

Violent Crime Prevention Through Public Funds: Analyzing Education and Poverty's Impact on Violent Crime Rates in Pennsylvania

4.2.a What is the source of your data? Include any links or references to the data source.

I am using education and poverty data that has been collected by the [US Census](#). I am also utilizing [Pennsylvania Uniform Crime Reporting System](#) to collect the total of violent crimes per county.

4.2.b What is your intended goal in analyzing this data set? Are you extending previous analysis or starting new analysis?

I am starting a new analysis that will analyze the impact education and poverty have on violent crimes across counties in PA. My intended goal in analyzing this data set is to see if public funds can be better utilized in preventing violent crimes across PA.

4.2.c What steps have you taken so far to get the dataset to its state for 4.1 above? What problems did you encounter?

I first had to compile the data into one manageable Excel spreadsheet. This was a very time-consuming process. To collect the information from the US census, there was not a simple way to output all 67 counties into one Excel spreadsheet. I needed to output the data 6 counties at a time. This resulted in me having a good number of spreadsheets that I needed to consolidate into one and then clean out data that I did not plan on utilizing. The Pennsylvania Uniform Crime Reporting System made it difficult to collect data in one excel spreadsheet. This again resulted in multiple spreadsheets (sixty-seven to be exact) that I needed to compile and remove unnecessary data. To ensure I collected all sixty-seven counties, every spreadsheet downloaded from the Pennsylvania Uniform Crime Reporting System I had to rename in order to ensure I would not miss one or duplicate multiple spreadsheets due to the name of the downloaded files being exactly the same. I took this extra step to ensure accurate reporting.

4.2.d How could the results of the analysis be used in a commercial or research setting?

It could be used to show public funds might be better put to education or funding welfare programs to reduce violent crimes across Pennsylvania.

4.2.e What machine learning / modeling techniques do you anticipate using?

Nominal classification, numeric estimation, other?

How do(es) the planned modeling technique(s) relate to 4.2.d?

I plan on using regression models on the data that has been collected. This will allow me to interpret the relationship between education/poverty and crime rates.

4.2.f Document any other aspect of the project that you feel is important to communicate.

It is important to note a few things with the data that has been collected. Violent crimes can be interpreted as a number of crimes depending on the person. The data that has been collected is based on violent crimes such as Murder/Nonnegligent Homicide, Rape, Robbery, and Aggravated Assault. The selected crimes are what Pennsylvania Uniform Crime Reporting Systems lists as “violent crimes”. The data that was available for High school graduate or higher is conducted every year, so the available percentage is based on five years of data. The “persons in poverty” percentage is based off of a five-year estimate. Both data sets range between the years of 2017 – 2021. In order to continue with the same years calculated – the data provided for violent crimes is also between that year range. The violent crimes data also is the number based on 100,000 people. This number was founded by taking the numbers of total crimes dividing by the population estimates from 2021 and multiplying by 100,000.

5.2.a What additional data did you collect during analysis, if any? Include any links or references to the data source.

I did not collect any additional data for analysis. If this project was a few more weeks, it would have been interesting to explore how public funds are currently being spent.

5.2.b Did you achieve your intended goal in analyzing this data set? Explain how analysis shows goal achievement or refutation.

I feel as though I did achieve my intended goal. I wanted to show how education and poverty may play a role in violent crime. Although education did not work well for what I was trying to prove, poverty kept and proved that poverty can/does play a role in violent crime per my model that was presented.

5.2.c What machine learning/modeling steps have you taken?

I utilized linear regression to achieve my intended goal. I did not run into really any issues that I think would affect my overall goal. I do think there was a lot we learned in class that could have been utilized to further my analysis. I do wish I had more time to collect data and act on it as this was a very interesting project.

5.2.d Use SMO, or SMOreg, or MultiLayerPerceptron, or clustering, OR at least one other technique not used in assignments 1-3.

I performed a clustering on my data for persons in poverty and violent crimes total. I am unsure if I am misunderstanding on to interpret the data but it doesn't seem to provide me any concrete information from my overall goal of the assignment.

5.2.e Revise explaining how could the results of the analysis be used in a commercial or research setting?

Prior in my other answers for 4.2.d I mentioned that the results could be used in order to better spend public funding. Although I still believe that to be true, I think it is important to also mention that this data could be used in a much smaller scale, to advocate for donations or volunteers to better a community.

5.2.f Document any other aspect of the project that you feel is important to communicate.

There are no other aspects of my project that I feel is important to communicate. Thank you!