Data Analytics for Information Professionals

Topic

Statistical Data Analysis of Crime Rates across U.S. Communities

Instructor*:* Yla Tausczik

Submitted by-

[Rajat Aghi, Rahul Bahadur, Akash Udani, Sheryl Mathias]

Problem Domain

The problem domain we aim to look at with this project is that of crime across various communities [1] in the United States. This is an important problem because ensuring the security of persons and property within a community is essential for the well-being of the people living in the community. Having low crime rates is also directly related to the prosperity of the state, making this one of the top priorities across all levels of government.

Our Motivation

The analysis of this data set can be used to identify trends in both high and low crime communities and discern the possible reasons for such trends. This analysis can help in identifying the factors which deter crime in communities and consequently, steps can be taken in relatively high crime communities to magnify these factors in whatever way possible.

Research Question

How do the socio-economic factors describing a community affect the total crime in the community?

The Data Set

Our data set[2] describes the problem of crime within various communities across the United States. The data set that we have chosen combines socio-economic data from the census and Admin Stats Survey and crime data from the FBI Uniform Crime Report. This data set can be used to identify trends in both high and low crime communities and discern the possible reasons for such trends. Initially, our data set had 147 variables about 2215 communities. Out of these 147, 129 were independent variables and 18 were dependent variables.

Data Cleaning

1. Converted CommViolPredUnnormalizedData.txt text file to CommunityandCrimeData Comma Separated Value file.

2. Out of the 18 potential goal attributes we retained only 2 of them i.e. ViolentCrimesPerPop (total number of violent crimes per 100K popuation) and nonViolPerPop (total number of non-violent crimes per 100K popuation) as they were representative of all the 8 crimes.

3. Out of 125 predictive variables (Independent variables), 21 variables were omitted as they had too many null values (values with ?)

4. The 2 outcome variables (ViolentCrimesPerPop and nonViolPerPop) were combined to form another outcome variable called “TotalCrime” in Excel.

5. Missing values which were marked with ? were omitted from our final data set by running the following R command:
   CrimeDataCleaned<- na.omit(Final_Crime_Set)

write.csv(CrimeDataCleaned, file = "FinalCrimeSet.csv", row.names = FALSE)

6. Two of our independent variables i.e. NumInShelters (number of people in homeless shelters) and NumStreet (number of homeless people counted on the street) were converted to percentage by using community wise population (population for community) as all other independent variables are described in percentage format. Also percentage is more representative of a sample compared to absolute values.

7. From the above step we calculated the independent variables "PctHomShelt" and "PctOnStreet" using the following R commands:
   *CrimeDataCleaned$PctHomShelt<-(CrimeDataCleaned$NumInShelters/CrimeDataCleaned$population)*100*

   *CrimeDataCleaned$PctOnStreet<-(CrimeDataCleaned$NumStreet/CrimeDataCleaned$population)*100*

From the 18 dependent variables, we shortlisted 2 variables, **ViolentCrimesPerPop** (total number of violent crimes per 100K population) and **nonViolPerPop** (total number of non-violent crimes per 100K population) as they were representative of all the 8 crimes described by the other dependent variables.
The 2 dependent variables were then combined to form our final dependent variable, **TotalCrime** (total number of crimes per 100k population).

Descriptive Statistics for Dependent Variable

For our model, the dependent variable is Total number of crimes committed community wise per 100K population. We calculated our dependent variable by adding total number of non-violent crimes and violent crimes per 100K population community wise respectively.

Mean	Median	Standard Deviation	Minimum	Maximum
5524.762	4908.91	3229.12	124.58	30594.25

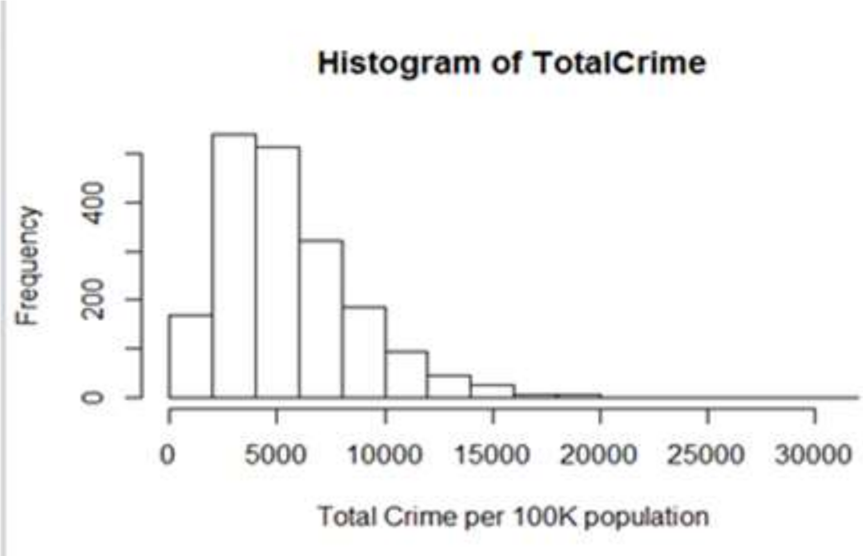


Fig. 1

The histogram has a positive skew. It gives us the frequency of the total crime per 100K population.

Descriptive Statistics for Independent Variables

Out of all the independent variables present initially in the data set, we shortlisted 15 variables which showed a significant effect on our dependent variable.

Sr. No.	Variable Name	Brief Description	Mean	Median	Std. Dev.	Min. Value	Max. Value
1	PctPopUnderPov	Percentage of people under the poverty level	11.66	9.38	8.47	0.64	48.82
2	PctNotHSGrad	Percentage of people 25 and over that are not high school graduates	22.66	21.54	11.07	2.09	73.66
3	PctHomShelt	Number of people in homeless shelters	0.06	0.00	0.13	0.00	1.95
4	PctImmig	Percentage	7.78	4.58	8.82	0.191	60.40

		of immigrants who immigrated within last 3 years					
5	MalePctDivorce	percentage of males who are divorced	9.17	9.20	2.81	2.13	19.09
6	PctOnStreet	Percentage of homeless people counted on street.	0.01	0.00	0.06	0	1.48
7	agePct65up	Percentage of population that is 65 and over in age	11.98	11.83	4.85	1.66	52.77
8	PctEmploy	Percentage of people 16 and over who are employed	61.87	62.47	8.09	24.82	84.67
9	PersPerFam	Mean number of people per family	3.13	3.10	0.25	2.29	4.64
10	PctKids2Par	Percentage of kids in family housing with two parents	71.03	72.23	11.85	26.11	92.58
11	PctWorkMomYoungKids	Percentage of moms of kids 6 and under in labor force	60.32	60.62	7.91	24.42	87.97
12	HousVacant	Number of vacant households	1733.55	584	5944.42	36	172768
13	RentMedian	rental housing - median rent	433.39	400	175.03	139	1001
14	PctBornSameState	percent of people born in the same	59.99	61.84	17.03	6.75	93.14

		state as currently living					
15	PopDens	Population density in persons per square mile	2804.22	2003.50	2945.49	10	44229.90

Table 1

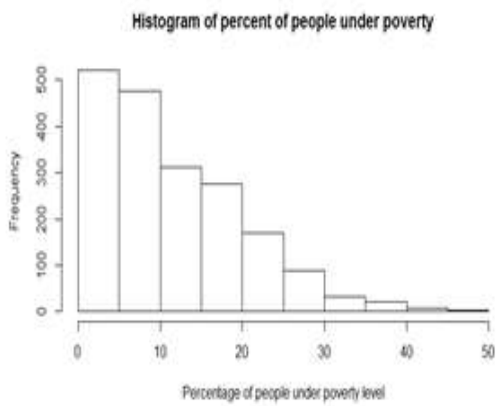


Fig. 2

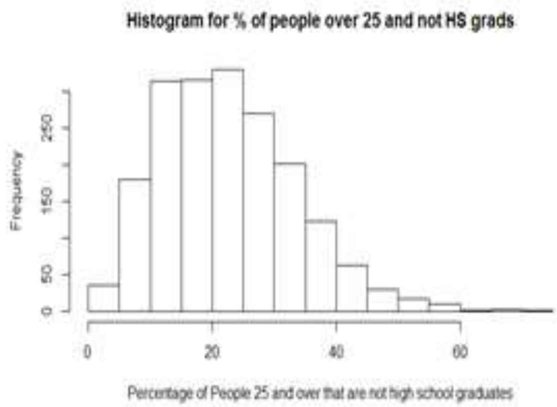


Fig. 3

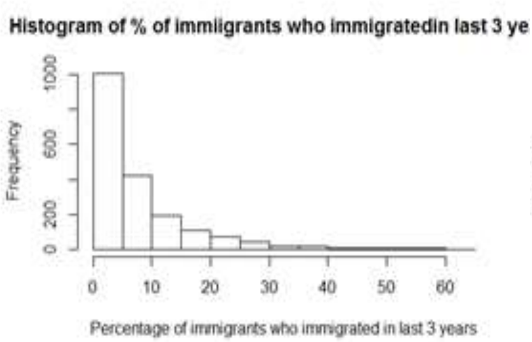


Fig. 4

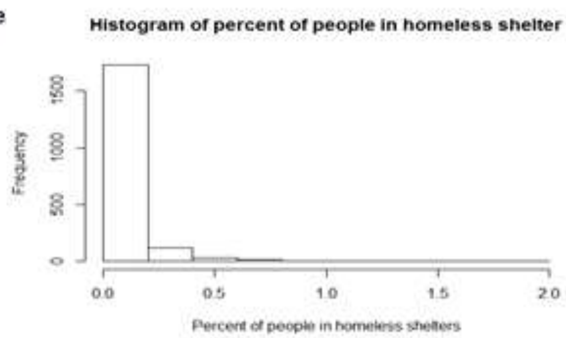
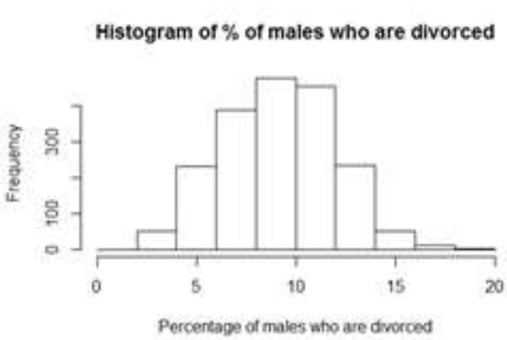


Fig. 5



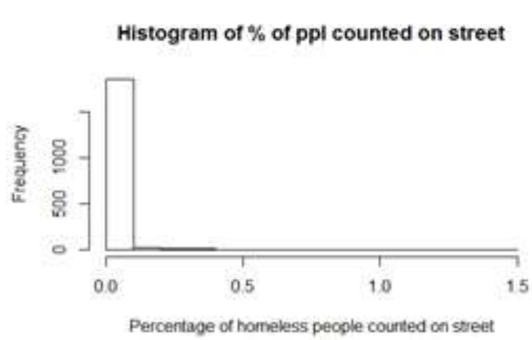
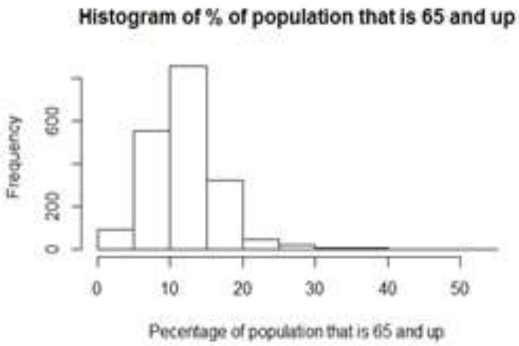


Fig. 6

Fig. 7



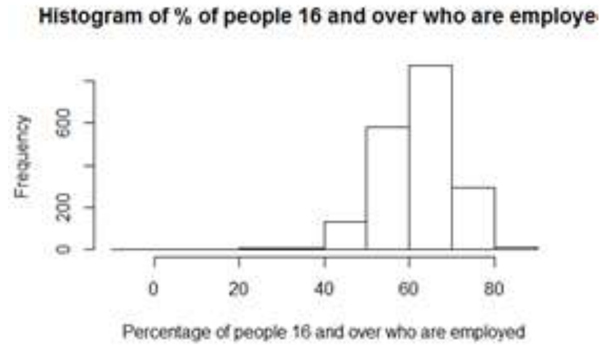
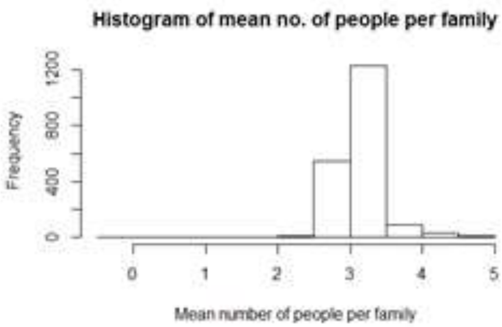


Fig. 8

Fig. 9



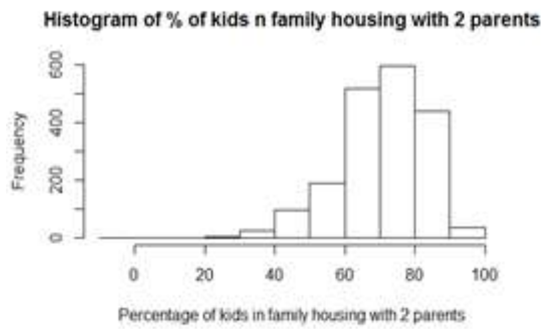
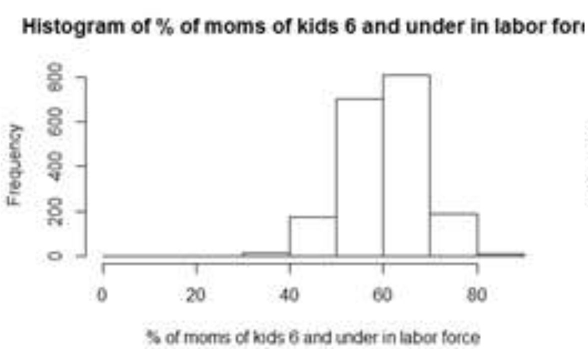


Fig. 10

Fig. 11



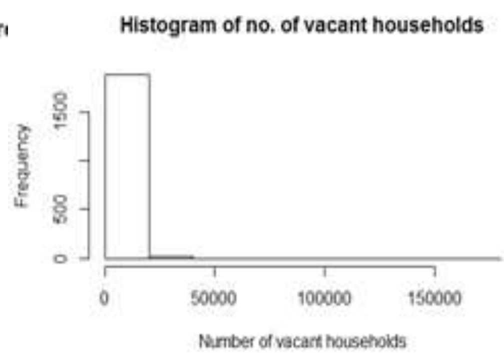


Fig. 12

Fig. 13

Fig. 14

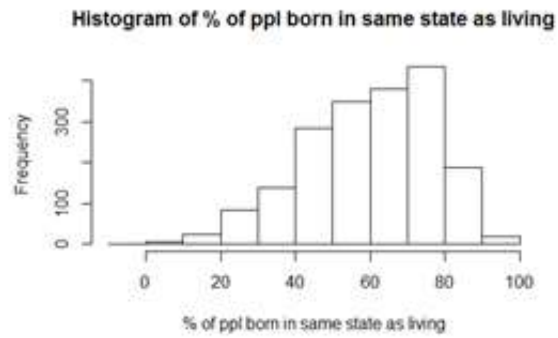


Fig. 15

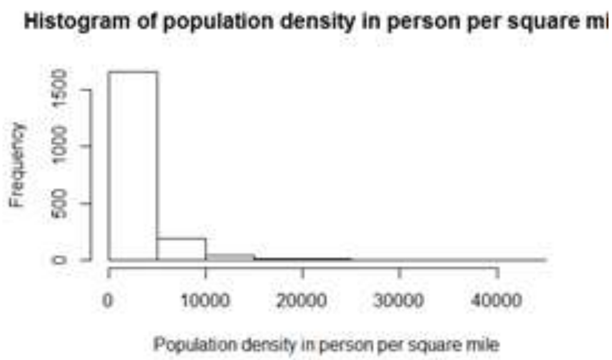


Fig. 16

Checking for Assumptions

1. **Independence of Errors**

We tested for duplicate values in our data set by ensuring that none of the rows have the same community details. Our data set didn't violate this assumption of multicollinearity.

2. **Testing Normality of Errors**

This assumption states that the errors between observed and predicted values are normally distributed. We plot a histogram to test this assumption
> *hist(resid)*

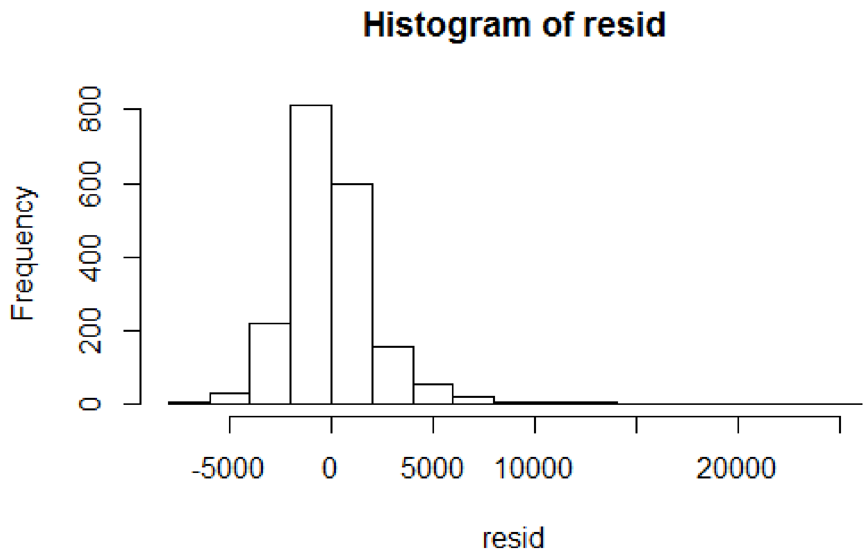


Fig. 17

It can been from the above histogram that the errors are normally distributed and hence our assumption is correct.

Apart from looking at histograms for each of our dependent variables as shown above, we also looked at the values of skewness[3] and kurtosis[3] for each of our independent variables to check for normality. The following table shows these values:

Sr. No.	Variable Name	Skewness	Kurtosis
1	PctPopUnderPov	1.03	3.79
2	PctNotHSGrad	0.75	3.79
3	PctHomShelt	5.04	47.03
4	PctImmig	2.42	10.35
5	MalePctDivorce	0.09	2.66
6	PctOnStreet	-0.97	1.88
7	agePct65up	1.68	12.09
8	PctEmploy	-0.46	3.65
9	PersPerFam	1.73	9.20
10	PctKids2Par	-0.65	3.24
11	PctWorkMomYoungKids	-0.16	3.22
12	HousVacant	16.56	401.87
13	RentMedian	0.79	3.25
14	PctBornSameState	-0.51	2.68
15	PopDens	4.47	40.46

Table 2

3. **Checking for outliers**

We check for outliers by getting observations that are 3 or more standard deviations away from the predicted values.

```
> out = lm(TotalCrime~PctPopUnderPov+Percent.in.homeless.shelters+PctOnStreet+agePct65up+MalePctD
ivorce+PersPerFam+PctWorkMomYoungKids+HousVacant+PctBornSameState+PctImmig,data=cleansample)
> pred=out$fitted.values
> resid=out$residuals
> resid.sd=sd(resid)
> resid[abs(resid)>=3*resid.sd]
       92        96       126       146       310       346       393       482       500 
16850.835  6983.430  7173.320 11235.722 11699.643  7978.218  8756.894  6921.179  7953.733 
      508       536       550       695       700       772       805       860       922 
-7421.613  7151.821  8313.353  9923.971  8322.349  6946.567 13771.524 15793.564  7756.582 
      984      1101      1287      1360      1625      1656      1722 
 7391.626  9499.094 12241.879  7114.961 25851.954  7432.887  6740.203 
```

There are a total of 50 outliers that are 3 or more standard deviations away from the predicted values.

4. **Constant error**

This assumption states that the variances of the errors of prediction are the same for all predicted values. To check for constant error, we can plot the predicted values with the residuals to see whether the residuals vary consistently with the predicted values

> plot(pred,resid)

> abline(a=0,b=0)

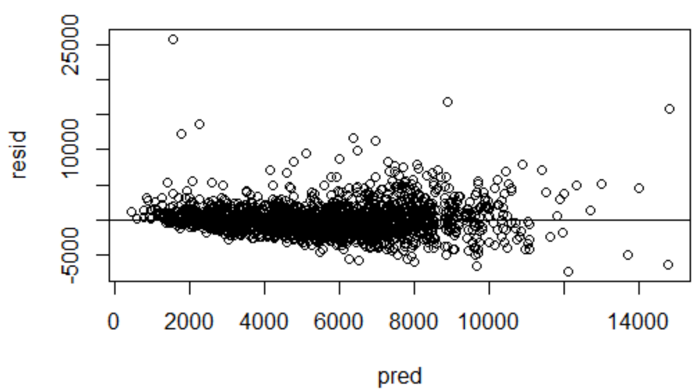


Fig. 18

We see that the residuals vary evenly with a few outliers

Inferential Test Used To Analyze the Data

Our data set has a number of variables that measure socio-economic factors to describe a community. We wanted to analyze the relationship between the total crime rate and all the socio-economic factors. Since there are many predictors (multiple independent variables) and one outcome (one dependent variable) we realized that the best inferential test to be used is multiple regression. Multiple regression is one of the most commonly used inferential tests in statistics.

The original dataset (CommunityCrimeData.xls) which we obtained from the website had a total of 2215 rows and 147 columns of which there were 129 independent variables and 18 dependent variables. However, on going through them, we found that a considerable number of the variables(columns) had most of their data missing and it would have been wrong to make any conclusions based on the little data that was available for those variables (for example: police

offices per 100K population had missing values in 1872 rows out of 2215 rows). We then removed about 25 columns which had their majority of the data missing.
On the resulting dataset we applied the Excel function CORREL taking into consideration all the independent variables and our dependent variable, which is TotalCrime. On analyzing the results we were able to shortlist 43 variables which had a significant correlation coefficient with our D.V.
We imported the resulting dataset in R and got rid of those rows which had blank or NAs assigned to them using the command 'na.omit'. A multiple regression model was built taking into account all these variables and the dependent variable – Total crime. A list of the details of all the Independent variables is provided below.
The results of the first model is listed in the appendix[4]. Most of the Independent variables which had a significant correlation with TotalCrimes were found to possess a p-value much higher than the significance level of 0.05. Following is the R command used to perform multiple regression for 43 independent variables.

```
Call:
lm(formula = TotalCrime ~ PctPopUnderPov + PctUnemployed + PctNotHSGrad +
    PctImmigRec10 + PctPersOwnOccup + PctImig + PctHomShelt +
    PctOnStreet + population + agePct12t21 + agePct12t29 + agePct16t24 +
    agePct65up + racePctAsian + racepctblack + racePctHisp +
    racePctWhite + medIncome + pctWwage + pctWFarmSelf + pctWInvInc +
    whitePerCap + blackPerCap + AsianPerCap + PctEmploy + PctEmplManu +
    MalePctDivorce + FemalePctDiv + PersPerFam + PctFam2Par +
    PctKids2Par + PctWorkMomYoungKids + PersPerRentOccHous +
    HousVacant + PctHousOccup + PctHousNoPhone + RentMedian +
    PctBornSameState + PopDens, data = CrimeDataCleaned)
```

A second model was then constructed in which all the IVs having a p-value of more than 0.05 were removed. Following is the R command used to perform multiple regression for 15 independent variables.
The output for second model is mentioned in appendix [5]
In the above model we can see that there were significant changes in the correlation coefficient for some of the variables. This led us to testing the independent variables for multicollinearity.

Checking for Multicollinearity

The 15 independent variables from our second model were compared with each other to test for multicollinearity and the following results were obtained.

	PctPopUn	PctNotHSC	PctImmig	PctHomSh	PctOnStre	agePct65u	PctEmploy	MalePctDi	PersPerFa	PctKids2Pa	PctWorkM	HousVaca	RentMedi	PctBornSa	PopDens
PctPopUnderPov	1.00														
PctNotHSGrad	0.66	1.00													
PctImmig	0.02	0.21	1.00												
PctHomShelt	0.22	0.14	0.07	1.00											
PctOnStreet	0.09	0.06	0.12	0.16	1.00										
agePct65up	0.07	0.23	-0.12	0.03	-0.01	1.00									
PctEmploy	-0.69	-0.60	0.04	-0.14	-0.04	-0.62	1.00								
MalePctDivorce	0.39	0.37	-0.11	0.29	0.17	0.16	-0.23	1.00							
PersPerFam	0.15	0.32	0.45	-0.02	0.01	-0.56	0.12	-0.25	1.00						
PctKids2Par	-0.75	-0.66	-0.03	-0.35	-0.16	-0.22	0.54	-0.71	0.03	1.00					
PctWorkMomYoungKids	-0.10	-0.07	-0.26	-0.05	-0.03	0.01	0.24	0.18	-0.26	-0.10	1.00				
HousVacant	0.13	0.06	0.09	0.17	0.11	0.02	-0.06	0.18	0.00	-0.22	-0.04	1.00			
RentMedian	-0.65	-0.54	0.45	-0.06	0.02	-0.19	0.52	-0.41	0.09	0.51	-0.19	-0.05	1.00		
PctBornSameState	0.14	0.25	-0.48	-0.06	-0.12	0.08	-0.16	-0.05	0.02	-0.06	0.13	-0.10	-0.41	1.00	
PopDens	0.08	0.22	0.63	0.17	0.11	-0.01	-0.02	0.08	0.19	-0.24	-0.12	0.18	0.19	-0.22	1.00

Fig. 19

In Figure 16, we can observe that the green colored cells denote a higher correlation (closer to 1) and the red cells denote a negative correlation(closer to –1). Thus it was found that there was significant correlation among 9 of the IVs. For example – Population under poverty (PctPopUnderPov) and population which was not high school graduate (PctNotHSGrad) had a correlation coefficient of 0.66. This is quite obvious and hence in our refined model only one of the two variables was taken. In this case we considered the variable Population under poverty as an independent variable in our model. Similar manipulations were done for rest of the IVs. The final model was created taking into account the effect of multicollinearity by eliminating 5 of the 9 highly correlated IVs.
Following is the final model with 10 independent variables.

```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)         -5.104e+03  1.209e+03  -4.221 2.55e-05 ***
PctPopUnderPov       1.302e+02  7.211e+00  18.052  < 2e-16 ***
PctHomShelt          2.383e+03  4.267e+02   5.585 2.68e-08 ***
PctOnStreet          2.659e+03  8.545e+02   3.112 0.001889 ** 
agePct65up           5.659e+01  1.374e+01   4.119 3.98e-05 ***
MalePctDivorce       4.764e+02  2.248e+01  21.193  < 2e-16 ***
PersPerFam           1.063e+03  3.202e+02   3.321 0.000915 ***
PctWorkMomYoungKids  2.732e+01  7.099e+00   3.848 0.000123 ***
HousVacant           4.124e-02  8.958e-03   4.604 4.42e-06 ***
PctBornSameState    -2.167e+01  3.782e+00  -5.731 1.16e-08 ***
PctImig              1.722e+01  8.196e+00   2.102 0.035728 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2243 on 1890 degrees of freedom
Multiple R-squared:  0.5198,    Adjusted R-squared:  0.5173 
F-statistic: 204.6 on 10 and 1890 DF,  p-value: < 2.2e-16
```

We received a final RSquared value of 0.5173 indicating that our final model was able to account for 51.73% of the total variations in crime.

Final Equation for our model

TotalCrime = -5.104e+03 + 1.302e+02*PctPopUnderPov* + 2.383e+03*PctHomShelt* + 2.659e+03*PctOnStreet* + 5.659e+01*agePct65up* + 4.764e+02*MalePctDivorce* + 1.063e+03*PersPerFam* + 2.732e+01*PctWorkMomYoungKids* + 4.124e-02*HousVacant* - 2.167e+01*PctBornSameState* + 1.722e+01*PctImig*

Discussion

Based on our final model, which is mentioned above, we can conclude that the variations in Total Crime(our Dependent variable) can be explained up to 51.73% by taking into consideration the following factors. The degree of effect that each factor has is also described.

1. Percentage of people under the poverty level – As expected, this factor has a positive correlation coefficient of 130.2 with total crime in an area. This means that each percentage increase in population under poverty in a community results in an increase of total crime by 130.2 units.

2. Number of people in homeless shelters – This factor has a high correlation with total crime rate and with each percent increase in people in homeless shelter the crime rates go up by 2383 units. However, this does not imply that the government should do away with homeless shelters. Rather, the policies should be directed in such a way that people are capable enough to own or rent a house, perhaps to instill a sense of belonging in them.

3. Percentage of homeless people counted on street.- As obvious, this factor also has a positive correlation coefficient with total crime. With each percent increase in people on the streets the total crime goes up by 2659 units.

4. Percentage of population that is 65 and over in age – This result was somewhat confounding to us as this implies that as the number of senior citizens increases the crime rates go up by 56 units. However, this could also imply that the government should take up some measures to ensure the safety of older aged people as they are an easy target.

5. Percentage of males who are divorced – With each percent increase in males who are divorced the crime rates go up by 476.4 units.

6. Mean number of people per family – Increase in mean number of people per family leads to an increase of 1063 units in Total crime. This is quite obvious as more family members put higher pressure on the income resources of a family.

7. Percentage of moms of kids 6 and under in labor force- We were baffled by this result. However, the correlation coefficient is quite small in this case and with each percent increase in working mom population having kids aged 6 or under the total crime increase by 27 units

8. Number of vacant households - The correlation coefficient is quite small for this factor being only 0.004 meaning that with an increase in vacant household the crime rates go up by 0.04

9. Percent of people born in the same state as currently living - with an increase in percent of people born in the same state as currently living the crime rates decrease by 21.67 units.

10. Percentage of immigrants who immigrated within last 3 years – With every percent increase in number of immigrants who immigrated in the last 3 years the crime rates go up by 17.22.

Conclusion

On the basis of the statistical analysis we conducted, we identified 10 socio-economic factors that significantly affect crime happening in the community.

Limitations

1. Data for all the communities within United States is not available. Communities not found in both the census and the crime data were not considered. Also, data for the communities present in the data set had missing values.

2. There are other factors which the census does not capture such as number of police personnel deployed that may affect crime in a community.

3. Sometimes, there are certain dynamics in a community that cannot be captured by data. Thus, it might not be possible to completely identify all factors that affect crime in a community.

Future Scope

1. Data sets other than the census such as police personnel data can be combined with the community wise crime data to identify other factors which significantly affect crime in a community.

2. Data for the missing communities can be collected and the accuracy of the model can be checked by predicting crime rate for these communities basis this model. Further changes in the model can be made basis data collected for these communities.

Appendix

[1] A community is a group of people living in the same place or having a particular characteristic in common.
[2] Lichman, M. (2011, March 2). UCI machine learning repository: Communities and crime Unnormalized data set. [Retrieved October 4, 2016], from UCI machine learning repository, https://archive.ics.uci.edu/ml/datasets/Communities+and+Crime+Unnormalized.
[3] Skewness is a measure of symmetry and kurtosis is a measure of whether the data is heavily tailed or lightly tailed as compared to a normal distribution. These 2 functions can be found in the Moments library in R.

[4] R Output for 1st model.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-4.606e+02	4.27E+03	-0.108	0.914153
PctPopUnderPov	7.656e+01	2.12E+01	3.612	0.000312 ***

PctUnemployed	-3.669e+01	4.00E+01	-0.918	0.358892
PctNotHSGrad	-3.085e+01	1.35E+01	-2.279	0.022780 *
PctImmigRec10	5.725e+00	4.13E+00	1.386	0.166026
PctPersOwnOccup	8.438e+00	1.05E+01	0.803	0.421898
PctImig	6.675e+01	1.69E+01	3.953	8.01e-05 ***
PctHomShelt	1.479e+03	4.18E+02	3.541	0.000409 ***
PctOnStreet	2.543e+03	7.91E+02	3.217	0.001319 **
population	-1.727e-03	7.02E-04	-2.459	0.014014 *
agePct12t21	2.616e+01	6.29E+01	0.416	0.677397
agePct12t29	2.023e+01	6.29E+01	0.322	0.747673
agePct16t24	-3.252e+01	8.79E+01	-0.370	0.711468
agePct65up	1.397e+02	4.30E+01	3.248	0.001183 **
racePctAsian	4.358e+00	2.14E+01	0.204	0.838354
racepctblack	2.501e+01	1.58E+01	1.587	0.112586
racePctHisp	1.146e+01	9.47E+00	1.210	0.226518
racePctWhite	8.157e+00	1.46E+01	0.560	0.575291
medIncome	-7.735e-03	2.17E-02	-0.357	0.721505
pctWWage	2.344e+01	2.60E+01	0.902	0.367136
pctWFarmSelf	-7.435e+01	8.36E+01	-0.890	0.373814
pctWInvInc	1.352e+01	1.25E+01	1.085	0.277997
whitePerCap	5.374e-02	2.54E-02	2.117	0.034402 *
blackPerCap	-7.792e-03	6.05E-03	-1.288	0.198072
AsianPerCap	-2.011e-03	5.71E-03	-0.352	0.724761
PctEmploy	8.438e+01	2.09E+01	4.042	5.52e-05 ***
PctEmplManu	-1.884e+01	7.88E+00	-2.390	0.016965 *
MalePctDivorce	2.445e+02	5.38E+01	4.548	5.75e-06 ***
FemalePctDiv	3.690e+01	5.22E+01	0.707	0.479516
PersPerFam	1.531e+03	5.92E+02	2.589	0.009710 **
PctFam2Par	3.488e+01	4.31E+01	0.809	0.418662
PctKids2Par	-1.595e+02	3.92E+01	-4.069	4.91e-05 ***
PctWorkMomYoungKid	-1.786e+01	8.34E+00	-2.143	0.032260 *
PersPerRentOccHous	-2.274e+02	2.82E+02	-0.807	0.419865
HousVacant	7.372e-02	2.52E-02	2.923	0.003508 **
PctHousOccup	-1.123e+01	1.29E+01	-0.869	0.384965
PctHousNoPhone	9.482e-02	2.90E+01	0.003	0.997389
RentMedian	-3.404e+00	8.80E-01	-3.867	0.000114 ***
PctBornSameState	-9.637e+00	4.48E+00	-2.152	0.031562 *
PopDens	-1.395e-01	2.51E-02	-5.567	2.97e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 2052 on 1861 degrees of freedom
Multiple R-squared: 0.6045, Adjusted R-squared: 0.5962
F-statistic: 72.93 on 39 and 1861 DF, p-value: < 2.2e-16

[5] R code output for second model which includes 15 independent variables.

```
Call:
lm(formula = TotalCrime ~ PctPopUnderPov + PctNotHSGrad + PctImig +
    PctHomShelt + PctOnStreet + agePct65up + PctEmploy + MalePctDivorce +
    PersPerFam + PctKids2Par + PctWorkMomYoungKids + HousVacant +
    RentMedian + PctBornSameState + PopDens, data = CrimeDataCleaned)

Residuals:
    Min      1Q  Median      3Q     Max
-7478.3 -1152.1  -237.4   932.4 24669.8

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          5.742e+03  2.241e+03   2.562 0.010469 *
PctPopUnderPov       7.721e+01  1.477e+01   5.227 1.91e-07 ***
PctNotHSGrad        -6.088e+01  9.399e+00  -6.477 1.19e-10 ***
PctImig              6.962e+01  1.123e+01   6.199 6.98e-10 ***
PctHomShelt          1.321e+03  4.050e+02   3.262 0.001126 **
PctOnStreet          2.409e+03  7.924e+02   3.040 0.002398 **
agePct65up           1.452e+02  2.201e+01   6.599 5.36e-11 ***
PctEmploy            1.017e+02  1.518e+01   6.701 2.73e-11 ***
MalePctDivorce       2.093e+02  2.952e+01   7.090 1.89e-12 ***
PersPerFam           1.531e+03  3.551e+02   4.312 1.71e-05 ***
PctKids2Par         -1.642e+02  1.016e+01 -16.155  < 2e-16 ***
PctWorkMomYoungKids -1.879e+01  7.524e+00  -2.497 0.012603 *
HousVacant           2.712e-02  8.431e-03   3.216 0.001321 **
RentMedian          -2.412e+00  5.418e-01  -4.453 8.96e-06 ***
PctBornSameState    -1.419e+01  3.828e+00  -3.708 0.000215 ***
PopDens             -1.751e-01  2.299e-02  -7.618 4.06e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2077 on 1885 degrees of freedom
Multiple R-squared:  0.5897,    Adjusted R-squared:  0.5864
F-statistic: 180.6 on 15 and 1885 DF,  p-value: < 2.2e-16
```