Database Link

Crime Database User Guide

Introduction

What is this collection of data sets?

The collection of data sets this user guide looks at consists of all data sets from data.gov that have a 'crime' tag and have been produced by local governments. By clicking on the hyperlink above, you can find this collection of 105 data sets detailing crime reports from across the United States, including states like Maryland, New York, Connecticut, Texas, and California. The data sets range from data about an entire state to sometimes just data for a specific city, such as one data set that is specifically for Austin, Texas. This user guide intends to show the user how they can search and filter the data sets for specific variables and how to download and open different file types that the data sets come in. Along with this, there is further information about where the data in these data sets come from, how it was collected, and what limitations might be applied when working with them. Finally, the visualizations tab provides statistics about what can be found in the data along with examples on how the data within the data sets can be used.

What can I find within these data sets?

Quick Navigation

Introduction

Table of Contents

Introduction How to Operate

Downloading the Data Sets

Things to Note About the Data

FAQ

Visualizations

Version History

This database showcases a variety of data revolving around crimes committed, with some data sets being state or county-specific while others focus on only violent crimes and some focus on only non-violent crimes. Some of the data sets detail how certain states organize their policing districts, allowing users to draw comparisons between how these area sections can be impacting the local crime types and frequency. To add to those evaluations, there are some data sets that report on actions taken to reduce crimes. Some data sets report on the demographics of the victims, which could be used to better understand motives behind the crimes committed or identifying who is most at risk. Users will be able to evaluate things like crime frequency, types of crimes committed, and locations they happened. In some cases, there are also data sets with police responses to crime, as well as data sets detailing the arrests that were made. Through this user guide, we will provide users with some different ways they may be able to interpret or

compare these data sets and the tools to analyze them further for their research or interest purposes. The data ranges from 1975-2024, with most data sets focused on data from 2001-2024.

Table of Contents

How to Operate

How to Search/Filter the Database

The filtering tags and search bar tool on the database are a great way of ensuring that the data set being found fulfills the purpose of the researcher(s). Data.gov's website uses the DCAT-US Schema v1.1 to tag and filter all of the data sets on this website. This schema has a specific set of keywords and vocabulary that is used for the tagging along with a specific number of fields that need to be filled out for each data set posting. There is a top search bar that is already populated with the word "crime", which can be deleted and replaced with any intended keyword (such as "Annual Crime") while still maintaining the "crime" tag. This will allow the user to target a specific search of interest for crime-related data. Another way to filter the data set is to navigate to the left sidebar of the database, where multiple filtering options are shown. The sidebar starts off with a "Filter by location" with an assistive mapping feature that the user can drag and denote exactly where in the world they want the data sets to be from. Users are able to filter by file type or by who published the data, such as limiting the data to only JSON and CSV files or by publishers of a certain state. If specific data needed to be found regarding the state of Maryland, the option of "State of Maryland" in the Organizations filter can be selected as well as "opendata.maryland.gov" in the Publishers filter (both returning the same eight data sets after being selected). Combining the filtering options along with the search bar will guarantee a precise selection of data sets intended for the purpose of the research.

How to Navigate the data sets?

After selecting a data set, there is information provided on the specific topic (which will most likely be "local government" in our case), publisher, and contact. Contact information is a great tool for being able to email the publishers of the data set and asking any questions regarding information that is not provided or not thorough enough. There will be a short description of the data set below the title of it, allowing for further information regarding what the data set is about before downloading. "Additional Metadata" is also provided at the bottom of this page, such as when the data was created/uploaded and who is the maintainer of the data to ensure the reliability of the data that is being presented. To download the data sets, there will be a header about halfway on the page that states "Downloads & Resources" - and from there, a download can be selected from any of the following formats below the header. Clicking on the text indicating the format of the file (such as "CSV") or the associated icon will navigate you to the next page. The description of the data set will appear once more and a download button and icon in the top right, prompting the user to download. The data set can then be opened in an associated interpreter for that file and read.

Downloading the Data Sets

What file types can I download the data sets in?

The data set can be downloaded in a number of different files, such as CSV (comma-separated value), JSON (JavaScript Object Notation), or XML (Extensible Markup Language). CSV files are typically used when conducting data analysis in Python due to the convenience of being able to import the data into a dataframe for easy visualization. JSON is a format typically used when pulling data from a source through an API. Though it can also be used in Python, it requires slightly different syntax than if using a CSV, and requires more from the user to visualize the data in the same data frame as if it had been imported as a CSV file. XML is a markup language that describes the text in a document. Once opened, users will be able to see the number of occurrences of different types of crime across the different counties in Maryland since 1975. The various columns show types of crime as well as relevant information about the year to year changes.

How to Export the data set

All of the different file types can be downloaded from http://data.gov and opened in any text-editing application to view the contents of the file. Common text-editing applications include VSCode, Spyder, or the built-in text-editing application on your computer.

Which programs can be used to open each file type

While you can use any text-editing application listed above to view the contents of the file in whatever filetype you downloaded, the easiest way to visualize the contents is to use Python to load in the data as a CSV, JSON, or XML file, Doing so will allow you to visualize the data in a data frame and conduct your analysis from there using the various functions built-in to the Pandas module. It is important to note that downloading as a CSV file is the recommendation as it is the most intuitive for working with Python. Choosing to download the file as a JSON or XML file will require more familiarity with working with those files in Python. Should you choose to download the data as an RDF file, there are a number of applications that can be used to open the file depending on your operating system. That list can be found here. Below you can find an example of what the data may look like once it has been opened in Python.

```
In [6]: df = pd.read_csv("Violent_Crime___Property_Crime_by_County__1975_to_Present.csv")
        df
```

Out[6]:

	JURISDICTION	YEAR	POPULATION	MURDER	RAPE	ROBBERY	AGG. ASSAULT	B & E	LARCENY THEFT	M/V THEFT	...	B & E PER 100,000 PEOPLE	LARCENY THEFT PER 100,000 PEOPLE	M/V THEFT PER 100,000 PEOPLE	MURDER RATE PERCENT CHANGE PER 100,000 PEOPLE	PER CH 1 PI
0	Allegany County	1975	79655	3	5	20	114	669	1425	93	...	839.9	1789.0	116.8	NaN	
1	Allegany County	1976	83923	2	2	24	59	581	1384	73	...	692.3	1649.1	87.0	-36.7	
2	Allegany County	1977	82102	3	7	32	85	592	1390	102	...	721.1	1693.0	124.2	53.3	
3	Allegany County	1978	79966	1	2	18	81	539	1390	100	...	674.0	1738.2	125.1	-65.8	
4	Allegany County	1979	79721	1	7	18	84	502	1611	99	...	629.7	2020.8	124.2	0.3	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
1075	Worcester County	2015	51566	1	15	27	102	273	1562	28	...	529.4	3029.1	54.3	0.4	
1076	Worcester County	2016	51255	3	17	39	93	289	1514	32	...	563.8	2953.9	62.4	201.8	
1077	Worcester County	2017	51408	2	26	43	106	220	1514	39	...	427.9	2945.1	75.9	-33.5	
1078	Worcester County	2018	51304	0	12	24	88	215	1178	34	...	419.1	2296.1	66.3	-100.0	
1079	Worcester County	2019	51606	0	12	24	94	186	1086	30	...	360.4	2104.4	58.1	NaN	

1080 rows × 38 columns

Things to Note About the Data

With the amount of data provided from this Database, there are several things to note and understand in order to properly utilize this data.

Data Value Ranges

With the amount of data sets underneath the crime filter, the ranges of data values and how the data is categorized ranges from set to set, depending on the city and department that they were collected from. Certain data sets, such as the "911 Calls For Service 2017" from Baltimore City, Maryland describe the types of calls received by the police department in the year, where they came from, and how they were prioritized. The data set "Citywide Crime Statistics" from the New York City Police Department provides statistics on the types of crime complaints they receive in a year, the amounts, and the percentage of overall crimes that category contributes to the yearly crime complaint number. It is important to note that not every data set is designed to be compared directly with one another due to these different values, but they can be used to be compared with other data sets collected in the city, state, or area to contextualize the data further.

Finding more about where the data is coming from

The majority of the data collected from this database were collected by police departments and their call centers. Data sets such as "Anne Arundel County Crime Totals By Type" from the State of Maryland emphasize this, with breakdowns of crimes committed by year, and the "911 Calls For Service 2017" from Baltimore City, Maryland showing their call center statistics from their police station call centers. There are several data sets where the data provided comes from a more specialized source, such as the data set "Micro Community Policing Plans" from the Seattle City GIS program, which defined policing boundaries for each neighborhood in Seattle, Washington based on residency perception of crime. The more specialized the source of data, the less globally applicable the data becomes, so it is important to understand what the purpose of research is before engaging with more complex data sets.

Limitations with the data

With the amount of data comes a few limitations, however. Several columns in the data sets are empty or unused by a majority of the data entries. For instances where there are columns with very few values, it is recommended to omit the data collected from these columns when looking at the data set as a whole. The data collected from these columns can be used to contextualize individual entries, however.

[Database Link](#)

FAQ

- *Why do I want to use this collection of data sets?*

If you're interested in evaluating crime statistics across multiple data sets from all around the United States. There's valuable information available regarding types of crimes committed, the rate at which they're committed, steps taken to penalize criminals and/or prevent future crime from occurring. There's multiple different types of crime data to explore, so any and all research or interest surrounding crime will be able to find something here.

- *How can I learn to use this website?*

You can check out for more information in this page "How to Operate" regarding this issue.

- *Is there a list where I can find all crime data sets available?*

Unfortunately, there is no list that displays all the crime-related data sets collected. The best way to look for data sets is to apply filters that would provide the user specific data sets he/she needs. For more information, visit the page "How to Operate" in section "How to Search/Filter the database

- *Are there any restrictions on how I can use the data on this website?*
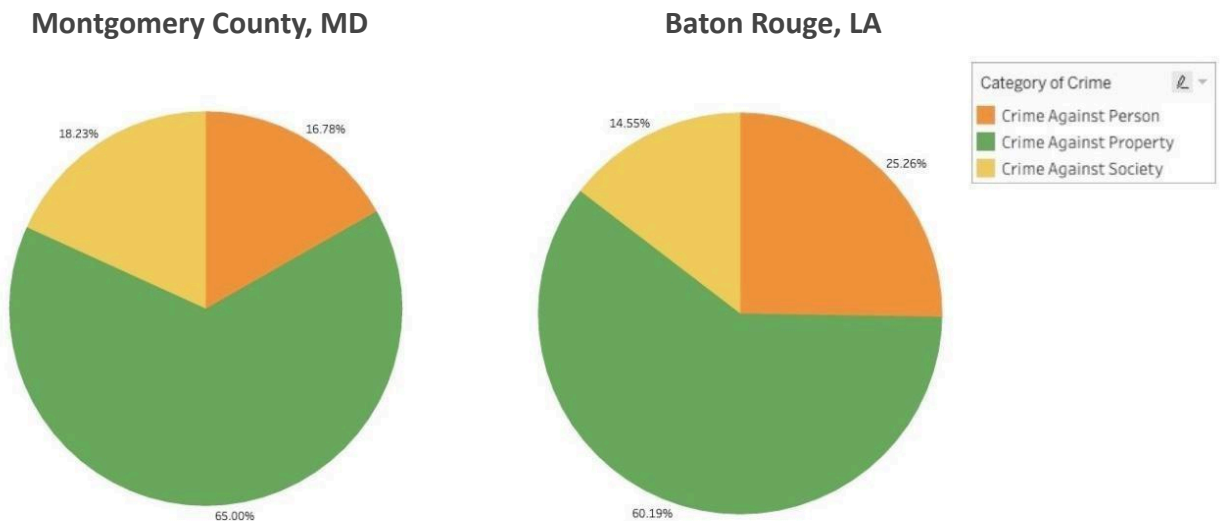
Since this is a website for open data, we do not plan on having any restrictions towards accessing the data

- *How were these data sets collected for Data.gov?*

These crime-related data sets were collected through links from local government agencies. Also, liaisons from these government agencies come in contact with the Data.gov team if their agency wishes to publish any data.

Visualizations

Montgomery County, MD   **Baton Rouge, LA**

Category of Crime
- Crime Against Person
- Crime Against Property
- Crime Against Society

Questions: "What is the most common category of crime in (Montgomery County, MD / Baton Rouge, LA)?", "What proportional differences are there in regards to the crime category between the two regions?"

These visualizations demonstrate how pie charts can be used to visualize the proportions of a numeric variable for each level of a categorical variable. In these visualizations, every slice of the pie represents the proportion of crimes committed for the selected crime category against the total crimes committed in the allotted time periods (Montgomery County: July 2016 Present, Baton Rouge: January 2021 - Present). Multiple visualizations allow users to compare similar variables across different data sets and use the observed similarities and differences to answer questions/draw conclusions on the data.

Database Link

Most Common Crime Locations in Montgomery County since July 2016

Category of Crime
- Crime Against Person
- Crime Against Property
- Crime Against Society
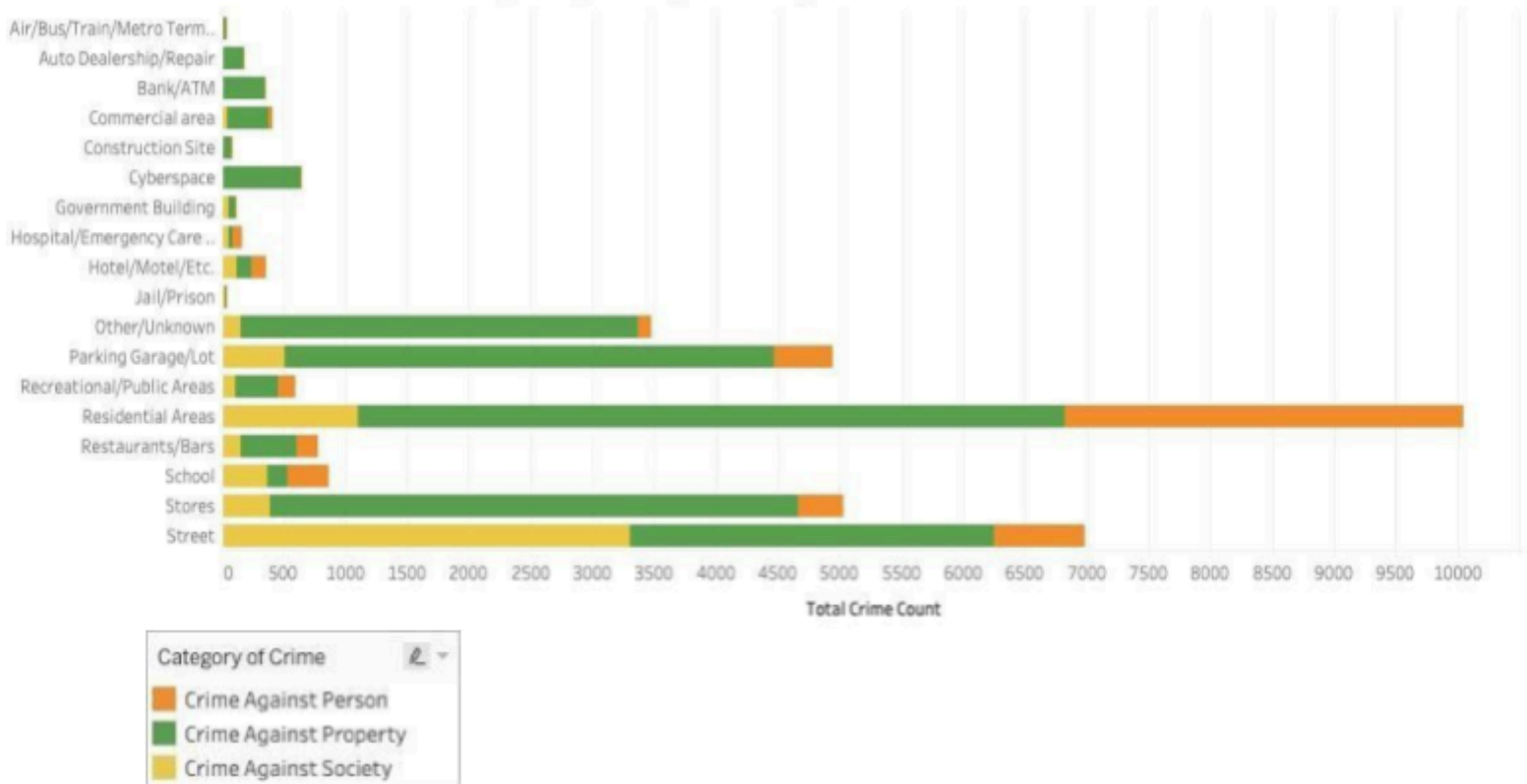
Question: "What is the most common category of crime at (location) in Montgomery County?", "Are you more likely to encounter a crime at (location one) or (location two)?"

This visualization demonstrates how stacked bar charts can be used to visualize multiple categorical variables against a numeric variable. In this case, the primary categorical variable (crime location) is split up based on the category of the secondary categorical variable (crime category). This could be beneficial for visualizations due to the fact that you can see both the distribution of the primary categorical variable against the numeric variable and the distribution of the secondary categorical variable in relation to the primary categorical variable.

Database Link

Most Common Crime Types in Baton Rouge since January 2021

Category of Crime

- Crime Against Person
- Crime Against Property
- Crime Against Society
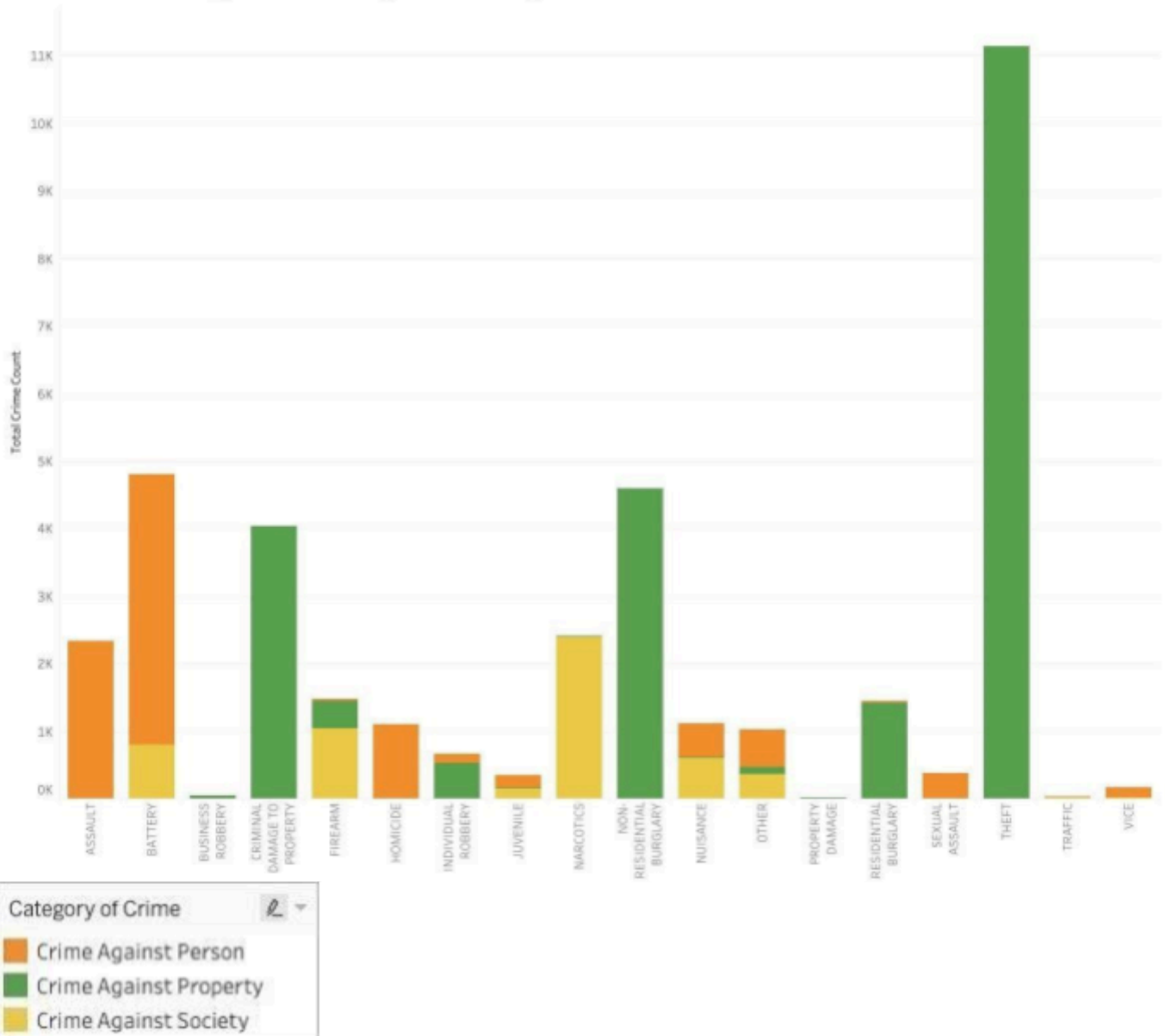
Question: "Were there more (type of crime)'s committed against (person, property, society) or (person, property, society)?", "What is the most common crime type in Baton Rouge, LA"

This visualization demonstrates the benefit of using visualizations that measure against multiple variables as opposed to one dimensional visualizations. Generally, visualizations with multiple variables allow for more questions to be asked, leading to more conclusions being drawn about the data. In the case of bar charts, stacked bar charts are used as opposed to regular bar charts

in order to allow users to branch off of the more obvious questions that simple bar charts are able to answer. For example, in this visualization, the obvious question would be "What is the most common crime type?" With the addition of another variable, this question can be expanded on in order for more questions to be answered and conclusions to be drawn.

[Database Link](#)

Version history

Version	Month and Year
1.0	May 2022
2.0	November 2022