Ryan Ermovick – DS2023 Midterm Report

<u>Introduction and Context:</u>

The primary dataset used for this project was a dataset from the Murder Accountability Project. The Murder Accountability Project is a public project that documents murder in the U.S. from the mid-20th century onwards, gaining data partially from the FBI. The secondary dataset used in this project was the population of each state from the U.S. 2000 census. Both sources are reputable because they come from large government organizations, such as the FBI and the U.S. census. The objective of analyzing these datasets was to determine if the murder rate in the U.S. and New York specifically went down after the terrorist attacks of 9/11. This project is relevant because it can help determine the attitudes of a country after a major attack and how it affects society. Perhaps there are certain ways that society reacts in response to a national tragedy, and these findings could be useful in determining how to handle the aftermath of other tragedies.

The visuals used in this project are maps of the U.S., with each map representing a different year (2000, 2001, 2002). The color of each state represents the most popular murder weapon used for that year. The intensity (or the alpha level) of the shading for each state corresponds to the murder rate per person. The murder rate per person is the total number of murders per year per state divided by the population of that state. This is where the census data was used. These maps were useful for visualizing if the murder rate per person went down following the events of 9/11 because one can quickly view if the intensity of the shading for a specific state or the U.S. went down throughout the years. The color coding of each state was also key, as it allowed the viewer to determine if certain weapons were more popular near the site of the attack or if certain weapons declined in usage after the attack. Having the format of the graph be a map was necessary because it is a common visual that many people can understand and interpret. Also, the color scale for these graphs is the Viridis color scale, which is an easy-to-interpret scale for people who are colorblind.

The murder dataset included many categorical and numerical variables, such as victim age, victim race, offender age, offender rage, and relation between offender and victim. For the purposes of this project, the data that was used for visualization was the state, year, and weapon used. The dataset was transformed by grouping together weapons into distinct categories such as knife or firearm. The data was further transformed by grouping the murders by year and state. Then for each year and each state, the most common weapon used was found. The format of the U.S. census data document was manually transformed (deleting errant text) so the file could be loaded into a data frame. No changes
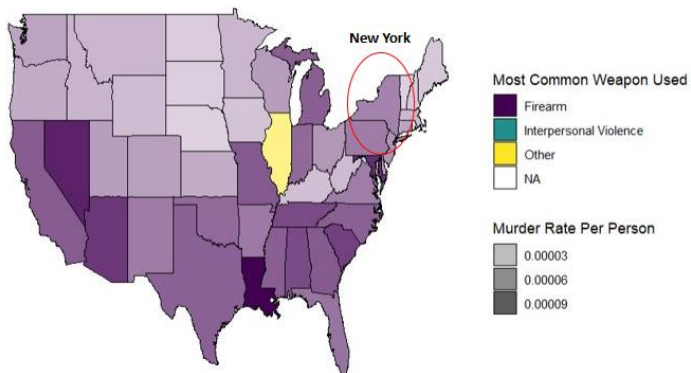
were made to this dataset. The murder rate per person for each state and year was found by dividing the number of murders for that year in that state by the population of that state.
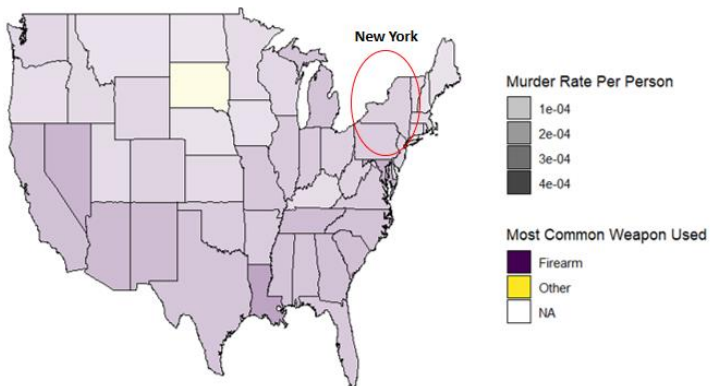
<u>Visuals:</u>

Murder Rate Per Person in the US: 2000 (Pre-9/11)
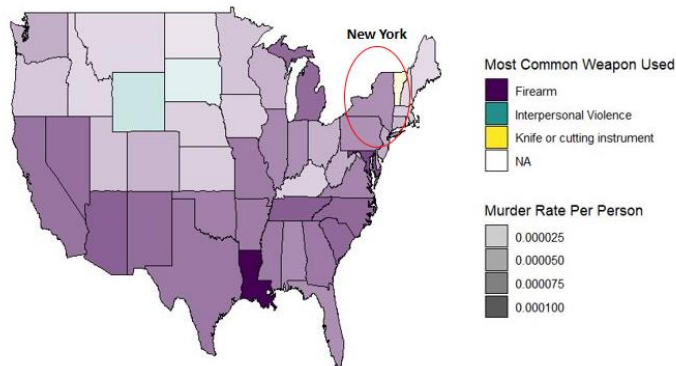
Murder Rate Per Person in the US: 2001 (Year of 9/11)

Murder Rate Per Person in the US: 2002 (Post-9/11)

<u>Analysis & Interpretation:</u>

While looking at the graphs above, we can see that in 2000 and 2001, New York was in the upper-mid range of murder rate per person in the U.S. In the rest of the U.S., it can clearly be seen that southern states have a higher murder rate per person than northern

states in 2000 and 2001. Firearms are also the most common weapon in New York and the U.S. in these years. In the year 2002, however, the graph drastically changes. The murder rate per person in New York decreases (as the intensity of the shading goes down), as well as in the rest of the U.S. Firearms are still the most common weapon used. The number of murders in New York was ~950 in 2001, but in 2002 the number of murders was ~900, about a 5% decrease. Overall, the major takeaways from the graphs are that the murder rate went down in the U.S. and New York on 9/11, as well as that the most popular murder weapon in the U.S. is firearms.

Conclusion:

Key insights from the research project include the murder rate per person in New York and in the U.S. having decreased following the events of 9/11. It cannot be said for certain that 9/11 affected the murder rate per person in the U.S. without further research, but they might be related. One possible effect that 9/11 had on the murder rate per person was that people might have been more hesitant to commit crimes, especially murder, for fear of greater repercussion from law enforcement. Another effect of 9/11 might have been a heightened sense of unity throughout the nation. Communities and neighborhoods, especially in New York, might have come together for support after such a major event. This would most likely deter people from committing violent crimes.

Limitations from this project include not having up-to-date population data per state for 2001 and 2002, and this may have skewed the results. The population data used was from the 2000 U.S. census, and this was the most ideal data source, as there is not a census every year. Another limitation of this is the change in the scale for the graph for 2002. The scale on the legend for this graph does not match the scale for 2000 and 2001, and this is because of the decrease in values for this year.

This project could be expanded in the future by gathering population data per year for each state. The graphs of this project could be improved by ensuring a constant scale of the legends. Also, utilizing different code packages could allow for the animation of the three graphs. This would mean all three graphs would be merged, and the animation switches the graph from year to year.

Sources Used

1. https://www.murderdata.org/p/data-docs.html
2. https://www.census.gov/data/tables/2000/dec/2000-resident-population.html