Lab 5

Amrutha Rao and Madison Goodwin

Research question: How do the types and severity levels of reported crimes in the Midwestern United States vary by neighborhood, and are there areas with higher rates of violent crimes compared to non-violent crimes?

Data Source 1: <Crimes_-_2001_to_Present.csv>

Website 1: [Crime Data | Open Minneapolis (minneapolismn.gov)](https://opendata.minneapolismn.gov/datasets/cityoflakes::crime-data/about)

The database shows reported incidents of crime (with the exception of murders where data exists for each victim). This database is specifically for Chicago, Illinois. The data is sourced from catalog.data.gov and the source of the data is the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. This continually evolving dataset currently accounts for 1,048,575 reported incidents of crime. The data is updated but doesn’t include the most recent seven days. The data is provided by the city of Chicago, and I contacted the Data Fulfillment and Analysis Division of the Chicago Police Department at DFA@ChicagoPolice.org to confirm its reliability. We chose to store this database as a CSV file. This database can also be stored as a JSON file, XML file, RDF file and Landing page. We chose to store this file as a CSV file for multiple reasons. CSV files are efficient for storing and transferring large datasets. CSV files are also easy to work with as their simple structure is easily understandable. CSV files also maintain data integrity as there is less of a risk of data corruption. CSV files are text-based and are more suited for GitHub. This makes it easier for the group to work on the assignment together. The dataset provided by opendata.minneapolismn.gov passes the “smell test”. The author is a veritable source to be contacted for answers to any questions one may have. As previously mentioned, the data is regularly updated and checked for errors, excluding the previous seven days. The dataset comes with information as to how it was acquired. Given my overall knowledge of the topic, the data does seem plausible. The data set can easily be opened in a simple document viewer, we used Microsoft Excel, without crashing the computer. The data appears to be properly labeled and organized, so you can easily pluck out each piece of information. The data is easily stored and can be moved if you need to use more than one laptop or computer to view the dataset. The data is real-time and accessible via an API, meaning you can get the data you need by requesting it live.

Data Source 2: <PDI__Police_Data_Initiative__Crime_Incidents.csv>

Website 2: [PDI (Police Data Initiative) Crime Incidents | Tyler Data & Insights (cincinnati-oh.gov)](https://data.cincinnati-oh.gov/safety/PDI-Police-Data-Initiative-Crime-Incidents/k59e-2pvf)

The database shows reported crime incidents in Cincinnati, Ohio. The data is sourced from data.cincinatti-oh.gov and the source of the data is the Cincinnati Police Department. The data was created through the City’s computer-aided dispatch (CAD) system and is stored in the City’s Records Management System (RMS). The crime incidents are the records of reported crimes, collated by an agency for management. This regularly updated dataset currently accounts for 515,270 reported crime incidents, and the data is updated daily. The data is provided by the city of Cincinnati, and I contacted the dataset owner to confirm its accuracy. We chose to store this database as a CSV file. This database can also be stored as an RSS file, XML file, RDF file and TSV file. We chose to store this file as a CSV file for a few different reasons. CSV files are efficient for collecting and transferring large datasets. CSV files are also easier to work with as their simple structure is easily comprehensible. CSV files also sustain data integrity as there is less risk of data corruption. CSV files are text-based and therefore more suited for GitHub. This makes it easier for a team of people to edit the file based on the information they need. This dataset passes the “smell test”. The author is a factual source to be contacted for anyone with concerns or inquiries about the dataset. As aforementioned, the data is often updated and checked for any mistakes. The dataset also comes with details about how the data was obtained. Given my overall knowledge of the topic, the data does seem probable. I’ve used this website in the past so I’m familiar with its credibility. The dataset can effortlessly be opened in a simple document viewer, we used Microsoft Excel, without crashing the computer. The data appears to be properly labeled and organized, so you can easily pick out each piece of information. The data is easily stored and can be moved if you need to use more than one laptop or computer while working on it. The data is real-time and accessible via an API, meaning you can get the data you need by requesting it live.

Data Source 3: <Crime_Data.csv>

Website 3: [Crimes - 2001 to Present - Catalog (data.gov)](https://catalog.data.gov/dataset/crimes-2001-to-present)

The database shows reported crime incidents in Minneapolis, Minnesota. The data is sourced from opendata.minneapolismn.gov and the source of the data is the Minneapolis Police Department. The data was created through the City’s computer-aided dispatch (CAD) system and is stored in the City’s Records Management System (RMS). This continually evolving dataset currently has 251,326 records of crime incidents, and the data is updated on a daily basis by 9:30 AM. The data is provided by the city of Minneapolis, and I contacted the dataset owner to confirm its legitimacy. We chose to store this database as a CSV file. This database can also be stored as a KML file, Shapefile, and GeoJSON file. We chose to store this file as a CSV file for a multitude of reasons. CSV files are efficient for gathering and transferring large datasets. CSV files are also facile to work with as their straightforward structure makes the information simpler and more perspicuous. CSV files also maintain data integrity as there is less risk of data corruption. CSV files are text-based and as a result, more suited for GitHub. This makes it easier for multiple people to work on the assignment together. This dataset passes the “smell test”. The author is a reliable source to get in touch with if more information about the dataset is desired. As previously mentioned, the data is frequently updated and checked for fallacies. The data comes with information as to how it was gathered. Given my overall knowledge of the topic, the data does seem credible. The data set can easily be opened in a simple document viewer, we used Microsoft Excel, without crashing the computer. The data appears to be appropriately labeled and organized, so you can easily understand each piece of information. The data is easily stored and can be moved if multiple people are using their own devices to view the dataset. The data is real-time and accessible via an API, meaning you can get the data you need by requesting it live.