WV Cities Crime Rate Analysis

12.1.2020

Benjamin Slates, Karen Truong, Aaron Tran, Alec McAtee,

Jeffrey Petty

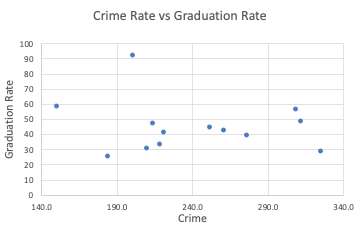
**Introduction**

Morgantown, West Virginia is a college town bustling with educational opportunities and facilities. One peculiar thing about such a town, which contains one of the most successful colleges and the most successful hospital in West Virginia is that it simultaneously contains one of the highest violent crime rates of the state. Because of this peculiarity, this study was conducted with the purpose of finding correlations between certain variables and the effects it had on violent crime rates as a means to provide insight into the common issue society is faced with every day. Perhaps through this study we would be able to find an answer or be led closer to one.

Using nearby towns in West Virginia for comparison to Morgantown, the data was analyzed based on certain variables to determine what factors may influence crime rate. In addition to this, Universities in West Virginia were utilized in order to analyze the impact of education on crime rates. In this study, it is hypothesized that violent crime rates are affected by several factors including diversity, average income, and education. Specifically, it is hypothesized that lower incomes and lower education levels, and higher levels of diversity will be correlated with higher rates of convictions.

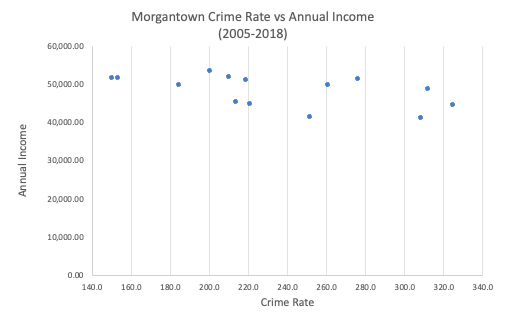
**Data**

For this study, data was collected for three variables to use as comparison to crime: education, diversity, and income. The collected data used in this study can be found in Appendix A. First, to determine whether education had an effect on violent crime, the graduation rates of each college in WV were compiled, along with the cities that house them as a means of comparison. In the first graph, a scatter plot was used to determine a correlation between crime rates and graduation rates. There was a very wide range of graduations in the sample ranging from 25% all the way to 91%. There turned out to be a very weak correlation between the 2 variables with an r squared value of 0.03153.



Second, to determine the correlation between income and crime, data was collected from datausa.io listing annual income estimates from Morgantown WV and the towns located nearby which include Wheeling, Bridgeport, and Fairmont. These annual income estimates were collected from each town through the years of 2005-2018, although some towns did not have observations for every year.

The scatterplot below shows annual income against annual crime rate (per 100,000 residents) for Morgantown between the years of 2005 and 2015. There was a wide distribution of crime rates ranging from 150.4% to 325.1%, and a range of annual incomes from $40,948 to $52,396. It can also be observed that there is an apparent linear relationship between an increase in annual income and a decrease in crime rate.



Lastly, to observe the variable of diversity, the diversity rates of each college were collected. In order to analyze the effects of diversity, dummy variables were utilized with diverse being 0 and non-diverse colleges being 1. To determine diversity, those colleges with 10% or more students recorded as being non-white were considered to be diverse and those with less than 10% were considered to be non-diverse.

**Methods**

Two models were created in order to fully approach the intents and purposes of this study.

Model 1 which showed the graduation rates, diversity, and crime rates, the method of choice used to fully analyze the data in this study was a multiple linear regression. Given the data, this choice was appropriate as there were multiple variables that needed to be evaluated in order to come to a conclusion. Graduation rates, and the level of diversity were used as predictor variables for the response variable (y variable) of crime rates. In other words, the goal was to predict violent crime rates through two factors: graduation rates and diversity.

Model 2 was created in order to present and analyze the relationship between income and crime rates. This model shows the annual average income, and crime rates in correspondence to the year and town. However, the original data that was collected for this model contained several missing variables. In order to complete it, a regression analysis was used in order to impute missing data into the model. The year and crime rate were used as predictors in determining the unavailable data. Through this it was possible to preserve the model and prevent losing several lines of important data. This model was analyzed through a simple linear regression. That is because, for the purposes of this study, it is important that the crime rate (response variable) is only being compared to the annual income.

**Results**

Shown in Appendix B are the following results produced from the data previously collected. Using a multiple linear regression on Model 1, it was found that graduation had a negative effect on crime rates with graduations. As shown in the first figure, graduation displayed a coefficient of -0.83. In other words, as graduation rates got higher, crime rates got lower. However, the P-value for this coefficient was rather high with a result of 0.86, meaning that the result was considerably statistically insignificant. As for diversity, it is shown to have a coefficient of 199.05, meaning that colleges which showed a lack of diversity demonstrated a positive relationship with crime. This however, is also statistically insignificant, as the correlating P-value was a 0.31. In terms of the R Squared value, the regression itself did not describe the variations in the data very well; the combination of a low R squared value and high P-values for both variables showed that the data altogether was not significant in explaining crime rates. As a result, it was concluded that diversity and race as well as college graduation rates had little to do with violent crime rates through the sample of data that was collected.

Through a simple linear regression on Model 2, it was found that the average annual income had a negative relationship with crime rate. As shown in the second figure, the average annual income displayed a coefficient of -0.0028. This shows that income had a slight negative relationship with crime; as income decreased, crime rates increased. Furthermore, the P-value for this variable was shown to be very statistically significant, with it being a small number of 0.00059. Although this model demonstrated an insignificant R squared value, the focus was to examine a correlation between income and crime rates. As such, the result was that income did indeed display some level of correlation with crime.

**Conclusions**

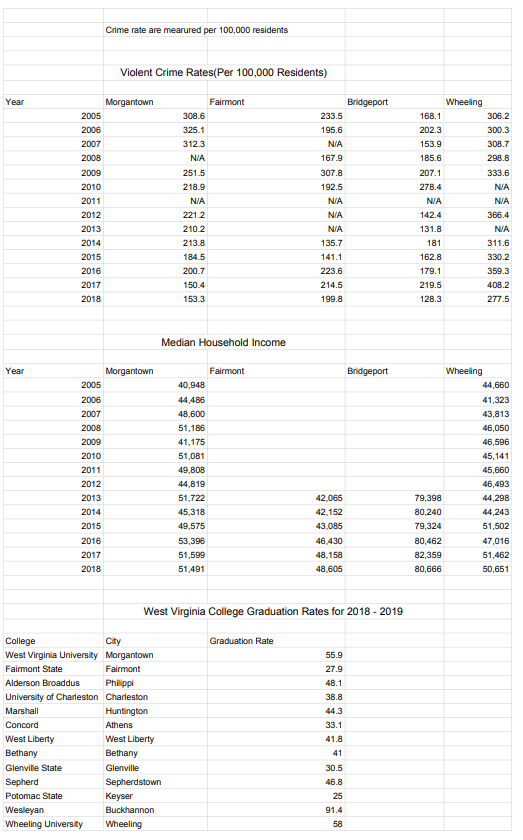
It was hypothesized in this study that lower income and education levels and higher diversity would lead to higher crime rates in Morgantown, WV. This study was conducted to further elucidate why this town has one of the highest crime rates in the state of WV despite certain advancements; those advancements being the presence of multiple high-quality hospitals, West Virginia University, and multiple other factors that increase Morgantown’s standard of living. In order to determine if this hypothesis was valid, 2 regression models were evaluated. This prediction was proven to be partially true based on the results of the models.

As shown in the results, Model 1 tested the relationship between education and diversity on the level of crime using a multiple linear regression. Unexpectedly, a relationship was not determined to exist among these factors. This model was unable to uphold the portion of the hypothesis stating that lower college graduation rates and higher diversity lead to higher crime rates. In comparison to this, Model 2, which demonstrated a simple linear regression showed that average annual income had a negative relationship with crime rates. Unlike Model 1, it was able to successfully support the hypothesis. Since Model 1 failed while Model 2 was successful, the original hypothesis was proven partially supportable.

In order to further determine the important variables impacting crime rates, future studies should be conducted. One change that could be made to ensure a more productive model is designed is to gather more data on the current variables. This study utilizes data from three sources over the course of 14 years, and the data was rather incomplete. It may be beneficial to obtain more complete data over a longer period of time. Another change that could be made is to alter model one to analyze the effect of high school graduation data rather than college completion. Since college graduation rates are expected to be lower, in general, than high school graduation rates, it may be true that there would be more variation in the education data. While the study was not completely successful in confirming important variables affecting crime in Morgantown, WV, the findings of Model 2 cannot be overlooked. This project brought to light the existence of a relationship between income and crime, and further studies should be conducted to enhance these findings.

**Appendix A**

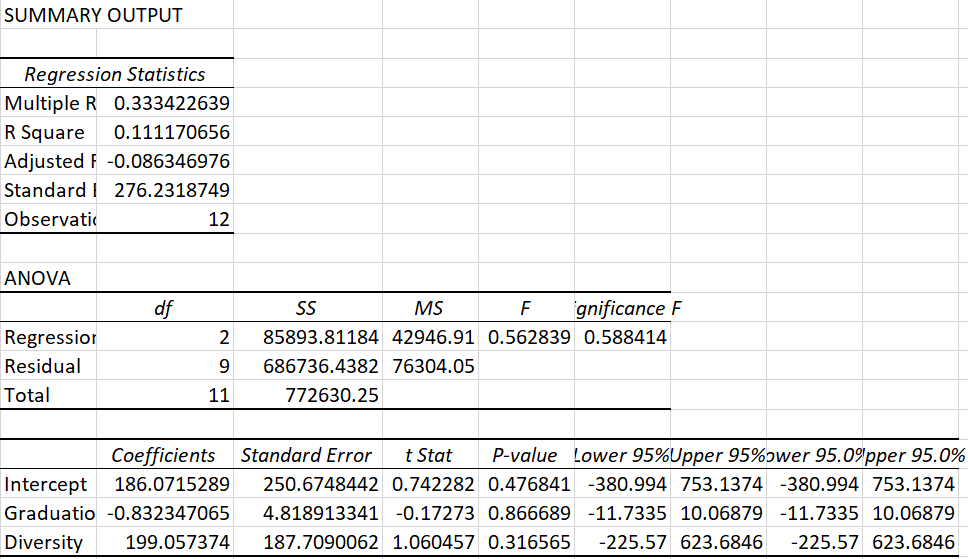
The data compiled and used for the purpose of this study included the variables of education, diversity, income, and crime rates. These data were obtained from the following sites: neighborhoodscout.com, disastercenter.com, and Data USA. Below are spreadsheets of data that were collected and used in this study. You will also find the results from the regression analyses performed for Model 1 and Model 2.



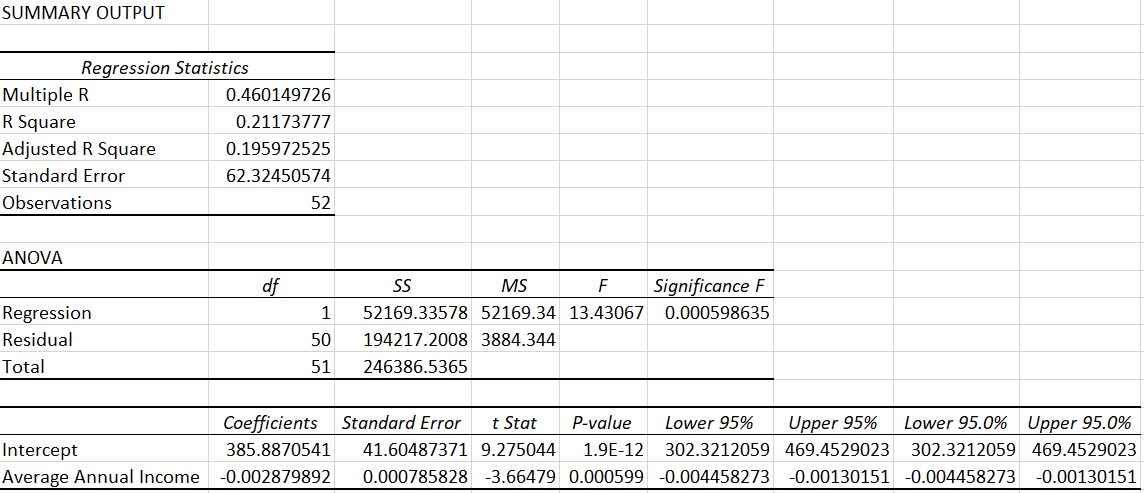
**Appendix B**

Shown below are the full results of the regression analysis completed for Model 1 and Model 2.

*Model 1: Graduation and Diversity on Crime*

**

*Model 2: Annual Income on Crime*

**