Based on the requirements for the datasets that will be used in this term project, it seems like it would be best to use data from a field that has a large amount of historical data available and is constantly updated. This is to ensure that there is a high likelihood that both flat files and APIs will be available for the data itself.

Given those requirements, I have decided to pursue crime data. There is plenty of historical data available, and there are also API’s available and many web portals that can show the crime rates in a given location.

The API that I intend on using is:

<https://crime-data-explorer.fr.cloud.gov/api>

This tool is an API that can be used to query the public FBI crime database

The csv file that I intend on using is:

<https://www.kaggle.com/katzwigmore/portland-oregon-crime-data>

This is a historical record of crime data in the Portland, OR area. This will serve as a reference during my project.

Website that I would like to pull data from:

<https://www.ucrdatatool.gov>

This is the website that the FBI uses to host a broad collection of crime data.

I think that the best relationship that can be created from each of these data sets is the datetime that is associated with the crime. One of the few things that is almost always included with a police report is a date, so tying these data sources together using the date will, I believe, be most beneficial. This will not be a unique key, of course, but it can serve as in index, I believe.

I currently live in Portland, Oregon. For that reason, I am interested in examining how my city is doing in relation to the rest of the country, in regard to crime rates. I do not know much about how crime rates are measured, so I look forward to dissecting what all of the data that is accessible in these sources is, and what it means. If I need to cut down on some data, then it may be beneficial to limit my research to only one type of crime, or simply look at crimes in Portland, Oregon as a whole. The majority of the data appears to be in a standard format, at least with the current data. The FBI uses a Universal Crime Reporting program to aggregate the data that it gets, so it (should) be in mostly the same format. I hope this will cut down on the data cleaning that will be necessary for this project.