Chris Ross, Kai Yoshino, Isabella Spaletta, Marisa Nanakul

Project Description

- **What is the purpose of your research project?**
  - Our research project's main purpose is to compare the different crime rates and types to the different GDPs of city and metropolitan regions. The crime data will allow us to look at the differences between larger population centers and smaller urban locations.
- **What is the dataset you'll be working with? Please include background on who collected the data, where you accessed it, and any additional information we should know about how this data came to be.**
  - GDP by city - https://www.bea.gov/regional/index.htm
    - The Bureau of Economic Analysis out of the US Department of Commerce has statistical data on the GDP by state and by metropolitan area. Since the FBI crime data statistics are split by city, using the GDP by metropolitan area data fit our analysis the best. We accessed this data at this data at the official governmental site for the Bureau of Economic Analysis.
  - Crime by city - https://ucr.fbi.gov/crime-in-the-u.s/2013/crime-in-the-u.s.-2013/tables/table-8/table_8_offenses_known_to_law_enforcement_by_state_by_city_2013.xls/view
    - The Federal Bureau of Investigation has an open data source with details about crime statistics for the last 20 years. We specifically were looking at the crime rates in each city in 2013 (the link given above). We accessed this data from the FBI's official website with the data files. One thing to note is that both data sets need to be similar date (year) wise.
- **Who is your target audience? Depending on the domain of your data, there may be a variety of audiences interested in using the dataset. You should hone in on one of these audiences.**
  - Our audience would be a group similar to the National Crime Prevention Council, whose focus is on different process or policy that focus on preventing crime instead of enforcement of laws. Looking at these data sets we can help to not only find outliers in terms of large crime centers by population, but also determine if there is some correlation between education and percent of crime committed. This would help crime prevention organizations focus their efforts on specific cities along with

different crimes. On top of this it can help find policies within areas that go to helping prevent crime.
    - http://www.ncpc.org/programs
- **What should your audience learn from your resource? Please list out at least 3 specific questions that your project will answer for your audience.**
    - How does population size affect a city's crime rate? Does it affect the rate of a specific crime more?
    - How does a city's GDP affect its crime rate? Does it affect the rate of a specific crime more?
    - How does population size of a city affect the probability that a crime is violent or nonviolent?
    - How does changes in population affect changes in specific crime rates over time?

Technical Description

- **What will be the format of your final product (Shiny app, HTML page or slideshow compiled with KnitR, etc.)?**
    - HTML page with different visualizations such as ploty or D3. For the backend we are not going to have anything overly complicated and will just have different CSVs relating to the different data sets.
- **Do you anticipate any specific data collection / data management challenges?**
    - The FBI data is given by city within the United States, while the BEA data for GDP's is by metropolitan areas so some cities are combined. This means that for the FBI data we will need to bring together some cities to match some of the specific data within the BEA GDP data.
    - If we want to bring in other interesting data to compare such as education or health statistics, we have to find these statistics per city again which can come to be difficult.
- **What new technical skills will need to learn in order to complete your project?**
    - Learning how to get visualizations using plotly or D3 and how to get them to appear on the HTML page.
    - Different forms of analysis that we can conduct on this data to have interesting visualization
    - Wrangle the data to fit our visualizations.
- **How will you conduct you analysis? Please include a detailed description of your intended modeling approach.**

    - Our first challenge will to collect the different information about the cities over the selected time frame and grouping them by city with a different row indicating a different year. In this way we will be able to take an over time approach to the data set. We will then need to aggregate the information to include anything that we find interesting each year and compare it over time to each other.
- **What major challenges do you anticipate?**
    - Defining city size based on population.
    - Producing meaningful and effective visualizations.
    - Exploring and expanding on these data sets to find interesting things to analyze. Possibly bringing in new data.