Analysis of DC Crime by Income Level

Taylor Poole, Heather Lampkin, Pasha Frederick

December 12, 2018

1 Introduction

Geographic Information Systems (GIS) can help users conduct spatial analysis [2] and answer questions that lead a well-rounded understanding of problem sets. Having better knowledge of a problem can help answer questions that allow for better application of resources. One application of GIS is crime analysis. For the purpose of the final project, we used ArcMap, Excel and R Studio to help understand where crime is committed in Washington District of Columbia (D.C.) and to see if it correlates with low-income neighborhoods. The national poverty level falls within incomes that fall at $19,000 or less annually. If crime did not correlate with neighborhoods that fell below the poverty level, we used ArcMap to help understand where the majority of crime was happening.

Crime analysis is important as it aids with the implementation of policy to better focus law enforcement efforts. The utilization of GIS tools can assist by enriching the application of policing methodologies to reduce the overall crime. This type of analysis can educate the public with a visual representation to help citizens clarify crime concerns and procure community involvement. The end state is to understand where crime is happening to better protect the citizens of Washington D.C. and prevent crime before it happens.

2 Objective and Research Questions

The objective of this research paper is to use the skills learned in GIS 311, to skillfully answer the following questions. Is there a higher amount of crime in neighborhoods with and annual median income rate of $19,000 or less? If the crime does not correlate with neighborhoods that fall below the poverty level, where does it occur? Lastly, Is it safe to just assume nothing bad will happen because it seems like a good neighborhood?

3 Methods

3.1 Excel

The 2016 crime data was obtained through the Washington D.C [1]. Metro Police Department website and downloaded a .csv file. The data was organized via Microsoft Excel to remove unnecessary data and kept all relevant columns along with potential primary key columns. This was done so we could link the data with other datasets required for analysis. The .csv file was imported into ArcMap. The table contained Latitude and Longitude info so the XY Event Layer tool was

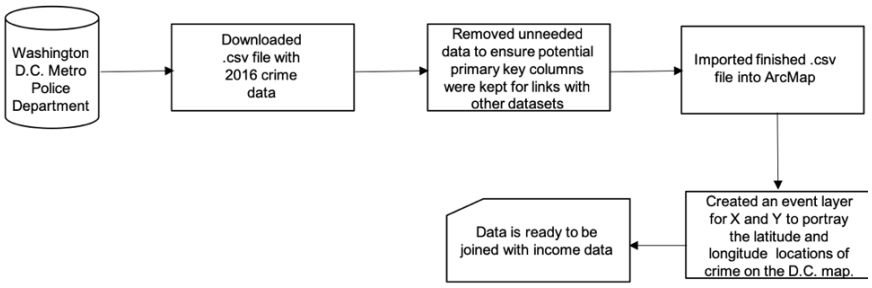

Figure 1: ACS Census Data using Excel

used in ArcMap to create a point map. This was accomplished by right-clicking the x,y layer name and click Data - Export Data. The Export Data dialog box then opens. Next, I set the output coordinate system which was NAD 1983 (2011) State Plane Virginia North FIPS 4501 (US Feet) and specified the location and name of the new feature class, then click OK to save. At this point the data was ready to merge with the other data sets.

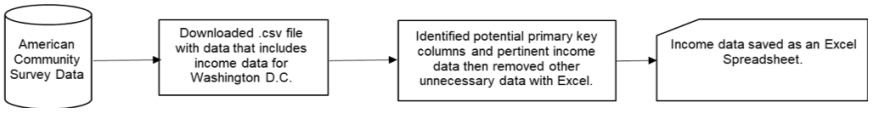

Figure 2: ACS Census Data using Excel

Information for the income data for Washington D.C. was obtained from the American Community Survey [4] as a data file named ACS_16_5YR_S1903. The information from this file was in a spreadsheet format and Excel was used to view the data.

The fields included information that was not going to be used for this project. All fields were removed in Excel except for the Household Median Income and the Census Tract listing fields. All text was also removed from the census tract data by using find and replace in Excel so that the census tract numbers were the only thing displayed in the Census Tract field. The income data was saved as an Excel spreadsheet.

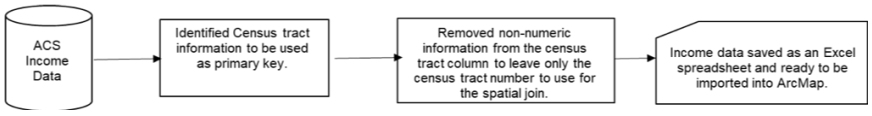

Figure 3: ACS Census Data using Excel and ArcMap

3.2 ArcMap

The census tract shapefile for Washington D.C. was obtained from the Census Bureau [3]. The data in this file included an attribute table with a census tract column that matched with our income data. All of the census tracts that were a part of Washington D.C. were selected and exported to create a data layer that was used for the base map of the city.

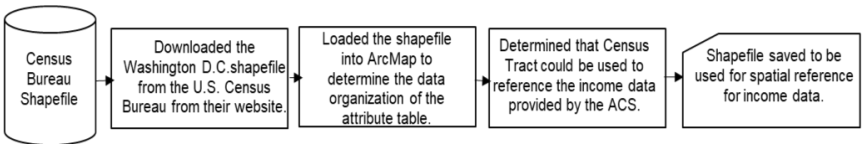

Figure 4: Census Tract Shapefile and Attribute Table in ArcMap

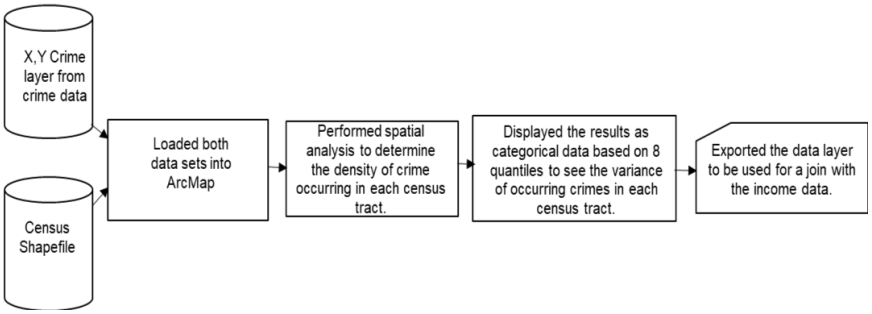

Figure 5: Spatial Analysis for Crime Data with ArcMap

The X,Y crime layer that was created from the latitude and longitude information in the crime data obtained from the MPDC, was loaded into Arcmap as a data layer in the same data frame as the census shapefile. Another data layer was then created by using the Point density (Spatial Analyst) tool in Arcmap and selecting all points that intersect the census tracts. The point density layer was then loaded into the data frame and display options were set to display the number of crimes in each census tract as a color ramp with eight quantiles. This helped to visually display the range of crime incidents within the city over the course of a year.

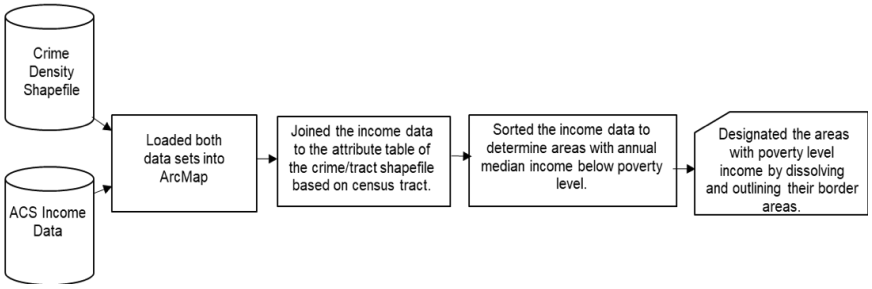

Figure 6: Spatial Join for Income Data with ArcMap

The Excel spreadsheet for the income data was loaded into ArcMap by using the Excel to Table (Conversion) tool. The format for the Household Median Income field was in a format that wasnt recognized in ArcMap so a new column was made and the information was copied into the new column using the data calculator. Spatial reference was made through the Point density layer attribute table by using the key field of census tract. The census tract shapefile displayed block tract data that was not on the income data set. After joining the income and crime density data in the attribute table, a dissolve was performed from the geoprocessing menu in Arcmap. The dissolve was based on the two attributes of census tract and Household median income in the attribute table.

This dissolve removed the block lines from the shapefile and kept the remaining census tracts that correlated with the income data set. The values from the Household median income that were greater than zero and less than 19,000 were selected and the results were exported into a new data layer to show the outlines of the census tracts with poverty level income. The resulting map displays the frequency of crime in Washington D.C. by the census tract area and also outlines the census tracts with median income below the poverty level.

3.3 RStudio IDE

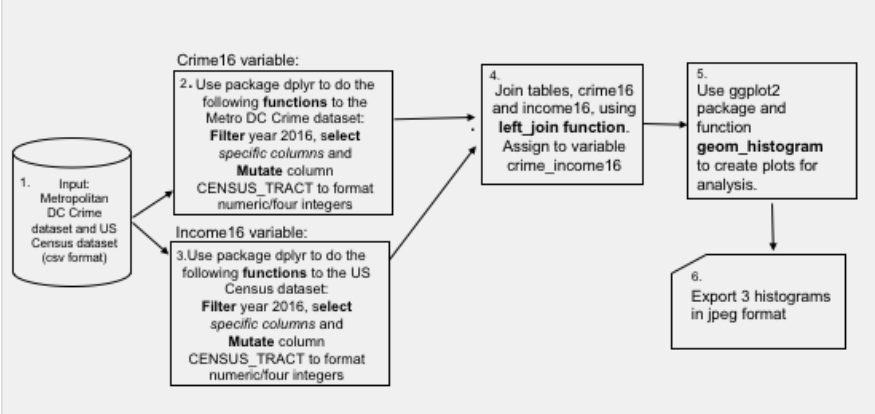

Figure 7: Data Analysis using RStudio

R studio Integrated development environment (IDE) is used for further analysis of crime in the District of Columbia. R is a programming language used primarily for statistics and graphics production. The IDE is fairly simple that allows the analysis to be performed quickly. **??**, shows the work flow of how Rstudio is used for the analysis of crime data in DC for the year of 2016.

The cylinder in **??** represents the input to be used in Rstudio. These inputs are the Metropolitan DC Crime dataset and the US Census dataset. After importation of these two datasets, both data sets are tidied independently using the package dyplyr. Dyplyr is a package component of Tidyverse. Dyplyr is used for common data manipulation by using a set of functions that are easy to learn for the user. Starting with the DC crime dataset, first the dataset is filtered by year. The function used here is the filter() function. Inside the parentheses, year == 2016 is place to filter out all the other years in the dataset. Next, the select() function is used to only take the columns that are necessary for answering the research questions. The specific columns chosen for the DC crime data set are "LATITUDE", "LONGITUDE", "REPORT_DAT", "Offensegroup" and "CENSUS_TRACT". The last function used is mutate(), because "CENSUS_TRACT" is a common variable amongst the crime dataset and census dataset, however they are not the same format. The mutate function is used to compute a new column using the data that is imported. For the crime dataset, the two lines of code used are mutate(CENSUS_TRACT = substr(CENSUS_TRACT, start = 3, stop = 6)) mutate(CENSUS_TRACT = as.numeric(CENSUS_TRACT)). The first part of code is stating that there should be a new column made named CENSUS_TRACT (essentially the first Census tract

column is being over-written) that takes the four ending characters and cuts off the rest of the preliminary census tract variable. The second line of code makes the CENSUS_TRACT variable an integer rather than a character string. Finally, the manipulated data is put in a table named Crime16. As seen in box 3 of **??**, the same process is used for the US census dataset, as it was for the DC crime dataset. First, the filter function is used to only have data for the year 2016. Next the select function is used to select specific columns, only in this case the columns selected are CENSUS_TRACT, year and income. Then the mutate function is used to create a new column for the census tract data to be presented in the same format as it is formatted in the DC crime dataset. Lastly, the manipulated data is placed in a table named Income16.

With the new creation of the tables, crime16 and income16, the tables are ready to be joined by the variable CENSUS_TRACT. To join the tables in Rstudio, the function left_join(). By using this function, income16 table is added to the crime16 table to create one table. The resulting table is named crime_income16. The table, crime_income16, will be used to further analyze the crime data and produce plots as shown in box five of **??**. However, it is important to point out that at this point one can export the new table and use it in other applications.

After joining the tables, crime_income16 can now be used for further analysis. For the purpose of this research, plotting the data in a histogram is the method chosen to complete the analysis of the results. A package commonly used to create histograms in R is ggplot2, which is also in the Tidyverse collection. Within the package ggplot2, the function geom_histogram() is used to create a histogram of the data used for analysis. The last step, block 6 of **??**, refers to the figures below in the Results section of this research paper. When exporting analysiss in Rstudio, formarts vary. However, for this assignment jpeg format is the chosen format, as it is simpler to add to the presentation format.

4 Results

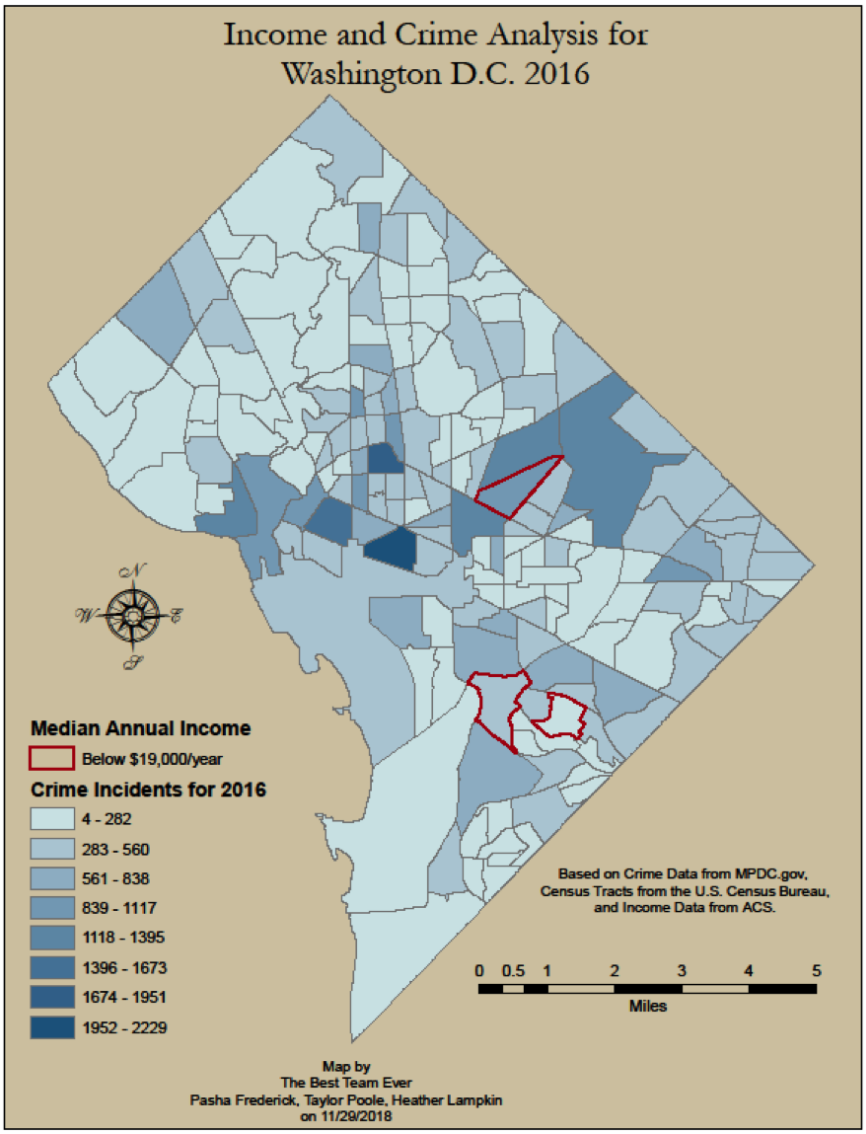

Figure 8: Map created using ArcGIS

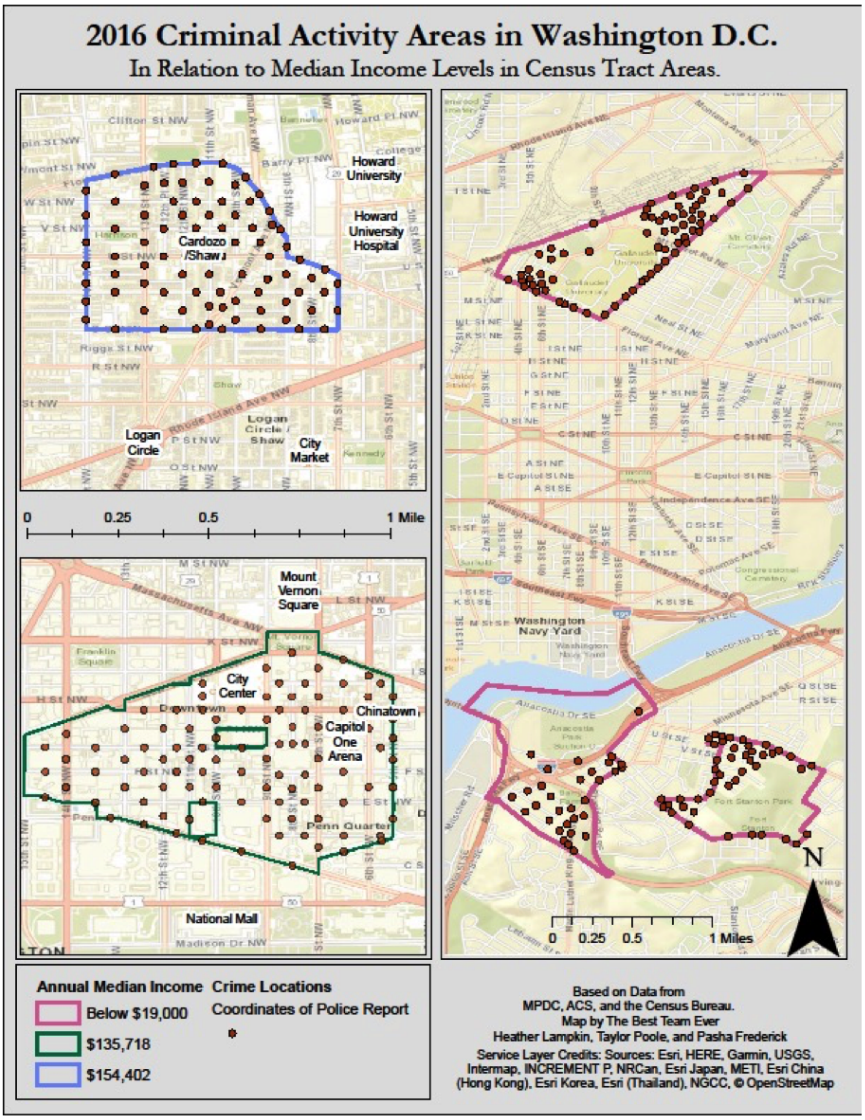

Figure 9: Map created using ArcGIS

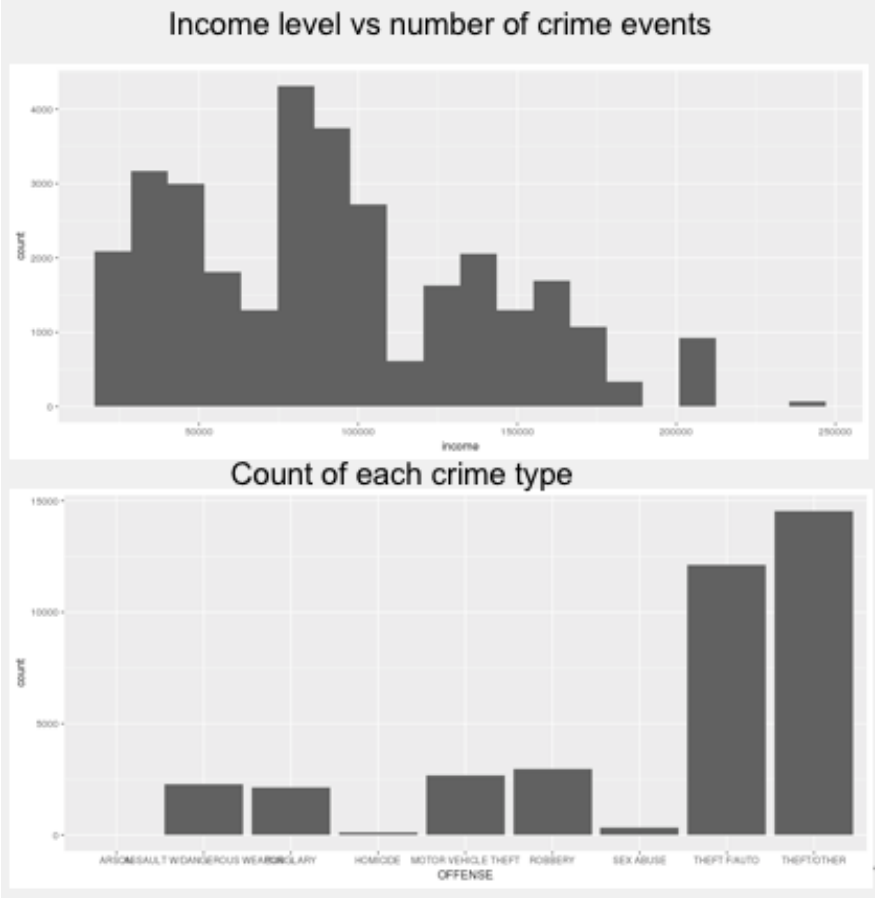

Figure 10: Crime Amount Based on Median Income

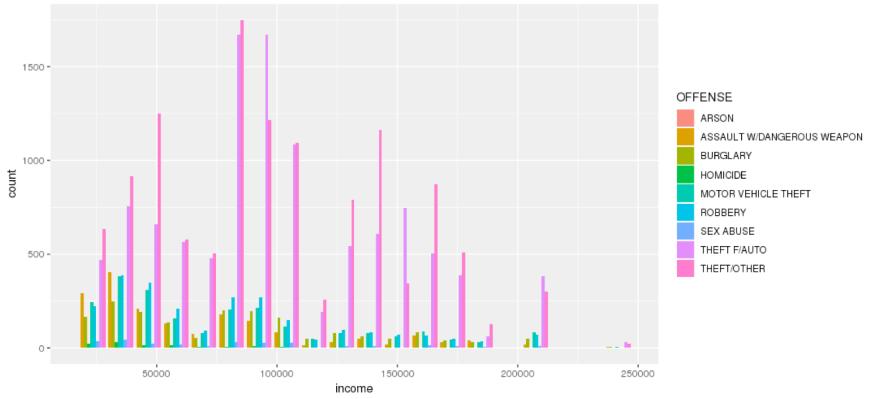

Figure 11: Crime Amount Based on Median Income, Separated by Offense Type

5 Conclusion and Further Research

5.1 Conclusion

The first question answered by the analysis conducted was if higher crime occurs in neighborhoods with an annual median income rate of $19,000 or less (poverty level). In 2016, there were 37,202 crimes reported in Washington D.C. and as seen in 8 the areas of income under $19,000 median income do not have the highest rates of crime. To answer the second question, If the crime does not correlate with neighborhoods that fall below the poverty level, where does it occur?, from 10 it concludes that high concentrated criminal activity appears to be located in neighborhoods with median incomes at about $75, 000. In the histogram, 11, shows the amount of crime occurrences by median income of the area where the crime occurs. In 11, we do see a correlation of more violent crimes occurring in areas with lower due to breaking the data up by offense type. Lastly, the third asks whether or not it is safe to assume and let ones guard down in a neighborhood that seems to be a good neighborhood. From the analysis done in this project, the answer would be no it is not a good idea to just assume crimes are unlikely to happen based on the population and median income.

5.2 Further Research

One way to take this research a step farther is to obtain data sets of DC crime and census data from multiple years in a row. The same processes from the Methodologies section of this report would be used for the new data sets. After completion of these process, analysis would take place on locating patterns of crime to determine if crime occurs in the same locations as they did in the 2016 dataset. After determining patterns of crime behavior over the years the data can be looked at by offense type and see if there are any hotspots of certain types of crime.

6 References

References

[1] Crime incidents in 2016.

[2] Paul Bolstad. *GIS Fundamentals: A First Text on Geographic Information Systems.*

[3] G. P. Branch. Tiger products.

[4] US Census Bureau. American community survey.