Crime Analysis

Shanmugavalli Muthu

2022-06-29

## Summary

Crime is one of the disturbing features of society that make life difficult for ordinary people. Although crimes are unpredictable, forecasting tools can help predict them. Theft, robbery, property damage, assault, and homicide can all be caused by a variety of factors. They also have a pattern in which the crime rate rises in a few months and then drops in others. This research or crime prediction is essential because it provides insight into why, when, and how criminal activities occur, aiding in improving the efficiency of criminal investigations and predicting the likelihood of future crimes. Government authorities can improve security in areas where crime is more prevalent by providing additional police coverage and surveillance cameras. In recent times, there has been a lot of crime analysis in order to reduce crimes in the future. The main aim of this study is to determine in which time period the crime rates peak and which socio-economic factors lead to most crime. The data is from the Chicago crime data set from 2008-2012 and used the linear regression method to analyze the data. The main findings would be which socio-economic index has a relationship with the crime rate and the time and pattern of the occurrence. The main aim of this study is to help the law enforcement authorities in hastening the investigation of crimes and identifying potential future criminal activity.

Keywords: Data analysis, Crime analysis, Crime prediction, Forecasting.

## Introduction

Crimes are usually unpredictable. Some regions naturally encourage crime and it fluctuates with societal trends in human behavior. The goal of this study is to identify criminal tendencies in the Chicago crime dataset from 2008 to 2012. The locations in Chicago areas with a high rate of criminal activity, the most common types of crimes, and the distribution of crimes over time are all important types of information. A socioeconomic indicator dataset, as seen in table 1, has also been provided to us and provides information on the standard of living in each community in Chicago. Linear regression is used to discover significant correlations between predictor variables (socioeconomic indicators) and crime rates, which aids in crime reduction in the affected areas.

## ID Case.Number Date Block IUCR  
## 1 11645833 JC213044 05/05/2012 12:25:00 PM 057XX W OHIO ST 1153  
## 2 11227247 JB147078 01/01/2012 09:00:00 AM 105XX S INDIANAPOLIS AVE 1153  
## 3 10225605 HY412867 07/11/2012 09:00:00 AM 017XX W ALBION AVE 1153  
## 4 10225648 HY412901 08/01/2011 08:00:00 AM 080XX S TRUMBULL AVE 1153  
## 5 11042125 JA376558 12/16/2011 12:00:00 AM 003XX W 64TH ST 1153  
## 6 11042582 JA377037 01/01/2011 12:01:00 AM 054XX S CALIFORNIA AVE 1754  
## Primary.Type Description  
## 1 DECEPTIVE PRACTICE FINANCIAL IDENTITY THEFT OVER $ 300  
## 2 DECEPTIVE PRACTICE FINANCIAL IDENTITY THEFT OVER $ 300  
## 3 DECEPTIVE PRACTICE FINANCIAL IDENTITY THEFT OVER $ 300  
## 4 DECEPTIVE PRACTICE FINANCIAL IDENTITY THEFT OVER $ 300  
## 5 DECEPTIVE PRACTICE FINANCIAL IDENTITY THEFT OVER $ 300  
## 6 OFFENSE INVOLVING CHILDREN AGG SEX ASSLT OF CHILD FAM MBR  
## Location.Description Arrest Domestic Beat District Ward Community.Area  
## 1 false false 1511 15 29 25  
## 2 RESIDENCE false false 432 4 10 52  
## 3 APARTMENT false false 2432 24 40 1  
## 4 RESIDENCE false false 834 8 18 70  
## 5 HOSPITAL BUILDING/GROUNDS false false 722 7 20 68  
## 6 APARTMENT true true 923 9 14 63  
## FBI.Code X.Coordinate Y.Coordinate Year Updated.On Latitude  
## 1 11 NA NA 2012 04/06/2019 04:04:43 PM NA  
## 2 11 NA NA 2012 02/11/2018 03:57:41 PM NA  
## 3 11 1163498 1943889 2012 02/10/2018 03:50:01 PM 42.00167  
## 4 11 1154779 1851149 2011 02/09/2018 03:44:29 PM 41.74736  
## 5 11 NA NA 2011 08/05/2017 03:50:08 PM NA  
## 6 02 NA NA 2011 08/13/2017 03:50:54 PM NA  
## Longitude Location  
## 1 NA   
## 2 NA   
## 3 -87.67386 (42.00167049, -87.673863642)  
## 4 -87.70842 (41.747362057, -87.708423712)  
## 5 NA   
## 6 NA

## Literature Review

Analysis of crime and prediction were explained by some approaches like KNN, decision trees, support vector machines, and extra trees. Decision trees, KNN, and extra tree give good accuracy of results. Authors discussed workflow-data collection, data preprocessing, classification and pattern identification, prediction, and evaluation (Gahalot et al.,2020). Research illustrates how Support Vector Machine, Fuzzy Theory, Artificial Neural Network (ANN), and multivariate time series help in crime analysis. They concluded that models’ results will be accurate only if they are highly trained. (Shamsuddin et al., 2017). The authors discussed crime distribution as in which months have more crimes and which months have fewer crimes. This was analyzed and visualized by using tableau software. They also looked at the crime rate by day and analyzed which days have higher crime rates. (Nwankwo et al., 2018). Data analytics plays an influential role in finding and making decisions in different circumstances. Appropriate data analytics tools and methods will help determine basic patterns and connections in datasets. (Sathyadevan et al., 2014). The study included many innovations for forecast models, with numerous driven by unique features such as GIS and Depp learning methods. They also covered a few topics under methodological models like exponential smoothing, Box Jenkins arima multivariate transfer models, neural networks, and cointegrated econometric models. (Gorr et al., 2003). This paper discussed CCMinner and its crime predictions and crime pattern analysis in order to provide support while making judgements regarding crime analysis. The authors discussed real-time crime case pattern analysis, including front-end and back-end users as well as the use of the Power BI tool. (Peiris et al., 2020).

## Research Questions

• Determine which time period(day) has higher crime rates compared to other days. • To find whether there is a correlation between the crime rate and socioeconomic indicators.

## Theory

This study is investigating the relationship between a crime rate and socioeconomic indicators of public health. This study includes: 1. Evaluate that there is 100% correlation between the crime rate and socioeconomic indicators. 2. Evaluate the correlation between the crime rate and socioeconomic indicators is null. It is the assumption that there is a relationship between criminal activities and the living condition in each community area. Linear regression is applied to find the correlations between interdependent variables (the socioeconomic indicators) and the dependent variable (crime rate).

## Data

For this study, Chicago crimes data to created from records of crimes cases by the Chicago Police Department from Crimes - 2001 to Present | City of Chicago | Data Portal. The data includes:

## ID Case.Number Date Block IUCR  
## 1 11645833 JC213044 05/05/2012 12:25:00 PM 057XX W OHIO ST 1153  
## 2 11227247 JB147078 01/01/2012 09:00:00 AM 105XX S INDIANAPOLIS AVE 1153  
## 3 10225605 HY412867 07/11/2012 09:00:00 AM 017XX W ALBION AVE 1153  
## 4 10225648 HY412901 08/01/2011 08:00:00 AM 080XX S TRUMBULL AVE 1153  
## 5 11042125 JA376558 12/16/2011 12:00:00 AM 003XX W 64TH ST 1153  
## 6 11042582 JA377037 01/01/2011 12:01:00 AM 054XX S CALIFORNIA AVE 1754  
## Primary.Type Description  
## 1 DECEPTIVE PRACTICE FINANCIAL IDENTITY THEFT OVER $ 300  
## 2 DECEPTIVE PRACTICE FINANCIAL IDENTITY THEFT OVER $ 300  
## 3 DECEPTIVE PRACTICE FINANCIAL IDENTITY THEFT OVER $ 300  
## 4 DECEPTIVE PRACTICE FINANCIAL IDENTITY THEFT OVER $ 300  
## 5 DECEPTIVE PRACTICE FINANCIAL IDENTITY THEFT OVER $ 300  
## 6 OFFENSE INVOLVING CHILDREN AGG SEX ASSLT OF CHILD FAM MBR  
## Location.Description Arrest Domestic Beat District Ward Community.Area  
## 1 false false 1511 15 29 25  
## 2 RESIDENCE false false 432 4 10 52  
## 3 APARTMENT false false 2432 24 40 1  
## 4 RESIDENCE false false 834 8 18 70  
## 5 HOSPITAL BUILDING/GROUNDS false false 722 7 20 68  
## 6 APARTMENT true true 923 9 14 63  
## FBI.Code X.Coordinate Y.Coordinate Year Updated.On Latitude  
## 1 11 NA NA 2012 04/06/2019 04:04:43 PM NA  
## 2 11 NA NA 2012 02/11/2018 03:57:41 PM NA  
## 3 11 1163498 1943889 2012 02/10/2018 03:50:01 PM 42.00167  
## 4 11 1154779 1851149 2011 02/09/2018 03:44:29 PM 41.74736  
## 5 11 NA NA 2011 08/05/2017 03:50:08 PM NA  
## 6 02 NA NA 2011 08/13/2017 03:50:54 PM NA  
## Longitude Location  
## 1 NA   
## 2 NA   
## 3 -87.67386 (42.00167049, -87.673863642)  
## 4 -87.70842 (41.747362057, -87.708423712)  
## 5 NA   
## 6 NA

Census data for socioeconomic indicators in Chicago, 2008 – 2012 was retrieved from Census Data - Selected socioeconomic indicators in Chicago, 2008 – 2012 | City of Chicago | Data Portal. The data includes:

## Community.Area COMMUNITY.AREA.NAME PERCENT.OF.HOUSING.CROWDED  
## 1 1 Rogers Park 7.7  
## 2 2 West Ridge 7.8  
## 3 3 Uptown 3.8  
## 4 4 Lincoln Square 3.4  
## 5 5 North Center 0.3  
## 6 6 Lake View 1.1  
## PERCENT.HOUSEHOLDS.BELOW.POVERTY PERCENT.AGED.16..UNEMPLOYED  
## 1 23.6 8.7  
## 2 17.2 8.8  
## 3 24.0 8.9  
## 4 10.9 8.2  
## 5 7.5 5.2  
## 6 11.4 4.7  
## PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA PERCENT.AGED.UNDER.18.OR.OVER.64  
## 1 18.2 27.5  
## 2 20.8 38.5  
## 3 11.8 22.2  
## 4 13.4 25.5  
## 5 4.5 26.2  
## 6 2.6 17.0  
## PER.CAPITA.INCOME HARDSHIP.INDEX  
## 1 23939 39  
## 2 23040 46  
## 3 35787 20  
## 4 37524 17  
## 5 57123 6  
## 6 60058 5

Before creating a new dataset from the existing ones by merging, a new dataset was created to count how many crimes occur in each community:

## # A tibble: 6 × 2  
## Community.Area Crime.Rate  
## <int> <int>  
## 1 0 25  
## 2 1 28096  
## 3 2 24055  
## 4 3 26025  
## 5 4 12641  
## 6 5 10957

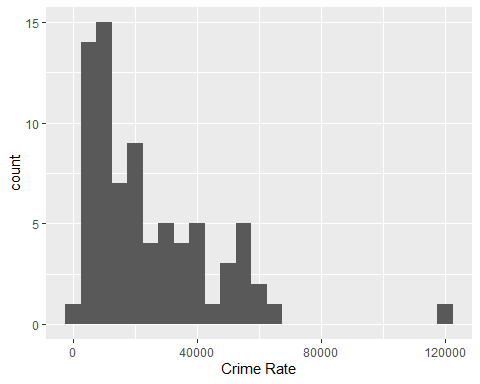
This data set was merged with the socioeconomic indicators data set which was used for regression model. All the N/A values were removed

## Community.Area Crime.Rate COMMUNITY.AREA.NAME PERCENT.OF.HOUSING.CROWDED  
## 1 1 28096 Rogers Park 7.7  
## 2 2 24055 West Ridge 7.8  
## 3 3 26025 Uptown 3.8  
## 4 4 12641 Lincoln Square 3.4  
## 5 5 10957 North Center 0.3  
## 6 6 38669 Lake View 1.1  
## PERCENT.HOUSEHOLDS.BELOW.POVERTY PERCENT.AGED.16..UNEMPLOYED  
## 1 23.6 8.7  
## 2 17.2 8.8  
## 3 24.0 8.9  
## 4 10.9 8.2  
## 5 7.5 5.2  
## 6 11.4 4.7  
## PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA PERCENT.AGED.UNDER.18.OR.OVER.64  
## 1 18.2 27.5  
## 2 20.8 38.5  
## 3 11.8 22.2  
## 4 13.4 25.5  
## 5 4.5 26.2  
## 6 2.6 17.0  
## PER.CAPITA.INCOME HARDSHIP.INDEX  
## 1 23939 39  
## 2 23040 46  
## 3 35787 20  
## 4 37524 17  
## 5 57123 6  
## 6 60058 5

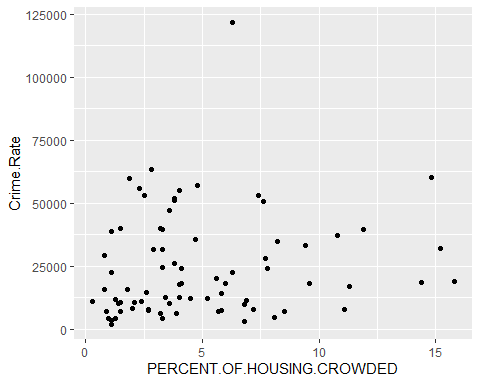
## Methodology

In this study, we use multiple linear regression to model the relationship between a crime rate and socioeconomic variables (dependent and independent variables). The equation of simple linear regression is similar to that of the slope formula. However, the equation for multiple regression line formula is Y =a\_1 X\_1+ a\_2 X\_2+ a\_3 X\_3+ C Where Y is dependent variable, X1 – X3 are independent variables, a1-a3 are slope and c is error/constant. Figure 1 (a) shows the distribution of crime rate from 2008 to 2012. Figure 1 (b) – (h) illustrates the correlation between socioeconomic indicators against the crime rate. For example, In Figure 1 (b) we can see that the there is a correlation between housing units with more than one person per room is plotted against the crime rate.

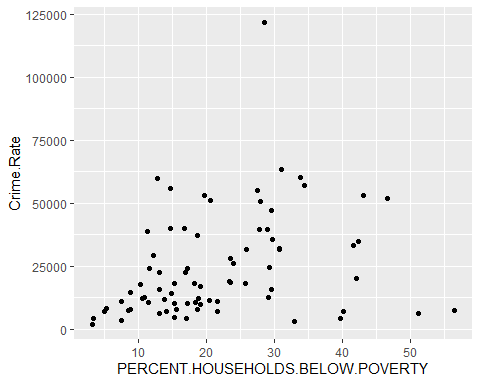
ggplot(DF, aes(x=Crime.Rate))+geom\_histogram(binwidth = 5000) + labs(x="Crime Rate")



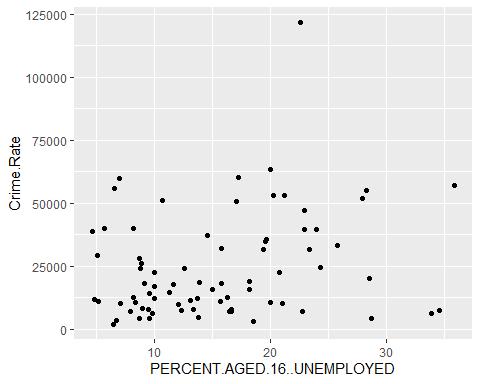
ggplot(DF, aes(x=PERCENT.OF.HOUSING.CROWDED, y=Crime.Rate)) + geom\_point()



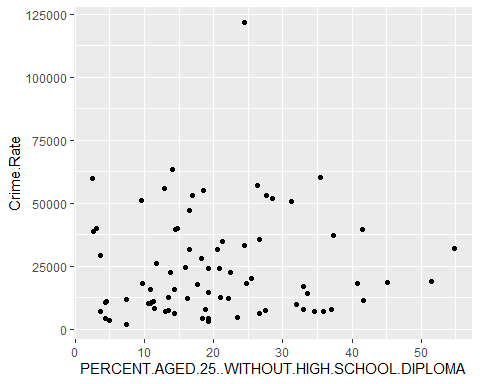
ggplot(DF, aes(x=PERCENT.HOUSEHOLDS.BELOW.POVERTY, y=Crime.Rate)) + geom\_point()



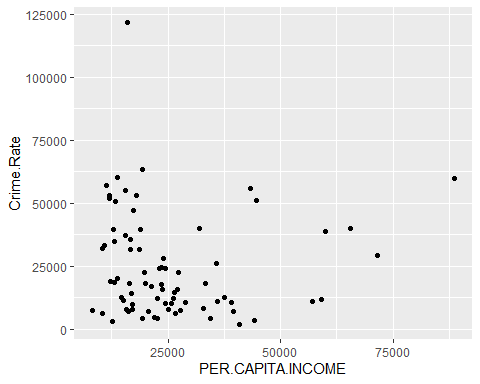
ggplot(DF, aes(x=PERCENT.AGED.16..UNEMPLOYED, y=Crime.Rate)) + geom\_point()



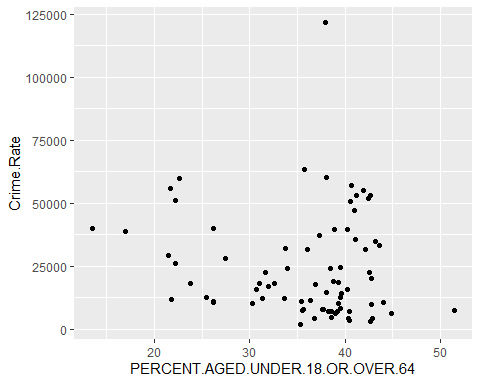
ggplot(DF, aes(x=PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA, y=Crime.Rate)) + geom\_point()



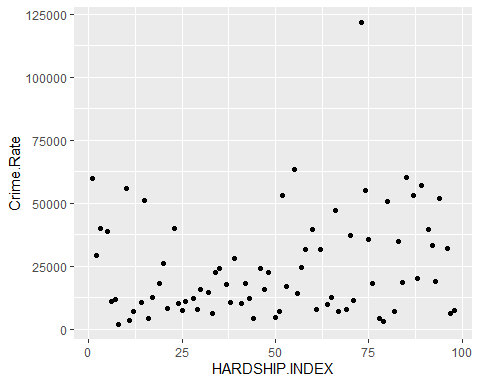
ggplot(DF, aes(x=PER.CAPITA.INCOME, y=Crime.Rate)) + geom\_point()



ggplot(DF, aes(x=PERCENT.AGED.UNDER.18.OR.OVER.64, y=Crime.Rate)) + geom\_point()



ggplot(DF, aes(x=HARDSHIP.INDEX, y=Crime.Rate)) + geom\_point()



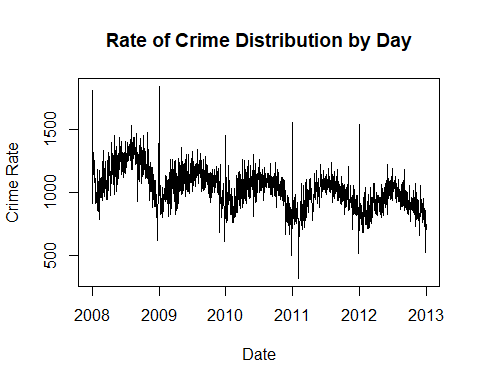
summary(DF)

## Community.Area Crime.Rate COMMUNITY.AREA.NAME PERCENT.OF.HOUSING.CROWDED  
## Min. : 1 Min. : 1856 Length:77 Min. : 0.300   
## 1st Qu.:20 1st Qu.: 9741 Class :character 1st Qu.: 2.300   
## Median :39 Median : 17970 Mode :character Median : 3.800   
## Mean :39 Mean : 24386 Mean : 4.923   
## 3rd Qu.:58 3rd Qu.: 35727 3rd Qu.: 6.800   
## Max. :77 Max. :121724 Max. :15.800   
## PERCENT.HOUSEHOLDS.BELOW.POVERTY PERCENT.AGED.16..UNEMPLOYED  
## Min. : 3.30 Min. : 4.70   
## 1st Qu.:13.20 1st Qu.: 9.20   
## Median :18.90 Median :13.90   
## Mean :21.77 Mean :15.37   
## 3rd Qu.:29.20 3rd Qu.:20.00   
## Max. :56.50 Max. :35.90   
## PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA PERCENT.AGED.UNDER.18.OR.OVER.64  
## Min. : 2.50 Min. :13.50   
## 1st Qu.:11.80 1st Qu.:32.00   
## Median :18.50 Median :38.10   
## Mean :20.34 Mean :35.75   
## 3rd Qu.:26.60 3rd Qu.:40.50   
## Max. :54.80 Max. :51.50   
## PER.CAPITA.INCOME HARDSHIP.INDEX   
## Min. : 8201 Min. : 1.00   
## 1st Qu.:15754 1st Qu.:25.00   
## Median :21323 Median :50.00   
## Mean :25563 Mean :49.51   
## 3rd Qu.:28887 3rd Qu.:74.00   
## Max. :88669 Max. :98.00

## Results

Crime dataset was used to visualize the patterns between 2008 to 2012 i.e. by crime rate by day. Before plotting the crime distribution by day, we have to format the date column which consists of date with timestamp to date. The present data is shown as:

dff <- read.csv("crime data.csv")  
dff$date <- as.Date(dff$Date, format= "%m/%d/%Y")  
  
dff <- dff %>% group\_by(date) %>% summarise(Crime.Rate = n())  
  
## Crime distribution by day from 2008 – 2012  
## From the plot, we find that crimes are more on new year, the first day compared to other days in a year.   
plot(dff$date, dff$Crime.Rate, type = "l",xlab = "Date",ylab = "Crime Rate",main = "Rate of Crime Distribution by Day")



Linear regression model:

##   
## Call:  
## lm(formula = Crime.Rate ~ PERCENT.OF.HOUSING.CROWDED + PERCENT.HOUSEHOLDS.BELOW.POVERTY +   
## PERCENT.AGED.16..UNEMPLOYED + PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA +   
## PERCENT.AGED.UNDER.18.OR.OVER.64 + PER.CAPITA.INCOME + HARDSHIP.INDEX,   
## data = DF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -34290 -11372 -1375 6927 82900   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) 61481.6457 28317.9313 2.171  
## PERCENT.OF.HOUSING.CROWDED 1709.6710 1335.3852 1.280  
## PERCENT.HOUSEHOLDS.BELOW.POVERTY -720.0732 518.7500 -1.388  
## PERCENT.AGED.16..UNEMPLOYED 774.5611 765.2996 1.012  
## PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA -1514.3973 651.7269 -2.324  
## PERCENT.AGED.UNDER.18.OR.OVER.64 -1940.8404 605.9423 -3.203  
## PER.CAPITA.INCOME 0.2760 0.2994 0.922  
## HARDSHIP.INDEX 1037.8876 516.4444 2.010  
## Pr(>|t|)   
## (Intercept) 0.03336 \*   
## PERCENT.OF.HOUSING.CROWDED 0.20473   
## PERCENT.HOUSEHOLDS.BELOW.POVERTY 0.16957   
## PERCENT.AGED.16..UNEMPLOYED 0.31503   
## PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA 0.02310 \*   
## PERCENT.AGED.UNDER.18.OR.OVER.64 0.00206 \*\*  
## PER.CAPITA.INCOME 0.35984   
## HARDSHIP.INDEX 0.04838 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 18040 on 69 degrees of freedom  
## Multiple R-squared: 0.3026, Adjusted R-squared: 0.2318   
## F-statistic: 4.276 on 7 and 69 DF, p-value: 0.0005634

## (Intercept)   
## 61481.6457439   
## PERCENT.OF.HOUSING.CROWDED   
## 1709.6710409   
## PERCENT.HOUSEHOLDS.BELOW.POVERTY   
## -720.0731713   
## PERCENT.AGED.16..UNEMPLOYED   
## 774.5611167   
## PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA   
## -1514.3972716   
## PERCENT.AGED.UNDER.18.OR.OVER.64   
## -1940.8403590   
## PER.CAPITA.INCOME   
## 0.2759838   
## HARDSHIP.INDEX   
## 1037.8875960

##   
## Please cite as:

## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

##   
## <table style="text-align:center"><tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"></td><td><em>Dependent variable:</em></td></tr>  
## <tr><td></td><td colspan="1" style="border-bottom: 1px solid black"></td></tr>  
## <tr><td style="text-align:left"></td><td>Crime.Rate</td></tr>  
## <tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left">PERCENT.OF.HOUSING.CROWDED</td><td>1,709.671</td></tr>  
## <tr><td style="text-align:left"></td><td>(1,335.385)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td style="text-align:left">PERCENT.HOUSEHOLDS.BELOW.POVERTY</td><td>-720.073</td></tr>  
## <tr><td style="text-align:left"></td><td>(518.750)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td style="text-align:left">PERCENT.AGED.16..UNEMPLOYED</td><td>774.561</td></tr>  
## <tr><td style="text-align:left"></td><td>(765.300)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td style="text-align:left">PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA</td><td>-1,514.397<sup>\*\*</sup></td></tr>  
## <tr><td style="text-align:left"></td><td>(651.727)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td style="text-align:left">PERCENT.AGED.UNDER.18.OR.OVER.64</td><td>-1,940.840<sup>\*\*\*</sup></td></tr>  
## <tr><td style="text-align:left"></td><td>(605.942)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td style="text-align:left">PER.CAPITA.INCOME</td><td>0.276</td></tr>  
## <tr><td style="text-align:left"></td><td>(0.299)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td style="text-align:left">HARDSHIP.INDEX</td><td>1,037.888<sup>\*\*</sup></td></tr>  
## <tr><td style="text-align:left"></td><td>(516.444)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td style="text-align:left">Constant</td><td>61,481.650<sup>\*\*</sup></td></tr>  
## <tr><td style="text-align:left"></td><td>(28,317.930)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left">Observations</td><td>77</td></tr>  
## <tr><td style="text-align:left">R<sup>2</sup></td><td>0.303</td></tr>  
## <tr><td style="text-align:left">Adjusted R<sup>2</sup></td><td>0.232</td></tr>  
## <tr><td style="text-align:left">Residual Std. Error</td><td>18,038.130 (df = 69)</td></tr>  
## <tr><td style="text-align:left">F Statistic</td><td>4.276<sup>\*\*\*</sup> (df = 7; 69)</td></tr>  
## <tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"><em>Note:</em></td><td style="text-align:right"><sup>\*</sup>p<0.1; <sup>\*\*</sup>p<0.05; <sup>\*\*\*</sup>p<0.01</td></tr>  
## </table>

The percentage of those below 18 and over 64, the percentage of those 25 and older without a high school diploma, and the hardship index have significant p-values and are strongly correlated with the crime rate. R-squared values range from 0 to 1. This model’s R-squared value is 0.30, which indicates that it accounts for about 30% of the variation in the crime rate.

## Implications

Formula: Y= 1709.67 X1 – 720.07 X2 + 774.56 X3 -1514.4 X4 – 1940.84 X5 + .276 X6 + 1037.88 X7 + 61481.64. This can be used for future crime predictions. (Y=crime rate, X1=Percent of housing crowded, X2=Households below poverty, X3=Aged 16 and unemployed, X4=Aged 25 above and no high school diploma, X5=Aged 18under or above 64, X6=Per capita income, X7=Hardship index.)

When we take these three particular variables alone into consideration, the results will be:

##   
## Call:  
## lm(formula = Crime.Rate ~ HARDSHIP.INDEX + PERCENT.AGED.UNDER.18.OR.OVER.64 +   
## PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA, data = DF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -35329 -11640 -2885 7252 87057   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) 68992.2 12488.7 5.524  
## HARDSHIP.INDEX 759.9 163.4 4.650  
## PERCENT.AGED.UNDER.18.OR.OVER.64 -1774.2 418.3 -4.242  
## PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA -924.3 314.2 -2.942  
## Pr(>|t|)   
## (Intercept) 4.83e-07 \*\*\*  
## HARDSHIP.INDEX 1.44e-05 \*\*\*  
## PERCENT.AGED.UNDER.18.OR.OVER.64 6.42e-05 \*\*\*  
## PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA 0.00437 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 18190 on 73 degrees of freedom  
## Multiple R-squared: 0.25, Adjusted R-squared: 0.2192   
## F-statistic: 8.112 on 3 and 73 DF, p-value: 9.837e-05

##   
## <table style="text-align:center"><tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"></td><td><em>Dependent variable:</em></td></tr>  
## <tr><td></td><td colspan="1" style="border-bottom: 1px solid black"></td></tr>  
## <tr><td style="text-align:left"></td><td>Crime.Rate</td></tr>  
## <tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left">HARDSHIP.INDEX</td><td>759.876<sup>\*\*\*</sup></td></tr>  
## <tr><td style="text-align:left"></td><td>(163.415)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td style="text-align:left">PERCENT.AGED.UNDER.18.OR.OVER.64</td><td>-1,774.233<sup>\*\*\*</sup></td></tr>  
## <tr><td style="text-align:left"></td><td>(418.280)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td style="text-align:left">PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA</td><td>-924.306<sup>\*\*\*</sup></td></tr>  
## <tr><td style="text-align:left"></td><td>(314.166)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td style="text-align:left">Constant</td><td>68,992.170<sup>\*\*\*</sup></td></tr>  
## <tr><td style="text-align:left"></td><td>(12,488.740)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left">Observations</td><td>77</td></tr>  
## <tr><td style="text-align:left">R<sup>2</sup></td><td>0.250</td></tr>  
## <tr><td style="text-align:left">Adjusted R<sup>2</sup></td><td>0.219</td></tr>  
## <tr><td style="text-align:left">Residual Std. Error</td><td>18,185.530 (df = 73)</td></tr>  
## <tr><td style="text-align:left">F Statistic</td><td>8.112<sup>\*\*\*</sup> (df = 3; 73)</td></tr>  
## <tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"><em>Note:</em></td><td style="text-align:right"><sup>\*</sup>p<0.1; <sup>\*\*</sup>p<0.05; <sup>\*\*\*</sup>p<0.01</td></tr>  
## </table>

From the output, we can see that 25% of the crimes are related to hardship index, meaning people aged under 18 or over 64 and those aged 25 without a high school diploma. Simple linear regression was carried out and results are shown below:

#Analyzing each independent variable

##   
## Call:  
## lm(formula = Crime.Rate ~ PERCENT.OF.HOUSING.CROWDED, data = DF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -23000 -14386 -6972 8784 96296   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 20658.2 3921.6 5.268 1.28e-06 \*\*\*  
## PERCENT.OF.HOUSING.CROWDED 757.2 639.3 1.184 0.24   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20530 on 75 degrees of freedom  
## Multiple R-squared: 0.01836, Adjusted R-squared: 0.005271   
## F-statistic: 1.403 on 1 and 75 DF, p-value: 0.24

##   
## Call:  
## lm(formula = Crime.Rate ~ PERCENT.HOUSEHOLDS.BELOW.POVERTY, data = DF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -35505 -12217 -5857 9675 93711   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 12833.4 4840.2 2.651 0.00978 \*\*  
## PERCENT.HOUSEHOLDS.BELOW.POVERTY 530.8 196.8 2.697 0.00863 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 19780 on 75 degrees of freedom  
## Multiple R-squared: 0.08842, Adjusted R-squared: 0.07626   
## F-statistic: 7.275 on 1 and 75 DF, p-value: 0.008632

##   
## Call:  
## lm(formula = Crime.Rate ~ PERCENT.AGED.16..UNEMPLOYED, data = DF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -30673 -13285 -6146 9473 92469   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 14029.6 5220.9 2.687 0.00887 \*\*  
## PERCENT.AGED.16..UNEMPLOYED 673.7 305.3 2.207 0.03039 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20080 on 75 degrees of freedom  
## Multiple R-squared: 0.06097, Adjusted R-squared: 0.04845   
## F-statistic: 4.87 on 1 and 75 DF, p-value: 0.03039

##   
## Call:  
## lm(formula = Crime.Rate ~ PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA,   
## data = DF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -21469 -15365 -7923 10678 96908   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) 22231.1 4712.6 4.717  
## PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA 105.9 200.6 0.528  
## Pr(>|t|)   
## (Intercept) 1.08e-05 \*\*\*  
## PERCENT.AGED.25..WITHOUT.HIGH.SCHOOL.DIPLOMA 0.599   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20680 on 75 degrees of freedom  
## Multiple R-squared: 0.003704, Adjusted R-squared: -0.00958   
## F-statistic: 0.2788 on 1 and 75 DF, p-value: 0.599

##   
## Call:  
## lm(formula = Crime.Rate ~ PERCENT.AGED.UNDER.18.OR.OVER.64, data = DF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -22677 -16013 -7450 11367 98045   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 36132.4 11749.6 3.075 0.00293 \*\*  
## PERCENT.AGED.UNDER.18.OR.OVER.64 -328.6 322.1 -1.020 0.31089   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20570 on 75 degrees of freedom  
## Multiple R-squared: 0.01369, Adjusted R-squared: 0.0005378   
## F-statistic: 1.041 on 1 and 75 DF, p-value: 0.3109

##   
## Call:  
## lm(formula = Crime.Rate ~ PER.CAPITA.INCOME, data = DF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -22366 -14735 -6513 11245 97235   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.466e+04 4.621e+03 5.336 9.71e-07 \*\*\*  
## PER.CAPITA.INCOME -1.067e-02 1.554e-01 -0.069 0.945   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20720 on 75 degrees of freedom  
## Multiple R-squared: 6.288e-05, Adjusted R-squared: -0.01327   
## F-statistic: 0.004716 on 1 and 75 DF, p-value: 0.9454

##   
## Call:  
## lm(formula = Crime.Rate ~ HARDSHIP.INDEX, data = DF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -25742 -13529 -7157 9247 94022   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 17398.99 4639.21 3.750 0.000345 \*\*\*  
## HARDSHIP.INDEX 141.14 81.21 1.738 0.086334 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20310 on 75 degrees of freedom  
## Multiple R-squared: 0.03871, Adjusted R-squared: 0.02589   
## F-statistic: 3.02 on 1 and 75 DF, p-value: 0.08633

## Conclusion

The findings show that more crimes occur on January 1st of every year compared to other days. The percentage of those aged 25 and older without a high school diploma, the percentage of those under 18 and over 64, and the level of hardship all have an impact on the prevalence of criminal activity in Chicago, as shown in a linear regression analysis. Hence, government officials can make plans and take action to reduce crime in the future. Future studies can use a more recent set of socioeconomic indicators with crime rates, as this study only used data from 2008 to 2012, which is available in the official site.

# References

1.Gahalot, A., Prathiba kumari., & Dhiman.S., (2020, February). Crime prediction and analysis. In 2nd International Conference on Data, Engineering and Applications (IDEA)

2.Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017, May). An overview on crime prediction methods. In 2017 6th ICT International Student Project Conference (ICT-ISPC) (pp. 1-5). IEEE.20

3.Nwankwo, C. S., Raji, M. K., & Oghogho, E. S. (2018). Application of data analytics techniques in analyzing crimes. Tech. Rep.

4.Sathyadevan, S., Devan, M. S., & Gangadharan, S. S. (2014, August). Crime analysis and prediction using data mining. In 2014 First international conference on networks & soft computing (ICNSC2014) (pp. 406-412). IEEE.

5.Gorr, W., & Harries, R. (2003). Introduction to crime forecasting. International Journal of Forecasting, 19(4), 551-555.

6.Peiris M.D.P.T, Shaminda M.B.D, Senarathne H.K, Dr. Gayana Fernando (2020 December) Crime Prediction and Analysis System (International Journal of Scientific and Research Publications, 546 ISSN 2250-3153 )

7.ISR Data | Chicago Police Department

8.Census Data - Selected socioeconomic indicators in Chicago, 2008 – 2012 | City of Chicago | Data Portal 9.Crimes - 2001 to Present | City of Chicago | Data Portal