Predicting Crime Rates in the Metro U.S. Based on Crowding and its Implications on the U.S. Prison System



Andy Ninh

Spring 2014

**Abstract**

**Introduction**

A social problem that has received a large amount of attention for many years, and is arguably one of the most important in 2014, is crowding, or overpopulation, and its implied deleterious effects on quality of life. Many studies have been conducted for at least the past 40 years, in all aspects of crowding and its effects, both physical and psychological. With regard to the law, one of the biggest areas seeing concerns with crowding is the U.S. prison system. Overcrowding in prisons causes many problems, which is why it is a serious concern. Once the causes of crowding have been established, researchers can begin to address the problems it causes and deal with them.

In the current study, a dataset comparing population density and crimes in metro cities in the U.S. from 1970 was used to complete a hierarchical clustering analysis in R. As expected, the analytical prediction method accurately clustered cities with similar population densities and relation to crime rates, supporting the hypothesis that crowding will increase the rate of crime. This is incredibly important because over the years, scientists have found confusing results on the topic of crowding, and crowding is an important issue with regard to the public at large, but specifically the prison system. As prisons across the country continue to increase in population, more violent crimes will be committed, making order difficult to maintain. This will completely defeat the purpose of the system. Let this finding be a warning to the justice system.

**Research Problem**

Understanding where crime happens can be a key to understanding why it happens. Models that predict the occurrence of crime by geographical area often use data on the characteristics of the inhabitants (e.g. income, race, family structure) of that area. However, population density has also received considerable attention as it relates to crime. Jane Jacobs, in 1961, contradicted the popular wisdom of city planners with her claim that crowded city streets and sidewalks could be effective deterrents to criminal behavior. A number of national studies tested the relationships between density and crime, with differing results. Some studies found positively correlated relationships between crime and density, while others found the opposite types of relationships. Still, others (e.g. Freedman, 1975) found non-significant relationships between the two variables.[[1]](#footnote-1) So, which is it?

It would seem intuitively that cramming people into a fixed area would not only statistically improve the possibility of increased crime, but also increase the likelihood due to negative effects on the psychology of the situation.

*Hypothesis 1:* Population density, or crowding, will be positively related to crime rate.

Also, high population density usually relates to a high volume of traditionally nonwhite or traditionally poorer neighborhoods. This would lead one to think that metro U.S. cities with high percentages of nonwhite people will have higher crime rates.

*Hypothesis 2:* The percentage of nonwhite persons in a population will positively affect the crime rate.

Although the many factors that might play into this hypothesis are outside the scope of this paper, it still important to assess this factor in the analysis. There is a general understanding, and proven track record, that cities like Detroit have one of the highest crime rates in the country, and it also has a high concentration of nonwhite population. Although this data originates from the 1970 U.S. census, it is likely that population data has not varied significantly for the major cities assessed as most major U.S. metros have maintained their appeal.

**Results + Methods**

The data collected for this study was collected in 1970 by J. Freedman, a famous psychologist who studied the effects of crowding on behavior.[[2]](#footnote-2) It was slightly modified from the original, and is missing some data, but this does not detract from the conclusions. It represents 100 U.S. Metro cities, defined as having a population greater than 250,000, and contained 4 variables: (1) Total Population in 1968 in thousands, (2) Percent of Nonwhite persons in the population in 1960, (3) Density, or Population per square mile in 1968, and (4) Crime rate per 100,000 in 1969. In order to assess the effect of density and crime, this study used an analytical prediction method known as hierarchical clustering in a program called RStudio, which uses a program language called R.

The variables assessed were measured on different scales, so the first step was to standardize the data. In order to do this, the “scale” function was used. The data frame passed through “scale,” along with two vectors. The first is called “center,” which is a vector of values, one for each column of the data frame to be standardized, which was subtracted from every entry in that column. The second, called “scale,” is similar to “center,” but is used to divide the values in each column. These vectors are created with the “apply” function that performs the same operation on each row or column of the data frame. The city names are not useful to portion of the analysis so they were removed. Notice that the “scale” function does not change the order of the rows of the data frame, so it was easy to identify observations using the omitted column from the original data. Refer to the Appendix for output of the Hierarchical Clustering.

The first step of the hierarchical method is to calculate a distance matrix. For a data set with n observations, the distance matrix will have n rows and n columns; the (i,j)th element of the distance matrix is the difference between observation i and observation j. In this experiment, the default of Euclidean distance was used, but it is possible to use other metrics.

**Discussion**

**Conclusion**

It is clear that crowding impacts crime and although our experiment did not yield perfectly significant results, it remains a powerful theory. This affects the U.S. prison system in a truly negative way, and unless a solution to the problem is created, it will fail.

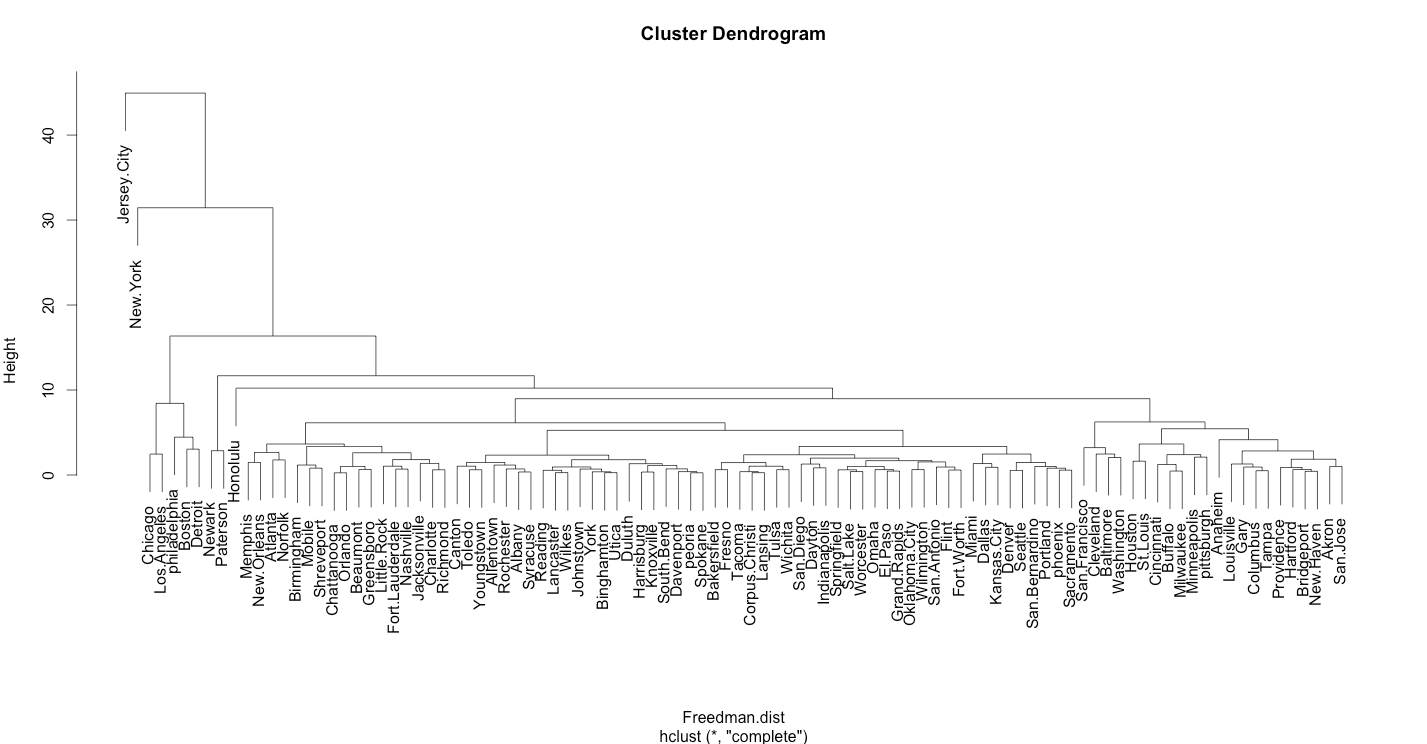
**Appendix**

*R Output for Paper*:

All output from R is beautifully displayed on my GitHub Account. A link to the R Markdown page is found here:

<https://github.com/A-Ninhja/Legal_Analytics_Spring_2014/blob/master/Final_Project/Crowding.md>

*The Dendogram Built from Hierarchical Clustering*



Click [here](https://raw.githubusercontent.com/A-Ninhja/Legal_Analytics_Spring_2014/master/Final_Project/Cluster.png) to view an enlarged version of the dendogram.

*R Output for Regression Analysis*

> reg\_1 <- lm(crime ~ density + nonwhite + population, data=Freedman)

> summary(reg\_1)

Call:

lm(formula = crime ~ density + nonwhite + population, data = Freedman)

Residuals:

Min 1Q Median 3Q Max

-2003.18 -657.16 70.94 603.73 2213.20

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2193.70088 143.04566 15.336 < 2e-16 \*\*\*

density -0.02145 0.06578 -0.326 0.745045

nonwhite 26.03770 8.76746 2.970 0.003764 \*\*

population 0.24495 0.06095 4.019 0.000116 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 877.2 on 96 degrees of freedom

Multiple R-squared: 0.2288, Adjusted R-squared: 0.2047

F-statistic: 9.493 on 3 and 96 DF, p-value: 1.496e-05

1. *See* Brian Christens & Paul W. Speer, *Prediciting Violent Crime Using Urban and Suburban Densities*, Behavior and Social Issues, 14, 113-127 (2005). [↑](#footnote-ref-1)
2. United States (1970), *Statistical Abstract of the United States*, Bureau of the Census. [↑](#footnote-ref-2)