ANALYSIS ON CRIME IN NYC PRECINCTS

CONTRIBUTORS: Neil Verosh D'Souza, Nikhil Kishore, Priya Singh Khokher, Sachin Verma

ABSTRACT

Crime is a major issue in most cities as it affects a primary quality of life factor in urban areas which is safety. Crime seems very random and hard to predict; but are there factors that affect the crime in a city? This project attempts to looks at different demographic factors to see which ones, if any, affect the total crime. NYPD crime statistics is aggregated based on precincts, while demographic data is collected on different geographical boundaries. To conduct the analysis we have spatially joined US Census data to the precincts shape file. This would make our findings easier to present to the NYPD with the appropriate context.

HYPOTHESIS

H_0: Total crime in a precinct is not affected by any demographic factors.

H_A: Total crime in a precinct is affected by factors like total population, number of educational establishments, average income, race, gender and age.

DATA

Data	Definition	Granularity
NYC Precinct[1]	NYC area divided according to Police Stations Jurisdiction	ShapeFile
Race, Income[2]	Demographic data for NYC	Census Tract

Age, Gender[3]	Demographic data	By Zip Code
Crime Statistics[4]	Data about Crime in NYC	By Precinct
Schools[5]	Point Locations of Schools in NYC	By Precinct

Data Preparation:

Crime data from NYPD is available only at the precinct level, whereas US Census does not provide data at precinct level. This made it difficult to compare the two datasets.

To minimize the errors in analysis following was done:

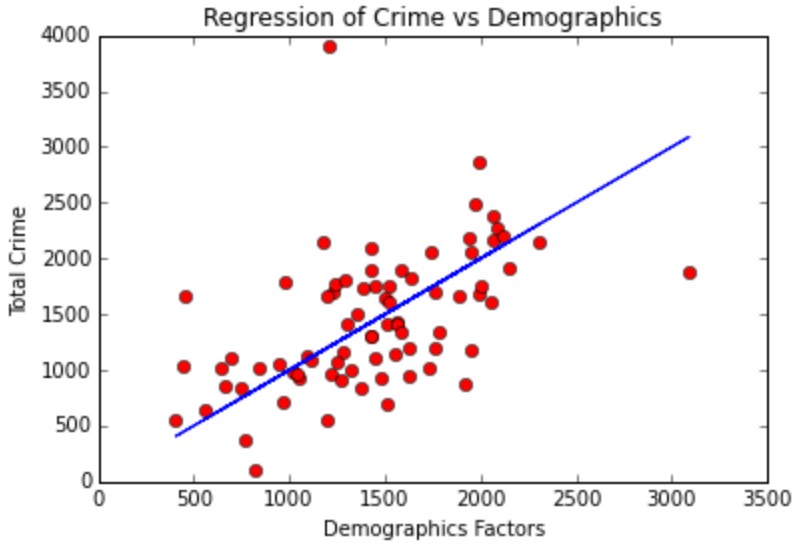
- For demographic data available at census tract, a spatial join of NYC Census tract shape file with NYC precinct shape file was performed. The join aggregated the statistics at precinct level.
- Joining demographic data available at Zip code level was trickier and error prone as zip codes cover larger areas. A given precinct can overlap multiple Zip codes. To circumvent this, centroid (i.e. the intersection point of latitude and longitude of the Zip code area) was calculated for each of the Zip codes and then a spatial join performed using this centroid provided the desired aggregation of data set at precinct level (with least error).

Zip codes with their centroids

Finally we converted the resulted output dbf file to csv using python and processed all the data into single csv for making our final data.

ANALYSIS

The first step was a **multivariate linear regression** to see if crime was influenced by the various factors. The regressors dealing with population were normalized using the total population to exclude some collinearity.



OLS Regression Summary

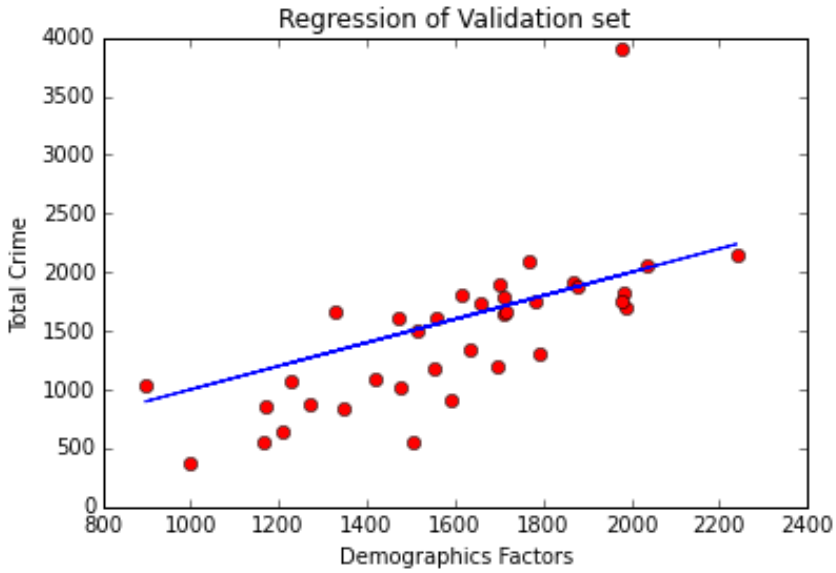
Dep. Variable:	Y	**R-squared:**	0.510
Model:	OLS	**Adj. R-squared:**	0.204
Method:	Least Squares	**F-statistic:**	1.666
Date:	Sun, 15 Nov 2015	**Prob (F-statistic):**	0.128
Time:	17:32:25	**Log-Likelihood:**	-297.16
No. Observations:	40	**AIC:**	626.3
Df Residuals:	24	**BIC:**	653.3
Df Model:	15		
Covariance Type:	nonrobust		

Names	X	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Total Population	**X1**	0.0010	0.005	0.195	0.847	-0.010 0.012
Number of Men	**X2**	1.309e+04	1.1e+04	1.190	0.246	-9617.892 3.58e+04
Number of Women	**X3**	1.545e+04	1.09e+04	1.413	0.170	-7114.622 3.8e+04
Whites	**X4**	-1.454e+04	9921.871	-1.465	0.156	-3.5e+04 5940.045
Blacks	**X5**	-1.284e+04	9870.526	-1.301	0.206	-3.32e+04 7527.009
American Indians	**X6**	8858.1465	4.63e+04	0.191	0.850	-8.67e+04 1.04e+05
Asians	**X7**	-1.303e+04	1.05e+04	-1.246	0.225	-3.46e+04 8557.376
Native Hawaiians	**X8**	-4480.4133	1e+05	-0.045	0.965	-2.11e+05 2.02e+05
Others	**X9**	-1.045e+04	9680.961	-1.080	0.291	-3.04e+04 9526.862
Men below 25	**X10**	-3.739e+05	1.58e+05	-2.364	0.026	-7e+05 -4.75e+04
Men between 25 -70	**X11**	3.475e+05	1.5e+05	2.320	0.029	3.83e+04 6.57e+05
Women below 25	**X12**	3.664e+05	1.82e+05	2.016	0.055	-8617.120 7.41e+05
Women between 25-70	**X13**	-3.398e+05	1.67e+05	-2.039	0.053	-6.84e+05 4216.327
Mean Income	**X14**	-0.0011	0.001	-1.081	0.291	-0.003 0.001
Median Income	**X15**	0.0015	0.001	1.904	0.069	-0.000 0.003
Number of Schools	**X16**	-19.6396	25.968	-0.756	0.457	-73.235 33.956

Omnibus:	0.193	**Durbin-Watson:**	2.019
Prob(Omnibus):	0.908	**Jarque-Bera (JB):**	0.214
Skew:	0.145	**Prob(JB):**	0.899

Kurtosis:	2.790	**Cond. No.**	8.89e+09

The regression did show a correlation between total crime and the regressors indicating that factors like age and median income have an impact (low p-values) but it was not possible to clearly interpret which regressors have a greater impact due to collinearity. The model did not perform so well on the validation set though slightly better than the training set (R-squared of 0.573).



To check for this, **feature selection (forward-stepwise)** was carried out on the regressors and it was found that precincts that have a greater population overall as well as a greater number of women, white people, Asians and Native Hawaiians have more crime. The most probable reason for number of white people factoring in is that they comprise a majority of the total population in most precincts. It needs to be emphasized that the people falling under these categories need not be perpetrators of crime but also (and most probably) victims of crime.

Output of Forward Stepwise Feature Selection

Dep. Variable:	Y	**R-squared:**	0.887
Model:	OLS	**Adj. R-squared:**	0.871
Method:	Least Squares	**F-statistic:**	55.01
Date:	Sun, 15 Nov 2015	**Prob (F-statistic):**	1.36e-15
Time:	17:25:06	**Log-Likelihood:**	-307.05
No. Observations:	40	**AIC:**	624.1
Df Residuals:	35	**BIC:**	632.5
Df Model:	5		
Covariance Type:	nonrobust		

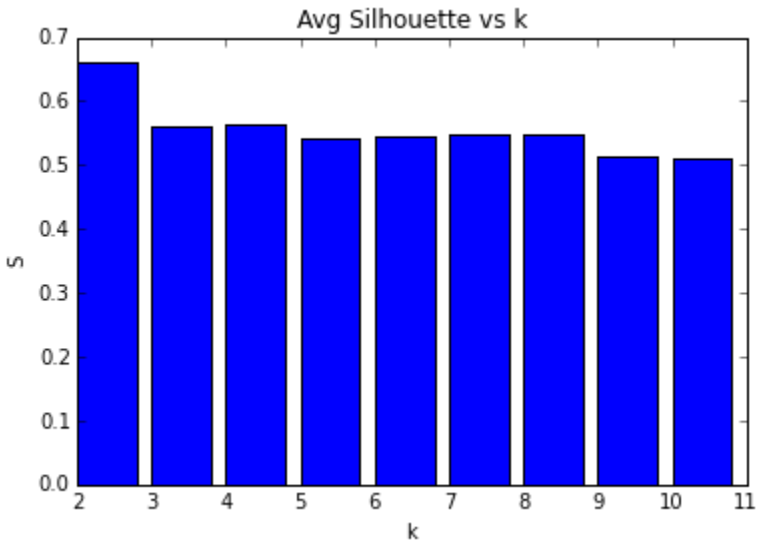
		coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
Total Population	**X1**	0.0036	0.002	2.360	0.024	0.001 0.007
Number of Women	**X3**	2354.8238	684.814	3.439	0.002	964.577 3745.071
Whites	**X4**	-797.0616	714.855	-1.115	0.272	-2248.295 654.172
Asians	**X7**	555.7551	1398.018	0.398	0.693	-2282.372 3393.882
Native Hawaiians	**X8**	-3.637e+04	8.41e+04	-0.432	0.668	-2.07e+05 1.34e+05

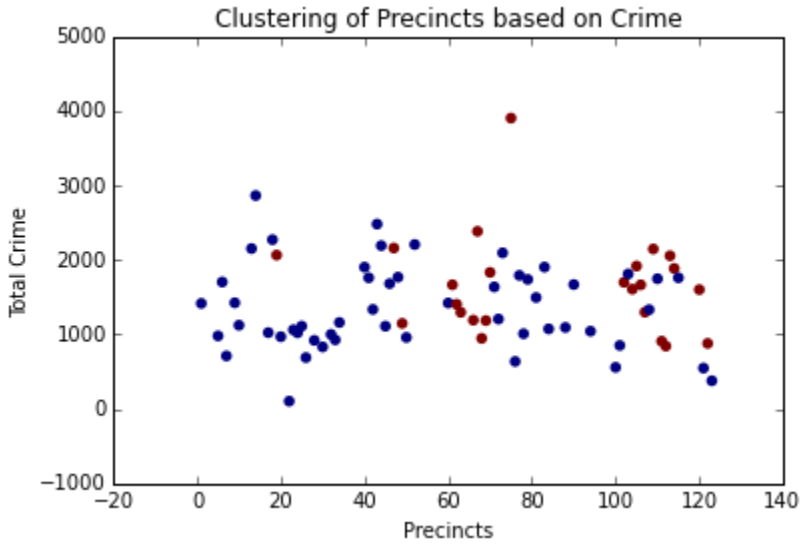
Omnibus:	6.051	**Durbin-Watson:**	1.757
Prob(Omnibus):	0.049	**Jarque-Bera (JB):**	4.910
Skew:	0.618	**Prob(JB):**	0.0859

Kurtosis:	4.191	**Cond. No.**	1.41e+08

Principal component analysis was performed but no valuable interpretation could be made. It only indicated that while individually factors may not have a major impact, as a combination, all the factors definitely influence crime.

An attempt was made to cluster precincts based on total crime in each precinct to see if any precincts display similar characteristics.





The **K-means clustering** suggests that two clusters is the optimum number based on silhouette values but visually it was hard to see a clear distinction between the two clusters which means that no group of precincts displays characteristics that are distinctly similar.

Individual Contribution

For this project, I contributed with the IPython notebook and the analysis of the data. I, also, did most of the written work and provided the conceptual ideas for the project and which data to use.

Conclusion

Through the various analysis performed it can be concluded that crime is influenced by demographic factors and mainly by the total population in a precinct. It is likely that precincts with a greater population of minorities will have more crime since opportunity for crime may be greater. Despite the different locations of precincts, there seems to be no significantly similar patterns or characteristics among them to form separate groups/clusters.

REFERENCES

[1] "NYC Precinct shapefile[1]",(2010), available at
http://www.nyc.gov/html/dcp/download/bytes/nypp_15c.zip

[2] "Race, Income[2]",Census Data (2010), available at
http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml

[3] "Age, Gender[3]", Census Data(2010), available at
http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml

[4] "Crime Statistics[4]" NYPD (2010), available at
http://www.nyc.gov/html/nypd/html/analysis_and_planning/historical_nyc_crime_data.shtml

[5] "Number of Schools[5]" New York Op10ern Data (2010), available at
https://data.cityofnewyork.us/Education/School-Point-Locations/jfju-ynrr