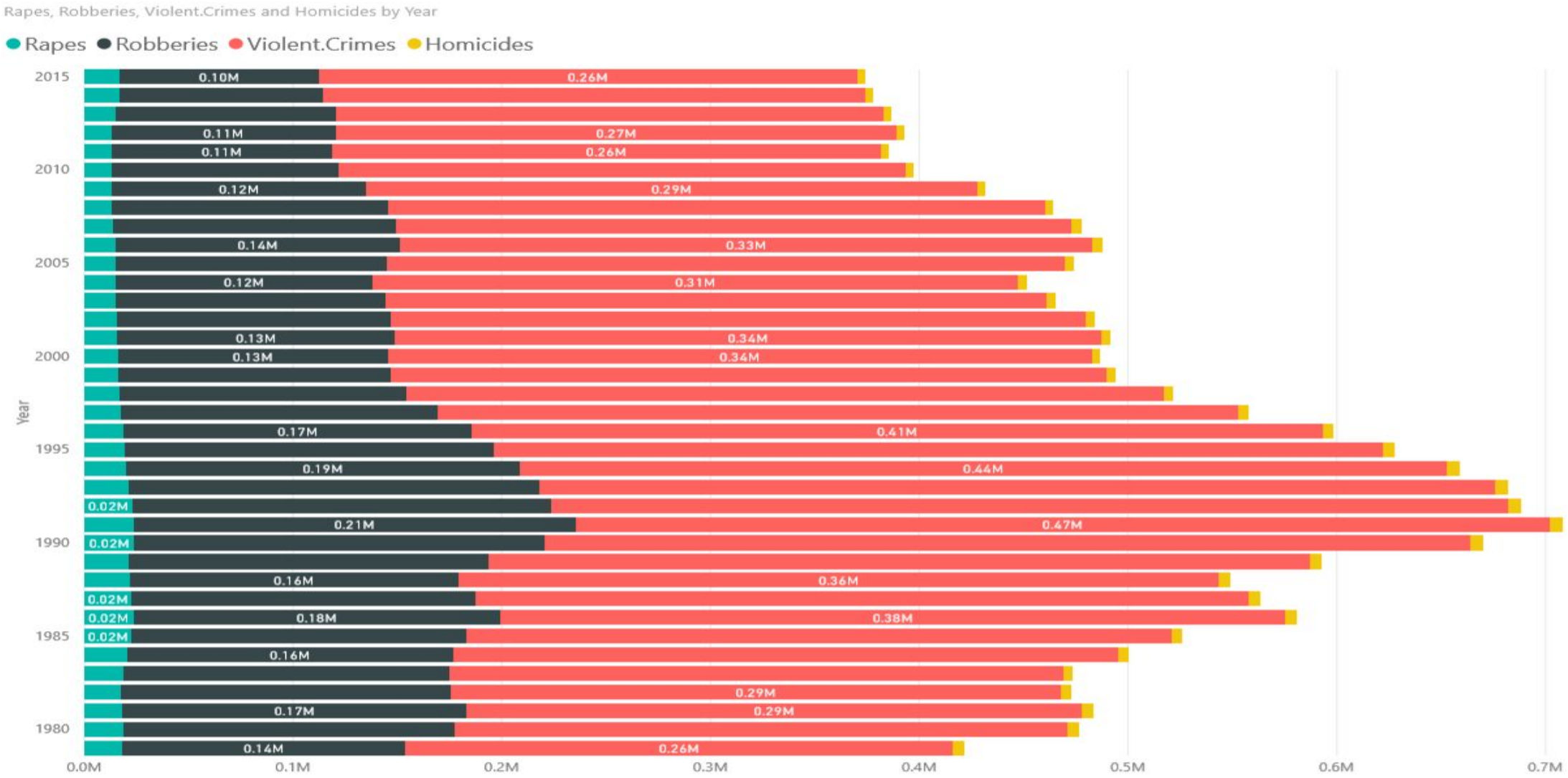
- Prescriptive analytics leads to optimization in production, scheduling and inventory in the supply chain to make sure that customers are at the right path.

VISUALIZATIONS

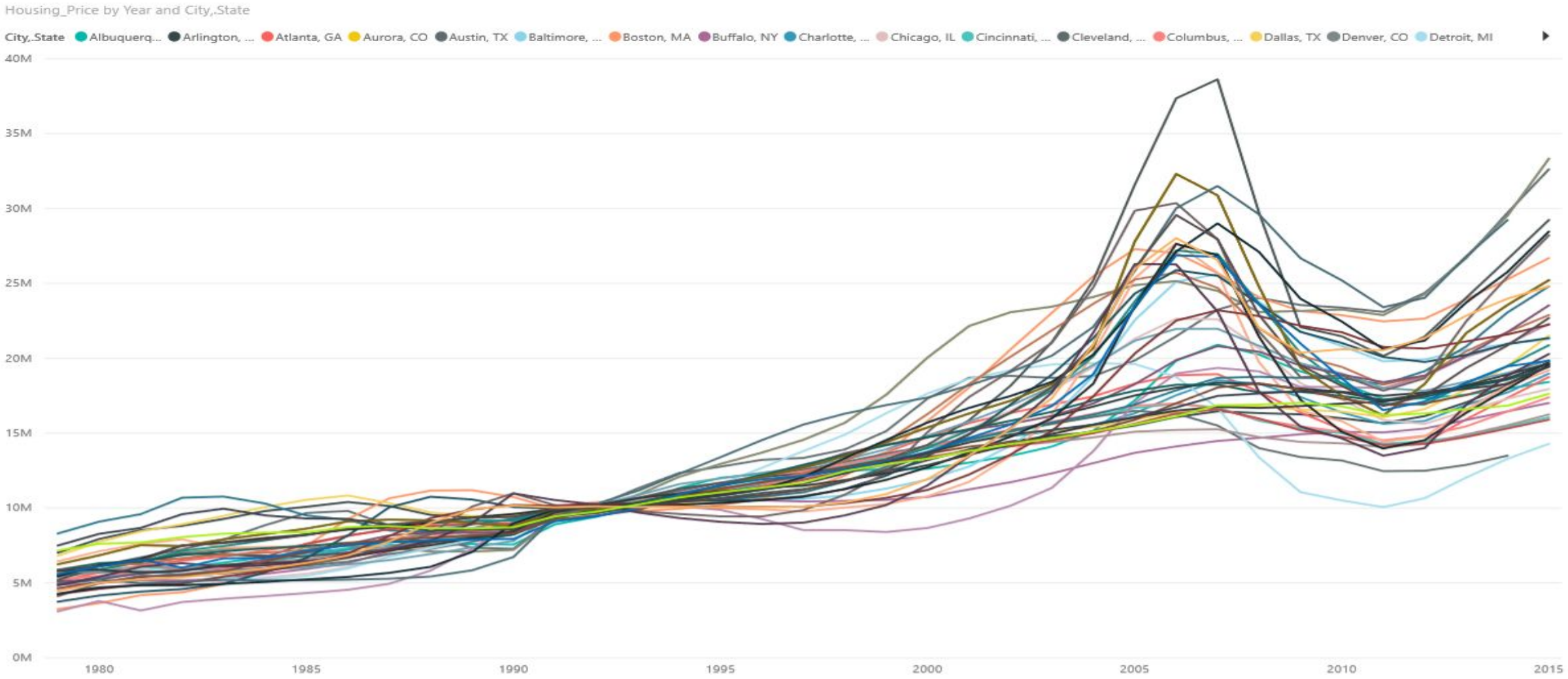
- Below Graph showing the Rape, Robberies, Violent Crimes and Homicides data by City.



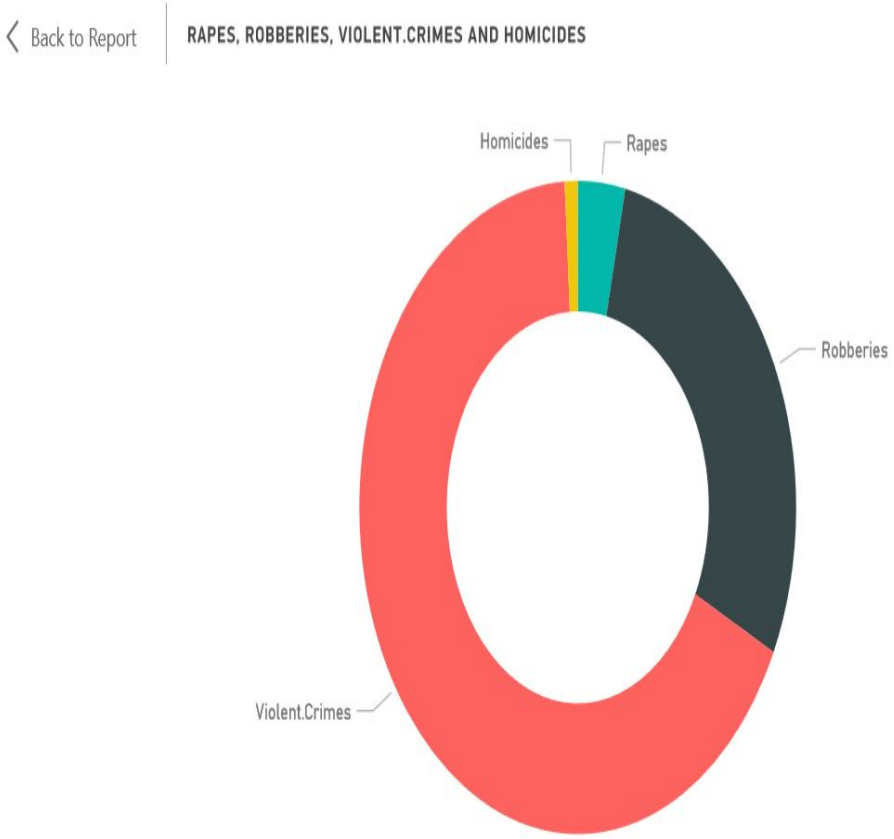
Below Graph showing the Rape, Robberies, Violent Crimes and Homicides data by Year.



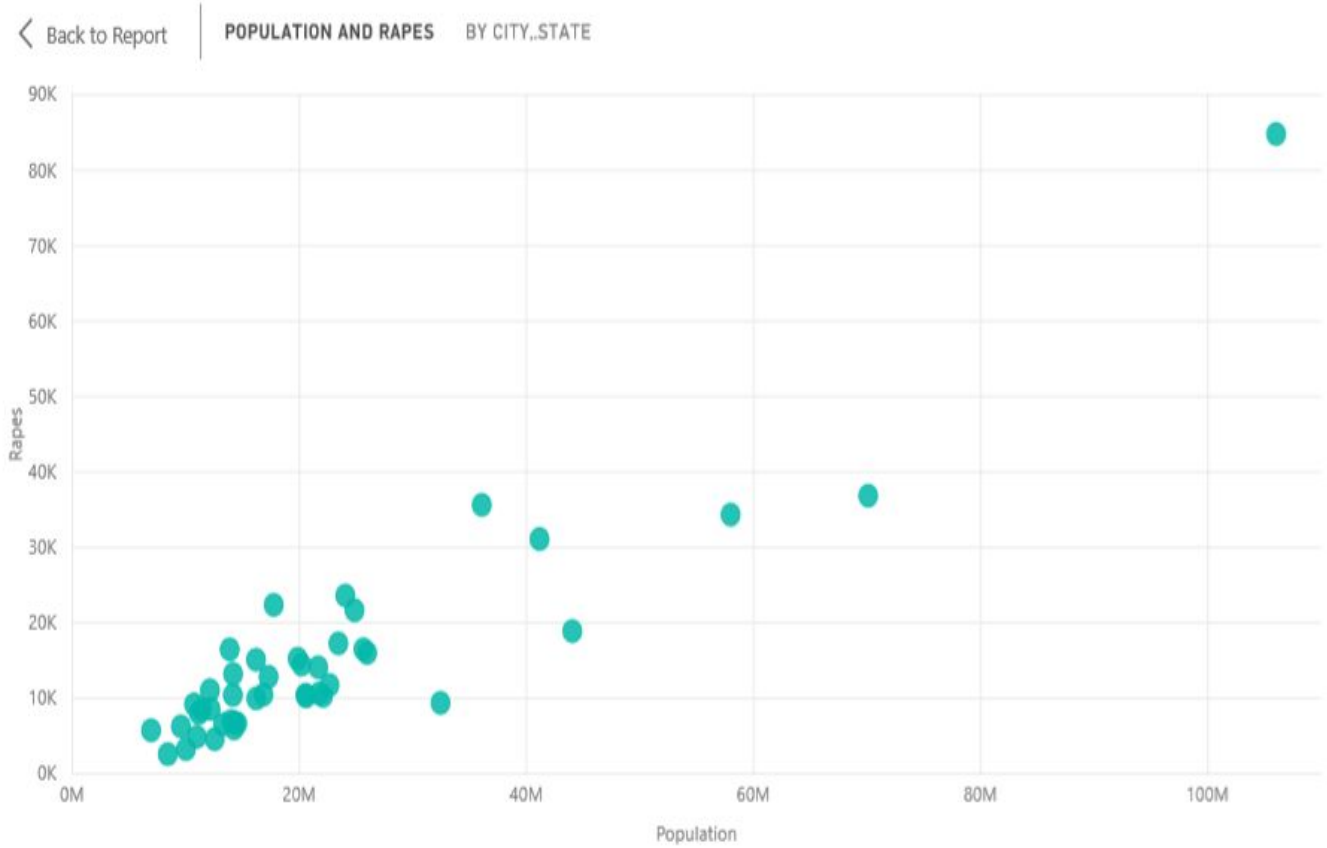
Below Graph showing the changes in House price by Year and by City.



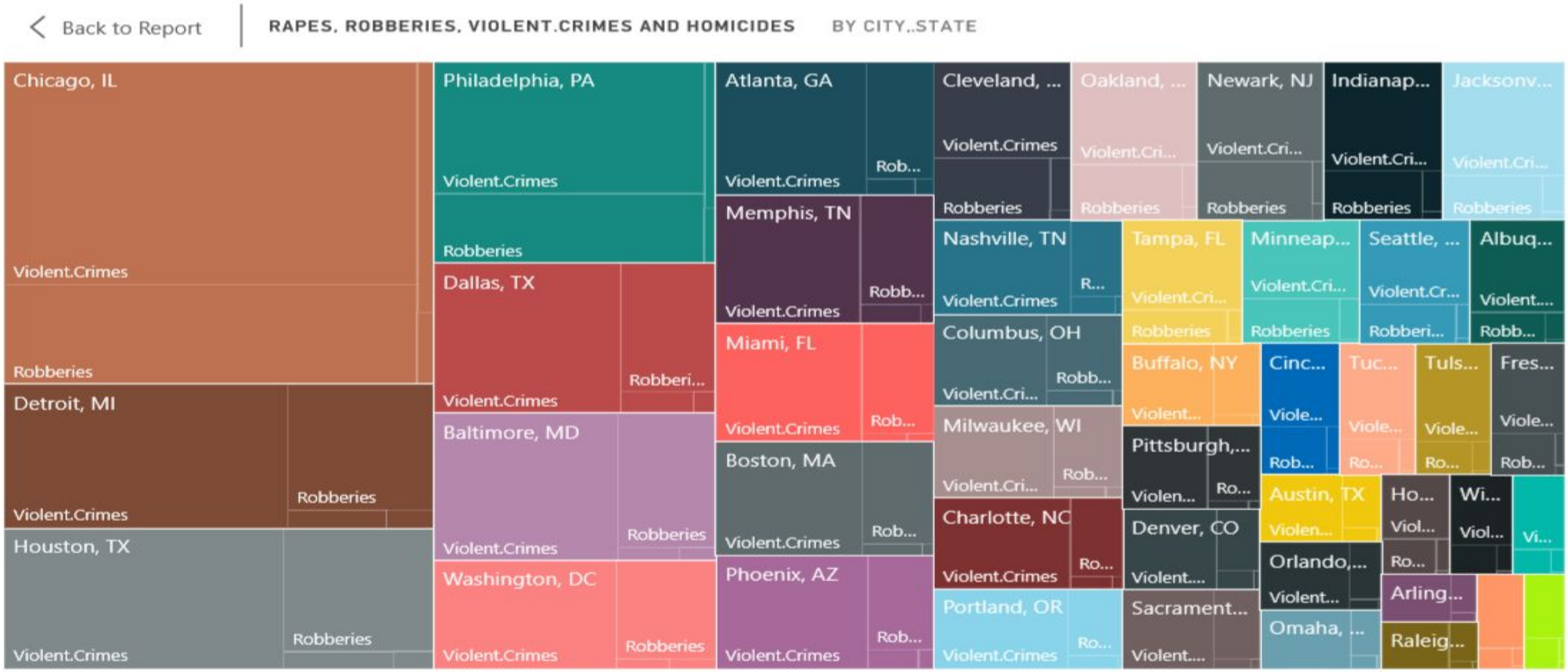
Below Graph showing the overall crime rate by Homicides, Rapes, Robberies and Violent Crimes.



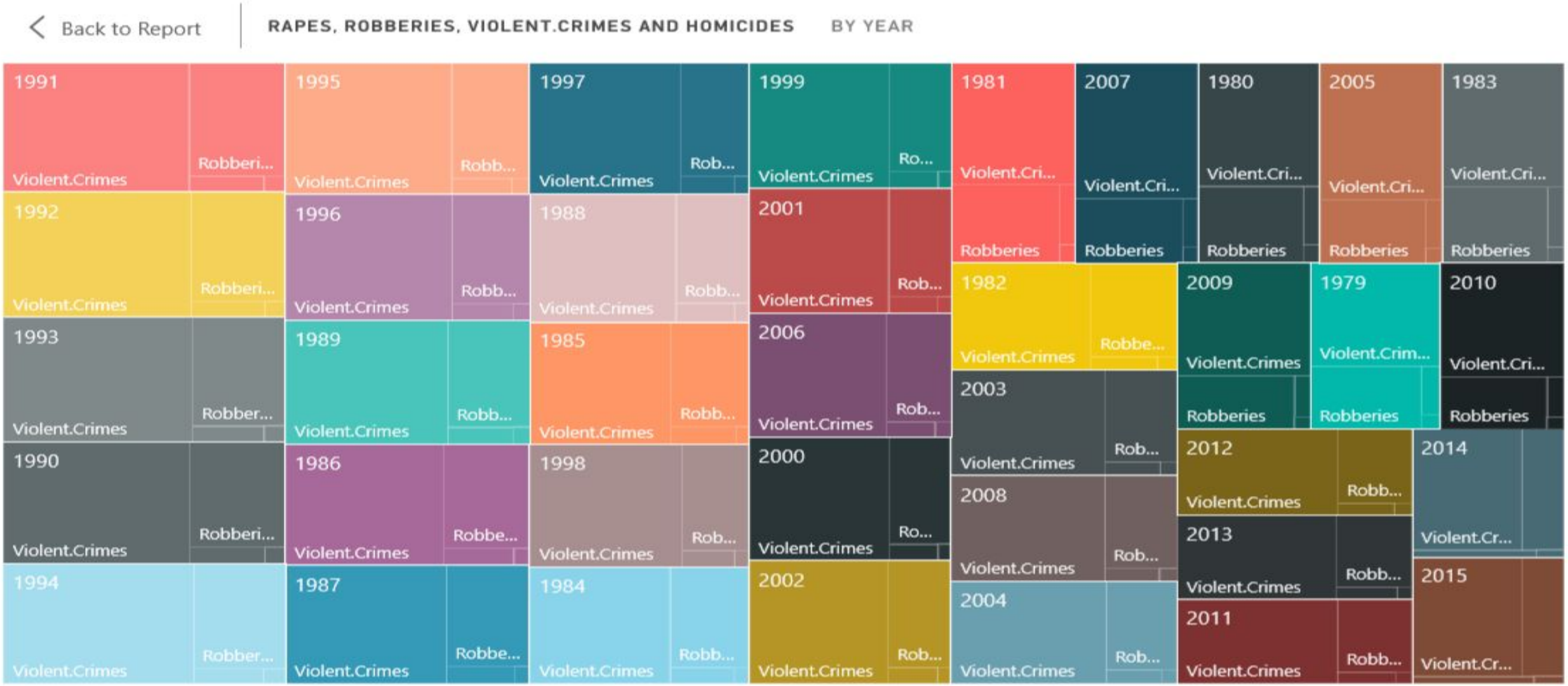
Below graph showing the relationship between Population and Rapes crime by City.



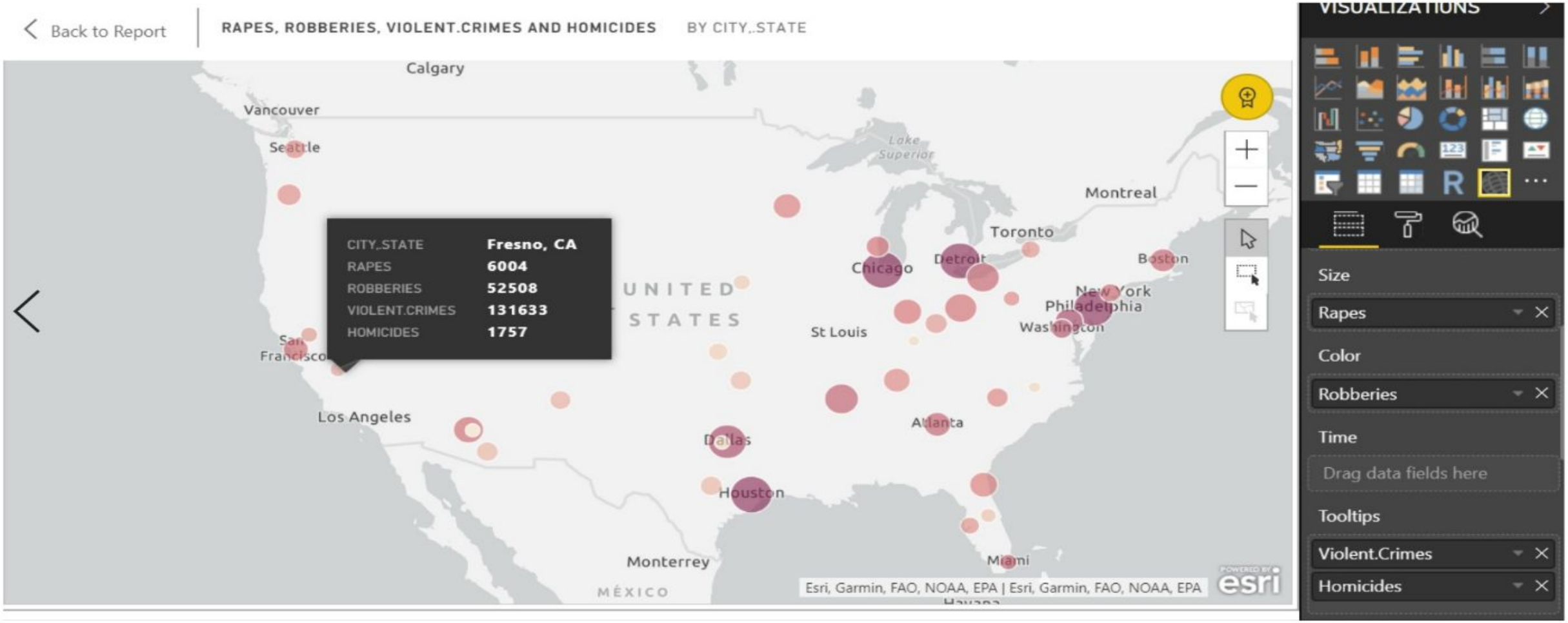
Tree map of Rapes, Robberies, Violent crimes and Homicides by city.



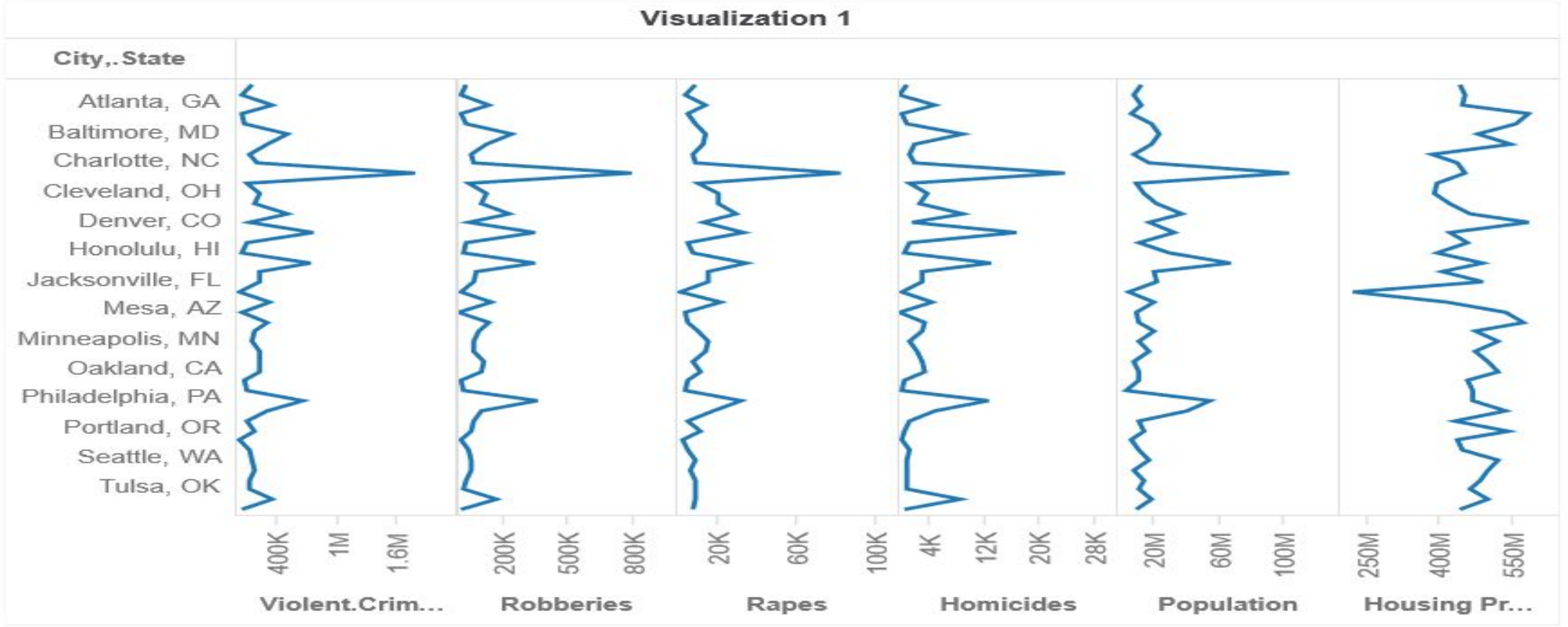
Tree map of Rapes, Robberies, Violent crimes and Homicides by Year.



Map graph showing the cities impacted by different crimes with Size of circle indicating Rapes and color indicating Robberies.



Below graph showing the chart by city indicating violent crime, robberies, rapes, Homicides, Population and Housing price.



CONCLUSION

- From the analysis we found that the housing prices for different cities are significantly different.

- Population has weak positive correlation on housing prices and violent crimes, rapes and robberies have weak negative correlation with the housing prices.

- As per the analysis, the Chicago has the highest rate of crime, and Detroit being the second. Also, violent crime and robberies have weak positive correlation with the housing price.

THANK YOU