Arjoneel Dhar Final Project – ISTA 350

The dataset that I chose for my project is a dataset that collects murder and homicide clearance data from police jurisdictions around the United States. My goal in the project was to understand how murder numbers exist compared across various police jurisdictions, and which jurisdictions within the US and within Arizona have the highest murder numbers. In addition, I wanted to see the trend for murder number within the city of Baltimore as it was a part of my upbringing and I wanted to see if the rate of murders had increased or decreased. In addition, one of the reasons I chose this dataset was that to source it I would have to use read\_csv via the link that was sourced through BeautifulSoup. Web scraping is fundamental and I wanted to be able to understand more about how html data is structured and how to extract data and specific elements from it.

Files used to generate images:

get\_html.py – does a full web scrape

figures.py – gets the csv from beautiful soup, makes data frames, and makes the figures.

<https://github.com/RoyalDakat/ista_350_final_project> - link to github repository containing project materials.

<https://www.murderdata.org/p/data-docs.html> - link to website for data scraping.

<https://www.dropbox.com/scl/fi/r0rsmar4a2c16hqmc9u9v/UCR65_22.csv?rlkey=9jnbkmpcbchwvkawd3hnvztpc&dl=1> –scraped link to csv file used for the dataset.

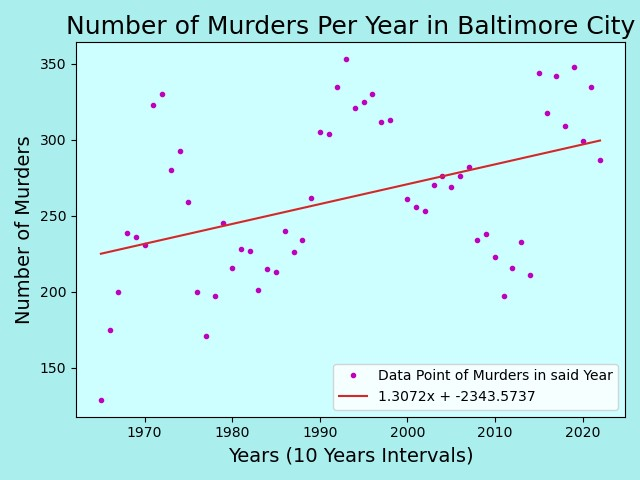
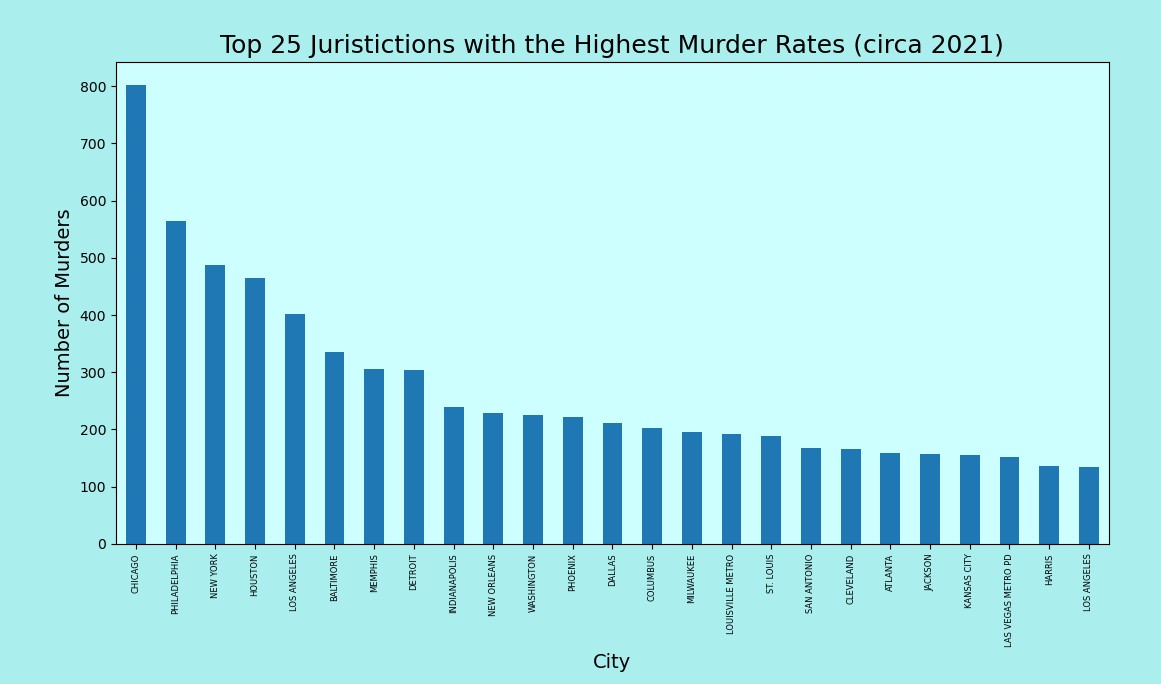
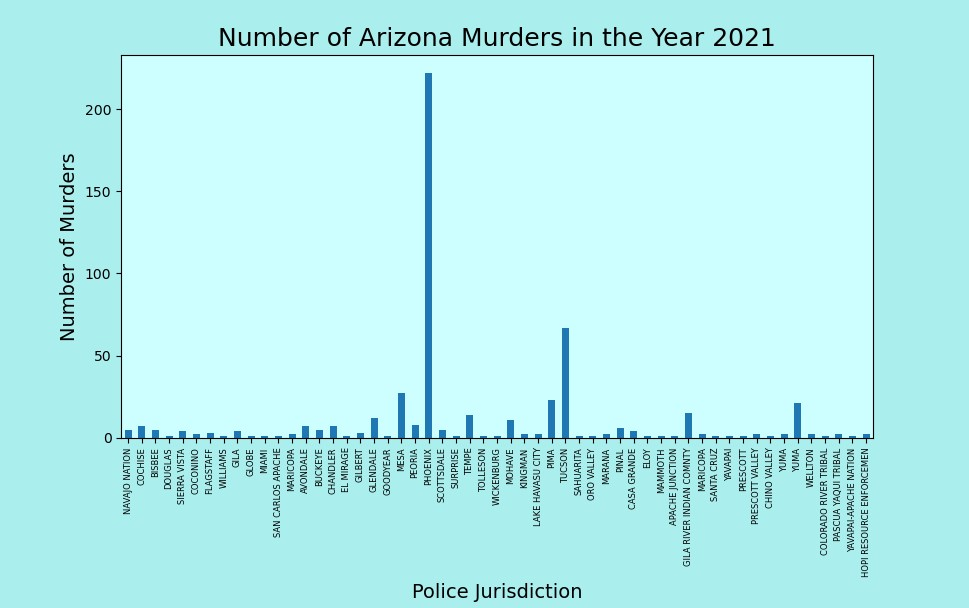


Figure 1 shows a graph of the murders in the city per year. The data starts in the year 1965 and goes until 2021. The data shows an interesting trend as the rate of murders is generally increasing at 1.3 more murders per year. It’s interesting to see that the fluctuation in the number of murders tends to coincide with historical events and time periods of excess. Ex. Drop in murders during the 2000’s and early 2010’s coinciding with the wars in Iraq and Afghanistan – A time of nation unity followed by economic downturn. versus the sharp rise in murders in the late 2010s corresponding with new conservative attitudes within the United States and the election of Donald J. Trump as president. To get a better perspective of the true nature of crime I think it would be important to factor in other variable such as population increase and density of murders per region as well as police force and budget.



The second figure shows the number of homicides in the top 25 homicide prone jurisdictions within the United States. I find it peculiar that the top 5 jurisdictions combined together could be compared to the homicide rates of the latter 20. Many of the cities listed have been historically hot spots for crime within the United States. An interesting perspective to consider could be how certain police jurisdictions have sub-jurisdictions that cover hot-spot areas. Such can the case for Chicago as the rates of murder in only certain parts of the city are extremely high, so to get a truly accurate understanding, it would be necessary to differentiate the various geographical regions during sampling and data collection.



The final figure shows the number of homicides comparing all police jurisdictions within the Arizona Area for the year 2021. Looking at the numbers it is appalling the level of homicides in the Phoenix jurisdictions, but it does make sense as the Pheonix area has most of Arizona’s underfunded and low-income areas, many of which are prone to crime. In addition, I think it’s interesting that aside from Tucson and Phoenix, some of the jurisdictions that have the higher murder rates, are ones that are proximate to Tucson and Phoenix; examples such as the Pima district and Mesa district, adjacent to Tucson and Phoenix respectively.

One final note, strangely enough the dataset had lots of holes for data from the years 2022 and 2023. I am unsure as to the reason for these holes, but an inference could be a side effect due to budget cuts post-COVID-19. This does not mean that the number of homicides isn’t being tracked, but that for many police jurisdictions, often times smaller ones especially, the custodians for the data have not updated the figures at the current time, possibly due to not hearing from them, or them being lower priority as the homicide numbers are generally low.