

Data.gov Traffic Data Instructional User Guide

May 10, 2023

Prepared by UMD iSchool Project Team 2

Table of Contents

Data.gov Traffic Data Instructional User Guide.....	1
Table of Contents.....	2
<i>Introduction.....</i>	<i>3</i>
What does this collection of data sets consist of?.....	3
What type of information or data is contained within these data sets?.....	3
<i>Navigating the System.....</i>	<i>4</i>
How to Utilize the Databases Search and Filter Features.....	4
How can one effectively navigate the data sets?.....	4
<i>Acquiring the Data Sets.....</i>	<i>5</i>
In what formats can the data sets be downloaded?.....	5
Exporting the Data.....	5
How can I determine which programs to use for opening different file types?.....	5
<i>Important Considerations Regarding the Data.....</i>	<i>7</i>
Value intervals of Data.....	7
Understanding the sources of the data.....	7
Data Restrictions.....	7
<i>FAQ.....</i>	<i>8</i>
How accurate and reliable is the Traffic Data available on Data.gov?.....	8
How can I combine multiple data sets from Data.gov to perform a comprehensive analysis?.....	8
How frequently is the crime data on Data.gov updated?.....	8
Can I use the traffic data on Data.gov for commercial purposes?.....	8
<i>Visualizations.....</i>	<i>9</i>
The distribution of motorcycle deaths by borough and year in New York City from 2012 to 2023.....	9
Baton Rouge Crime Incidents.....	10
A Collection of Montgomery County Traffic Related Datasets.....	11
<i>Version History.....</i>	<i>13</i>

Introduction

What does this collection of data sets consist of?

The data sets we're discussing here constitute a comprehensive collection of traffic-related data available on data.gov. This repository offers a plethora of data sets covering a wide spectrum of traffic aspects, including accidents, volume, congestion, and safety measures. With the data sets spanning across various regions in the US, such as California, Texas, and New York, among others, it offers a wealth of information for researchers, analysts, and policymakers.

Our user guide aims to assist you in navigating this vast data repository with ease. The guide offers practical guidance on how to search, filter, download, and open different file types available. We also provide comprehensive information on the data's origin, collection methodology, and limitations associated with working with it. Moreover, our extensive visualizations and examples provide further insight into the data, allowing you to maximize its potential.

This data collection is a valuable resource for researchers, analysts, and policymakers alike. Our user guide is designed to enable you to leverage the data's insights to make informed decisions that impact traffic and transportation in your community.

What type of information or data is contained within these data sets?

This database provides a rich collection of traffic-related data sets that offer insights into various aspects of traffic safety, including traffic volume, congestion, accidents, and fatalities. The data sets are sourced from a range of local and state government agencies across the United States and span several decades, with some data sets dating as far back as the 1980s. The data sets cover a wide range of topics and can be used to identify trends and patterns in traffic behavior, evaluate the effectiveness of traffic safety interventions, and inform policy and planning decisions.

For example, some of the data sets focus on traffic volume and provide detailed information on the number of vehicles on different roads and highways at different times of the day or year. These data sets can be used to identify areas of high congestion and to plan road improvements or expansions to accommodate growing traffic volumes.

Other data sets focus on accidents and provide information on the location, time, and nature of accidents, as well as the number and severity of injuries or fatalities. These data sets can be used to identify accident hotspots and to develop targeted interventions to improve safety in these areas.

Navigating the System

How to Utilize the Databases Search and Filter Features

To search for specific traffic-related data sets on the Data.gov website, users can utilize the search bar or filtering options. The search bar at the top of the page can be used to input keywords related to the data set of interest, such as “traffic fatalities” or “traffic congestion.” Once entered, the search results will display all relevant data sets containing the keyword(s). Alternatively, users can use the filtering options on the left sidebar of the page to narrow down the search results. The filtering options include filtering by file format, publisher, organization, and more. For example, users can filter by file type to limit the search results to only CSV or JSON files.

Another helpful filtering option is filtering by publisher, which allows users to see datasets published by specific organizations. For instance, a user interested in traffic data for the state of New York can select the “New York State Department of Transportation ” from the publisher filter to view all relevant datasets published by that organization. Additionally, users can filter by location using an interactive map. By clicking and dragging the map, users can select specific areas and limit the search results to data sets from those locations. Overall, utilizing the search bar and filtering options on the Data.gov website allows users to efficiently find and access traffic-related data sets of interest.

How can one effectively navigate the data sets?

When it comes to effectively navigating the data sets available on the traffic-related section of the catalog.data.gov website, there are a few key steps to keep in mind. By following these steps, users can quickly and easily access the information they need to conduct their research or gain insights into traffic patterns in their area.

STEP 1: SELECTING A DATA SET After accessing the traffic section of the website, users should take the time to carefully review the available data sets to determine which one best suits their needs. Each data set is accompanied by a short description that provides some basic information about its contents, as well as additional metadata that can help users evaluate the reliability of the data.

STEP 2: DOWNLOADING THE DATA SET Once a data set has been selected, users can access the Downloads & Resources header about halfway down the page. From here, users can choose to download the data in a variety of formats, including CSV and JSON. The downloaded file can then be opened in a compatible program and analyzed further.

STEP 3: ASKING FOR HELP If users have any questions about the data sets or require further assistance, they can use the provided contact information to reach out to the publisher of the data. This can be especially helpful if users require additional information that is not provided in the data set's description. By following these steps, users can quickly and easily navigate the traffic-related data sets available on catalog.data.gov. With access to this wealth of information, researchers, policymakers, and other interested parties can gain valuable insights into traffic patterns, congestion, and related topics.

Acquiring the Data Sets

In what formats can the data sets be downloaded?

The datasets available at the provided link can be downloaded in a variety of file formats, including CSV (comma-separated value), JSON (JavaScript Object Notation), XML (Extensible Markup Language), HTML (Hypertext Markup Language), ZIP (compressed file format), KML (Keyhole Markup Language), ArcGIS Geo Services REST API, and GeoJSON. Each file format has its own unique advantages and use cases. For example, HTML files can be used for web development, ZIP files can be used to compress and package multiple files, KML files are used for displaying geographic data in an Earth browser such as Google Earth, and GeoJSON files are used for encoding geographic data structures. Additionally, the ArcGIS Geo Services REST API can be used for geospatial data analysis, and it provides access to a wide range of spatial data and tools. The datasets cover various traffic-related information, including traffic volume, accident rate, traffic flow, and other relevant details, across different regions and time periods, allowing users to gain insights and visualize the data in their preferred format.

Exporting the Data

To export the datasets available at the provided link, the process is quite straightforward. Simply locate the dataset of interest, confirm compatibility with the desired file format, and download the file by clicking the corresponding download button. Once the file is downloaded, you can use a reliable text-editing application such as Visual Studio Code (VSCode) or Spyder to open and explore the data. Both applications offer advanced features and support for various programming languages, making them useful for data analysis and exploration. Additionally, many operating systems come with built-in text editors that can be used to view the data as well.

How can I determine which programs to use for opening different file types?

Opening different file formats requires software that can parse and interpret the structure and contents of the file. While basic text editors can open many file formats, working with more complex data sets may require specialized software. For example, when working with CSV, JSON, or XML files, it's recommended to use Python and its Pandas library for easier data manipulation and

analysis. Some file formats, such as ZIP archives, can be extracted with built-in system utilities or third-party software. KML files can be opened in Google Earth or other software that supports the format. When dealing with ArcGIS Geo Services REST API or GeoJSON files, GIS software such as ArcGIS or QGIS can be used. It's important to choose software that can handle the specific file format and its associated features to work with the data effectively. Example:

```
In [7]: import pandas as pd

file_path = r'C:\Users\LopezIngmar\Downloads\Traffic_Violations.csv'
df = pd.read_csv(file_path)
df

C:\Users\LopezIngmar\AppData\Local\Temp\ipykernel_20292\4001856591.py:4: DtypeWarning: Columns (19,20,21,22,23,24,25,34) have mixed types. Specify dtype option on import or set low_memory=False.
df = pd.read_csv(file_path)
```

```
Out[7]:
```

	SeqID	Date Of Stop	Time Of Stop	Agency	SubAgency	Description	Location	Latitude	Longitude	Accident	...	Charge	Article
0	fb324ab-bc8d-4743-ba23-7f9f370005e1	08/11/2019	20:02:00	MCP	2nd District, Bethesda	LEAVING UNATTENDED VEH. W/O STOPPING ENGINE, L...	CORDELL ST @ NORFOLK AVE.	38.989743	-77.097770	No	...	21-1101(a)	Transportatic Artic
1	a8d904ec-d888-4bc3-8884-f37a4b31854d	08/12/2019	13:41:00	MCP	2nd District, Bethesda	EXCEEDING POSTED MAXIMUM SPEED LIMIT: 85 MPH L...	NBI270 AT MIDDLEBROOK RD	39.174110	-77.248170	No	...	21-801.1	Transportatic Artic
2	54a848a-df28-4b68-a338-08883886aa40	08/12/2019	21:00:00	MCP	5th District, Germantown	DRIVING VEH W/ TV-TYPE RECEIVING VIDEO EQUIP T...	MIDDLEBROOK AN 355	39.182015	-77.238221	No	...	21-1129	Transportatic Artic
3	cf5479b8-9bc7-4218-a7b2-99e57ae932af	08/12/2019	21:43:00	MCP	5th District, Germantown	DRIVING VEHICLE ON HIGHWAY WITH SUSPENDED REGI...	GERMANTOWN RD AND ALE HOUSE	39.180508	-77.284023	No	...	13-401(h)	Transportatic Artic
4	5801ca35-8ee7-4f8e-9208-d89cde98d489	08/12/2019	21:30:00	MCP	2nd District, Bethesda	FAILURE OF LICENSEE TO NOTIFY ADMINISTRATION O...	EASTWEST/ 355	38.984247	-77.090548	No	...	16-116(a)	Transportatic Artic
...
1850587	970e89d3-6bd4-4fb4-b8dd-c43a7d78e529	03/17/2023	07:15:00	MCP	2nd District, Bethesda	EXCEEDING THE POSTED SPEED LIMIT OF 35 MPH	GEORGIA AVE @ HOME DEPOT	39.186717	-77.270578	No	...	21-801.1	Transportatic Artic
1850588	f8344fd2-d0a4-4e76-be94-734d65c924a8	03/17/2023	07:23:00	MCP	2nd District, Bethesda	EXCEEDING THE POSTED SPEED LIMIT OF 35 MPH	GREAT SENECA HWY @ GREY EAGLE CT	39.185297	-77.272923	No	...	21-801.1	Transportatic Artic
1850589	2a589e4c-2b4f-41ef-a812-4f907560c82c	03/17/2023	07:29:00	MCP	2nd District, Bethesda	EXCEEDING THE POSTED SPEED LIMIT OF 35 MPH	GREAT SENECA HWY @ GREY EAGLE CT	39.188082	-77.268558	No	...	21-801.1	Transportatic Artic
1850590	88798c5e-0e08-4bcb-8caf-de3cbf443a12	03/16/2023	08:07:00	MCP	Headquarters and Special Operations	HEADLIGHTS (*)	BETHESDA CHURCH RD @ SCHOOLYARD CT	39.284910	-77.210890	No	...	55*	Na
1850591	3d3b4cdb-f944-4959-84df-cb01541e98f3	03/17/2023	08:52:00	MCP	Headquarters and Special Operations	WINDOW TINT	MD28 AND SHADY GROVE RD	39.101120	-77.188433	No	...	61	Na

1850592 rows x 43 columns

Important Considerations Regarding the Data

Value intervals of Data

The dataset consists of traffic-related data collected by different government departments and agencies in the United States. The dataset includes various types of traffic data, such as traffic volume, speed, accidents, and more, from different locations in the country. The value intervals of the data vary depending on the type of data and the location it was collected from. For example, the data on traffic volume in New York City may have different value intervals compared to the data on traffic volume in Los Angeles. Therefore, it is important to understand the context and source of the data when analyzing and interpreting it. It is crucial to note that not all datasets are designed to be directly comparable due to their different value intervals and data categorization. However, these datasets can still be used to contextualize other data collected within the same area or state. It is also important to have a clear understanding of the data's limitations, potential biases, and any other relevant factors that may impact its accuracy or interpretation. This will help to ensure that the data is used appropriately and that any conclusions drawn from it are reliable and accurate.

Understanding the sources of the data

The data available in the provided link, particularly those related to traffic, have been collected from various sources, primarily from transportation departments and agencies. For instance, the "National Highway Traffic Safety Administration's Fatality Analysis Reporting System (FARS)" provides data on fatal traffic crashes, while "National Transit Database" offers information on transit service provided by transit agencies across the United States. Additionally, some datasets like "Traffic Volume Counts" are collected directly by state and local transportation agencies. It is important to note that the origin of the data can affect its scope and applicability, and users should consider this when selecting and analyzing datasets for their research purposes.

 Metadata

Data Restrictions

The Traffic data collection may have some restrictions due to the nature of the data collected. For instance, some of the columns in the data sets may have limited values or be irrelevant to the majority of the entries. To avoid any misinterpretation or inconsistency when analyzing the data set, it is recommended to focus on the relevant and meaningful columns. This will ensure that the data analysis process is both efficient and accurate.

FAQ

How accurate and reliable is the Traffic Data available on Data.gov?

The accuracy and reliability of the crime data available on Data.gov can vary depending on the source of the data and the methods used to collect and report it. It is important for users to thoroughly review and verify the data before using it for any serious research or analysis. Additionally, it is important to keep in mind that crime data is often subject to various biases, such as underreporting, inconsistent reporting practices, and variations in crime definitions and categorization. Users should take these factors into account when using the data.

How can I combine multiple data sets from Data.gov to perform a comprehensive analysis?

Combining multiple data sets from Data.gov can be a complex and time-consuming process, depending on the size and complexity of the data sets involved. Users will need to have appropriate data management and analysis skills, as well as access to specialized software tools for merging, cleaning, and analyzing the data. Some popular tools for data analysis include R, Python, and Excel.

How frequently is the crime data on Data.gov updated?

The frequency of updates to crime data on Data.gov can vary depending on the source of the data and the agency responsible for maintaining it. Some data sets may be updated on a regular basis, such as monthly or annually, while others may only be updated sporadically or not at all. Users should check the metadata for each data set to determine the last date of update, and contact the responsible agency for more information if necessary.

Can I use the traffic data on Data.gov for commercial purposes?

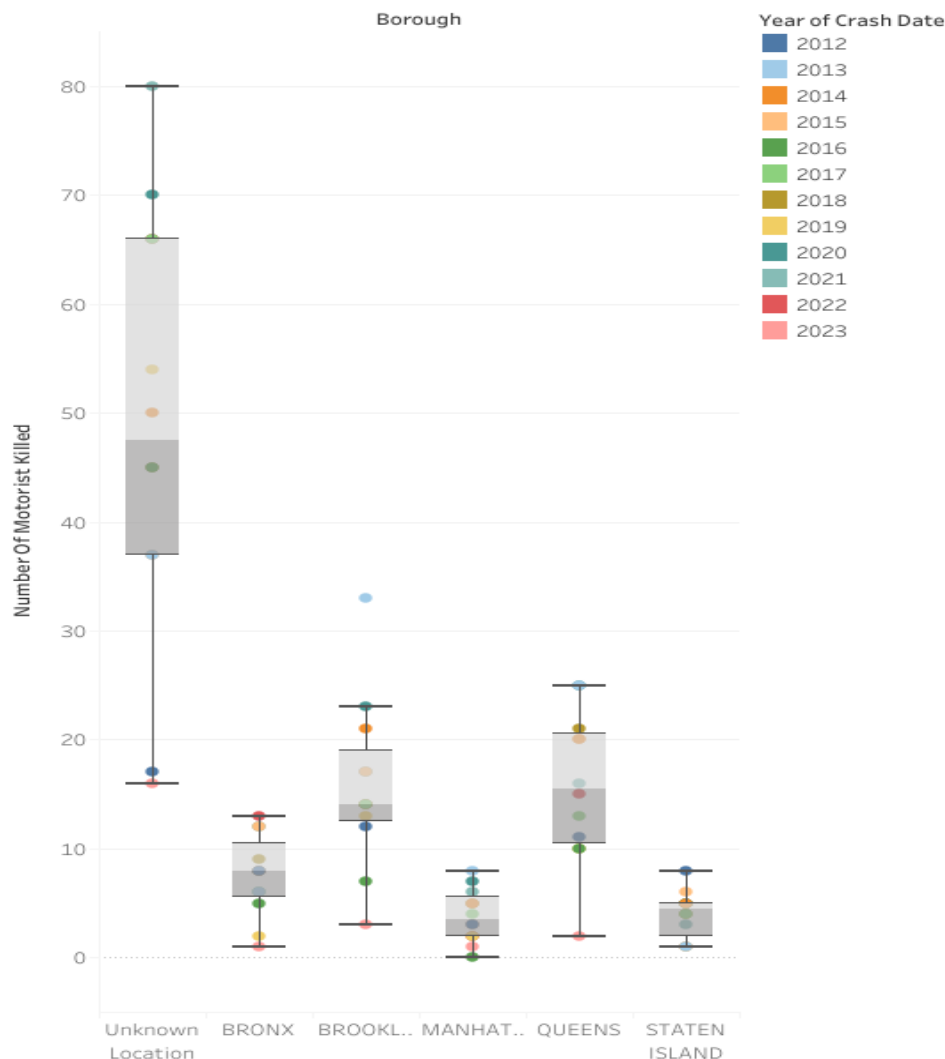
The use of traffic data on Data.gov for commercial purposes may be subject to certain restrictions, depending on the specific data set and the agency responsible for maintaining it. Users should carefully review the terms of use and licensing information for each data set before using it for any commercial purposes. Additionally, some data sets may require attribution or payment of fees for commercial use. Users should contact the responsible agency for more information on the specific requirements and restrictions.

Visualizations

The distribution of motorcycle deaths by borough and year in New York City from 2012 to 2023

- ❖ The x-axis represents the boroughs, while the y-axis represents the number of motorcycle deaths.
- ❖ Each box represents the distribution of motorcycle deaths for a specific year in a borough.
- ❖ The top and bottom edges of the box represent the upper and lower quartiles of the data, respectively.
- ❖ The horizontal line inside the box represents the median (50th percentile) of the data.
- ❖ The whiskers (vertical lines extending from the box) represent the minimum and maximum values within 1.5 times the interquartile range (IQR) of the lower and upper quartiles, respectively.
- ❖ Any data points outside of the whiskers are considered outliers and are plotted as individual points.
- ❖ The chart shows that the number of motorcycle deaths varies widely between boroughs and years, with some boroughs having consistently higher or lower numbers of deaths than others.
- ❖ https://public.tableau.com/views/DistributionofMotorcycleDeathsbyboroughYear/Sheet1?language=en-US&:display_count=n&:origin=viz_share_link

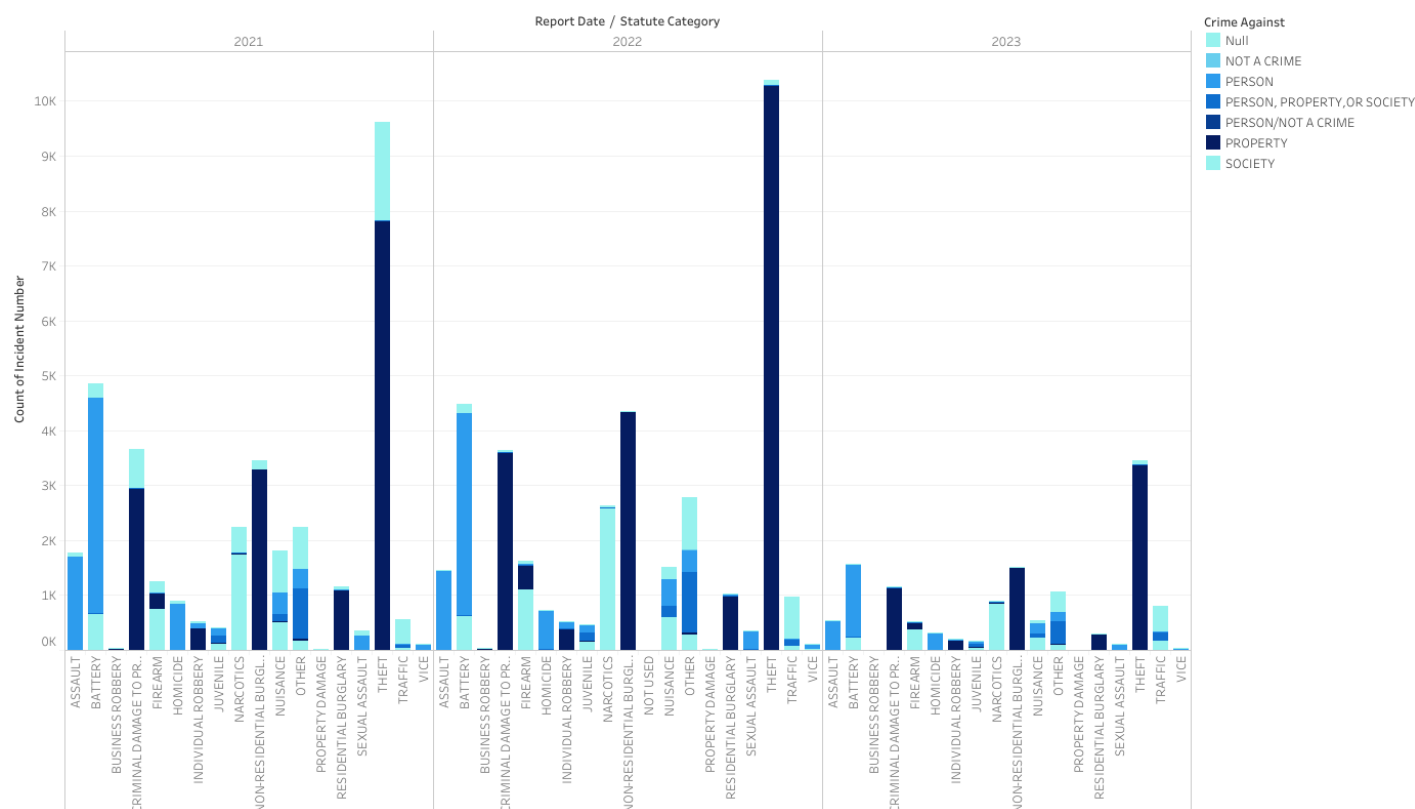
Sheet 1



Baton Rouge Crime Incidents

- ❖ The Baton Rouge Crime Incidents visualization is a stacked side-by-side bar chart that analyzes crime incidents in Baton Rouge by YEAR (Report Date), Statute Category, and CNT (Incident Number).
- ❖ By filtering the Statute Category and YEAR (Report Date), you can drill down into specific areas of interest and get insights into the types of crimes that are occurring in the area.
- ❖ The visualization's design is eye-catching and user-friendly, with different colors used to represent different types of crimes, making it easy to compare and contrast data.
- ❖ With this visualization, you can easily filter out crimes and years to tailor your analysis to meet your specific needs.
- ❖ https://public.tableau.com/views/BatonRougeCrimeIncidents/Sheet1?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

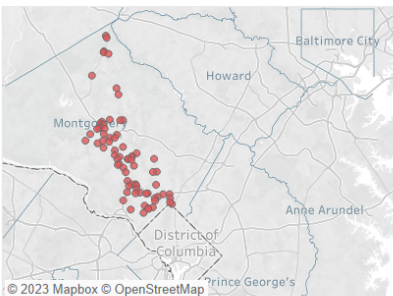
Sheet 1



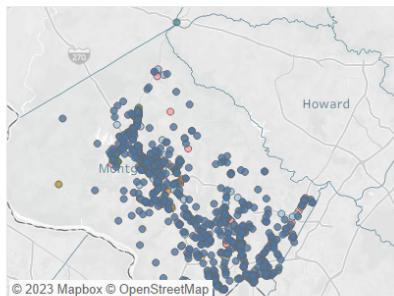
A Collection of Montgomery County Traffic Related Datasets

- ❖ These collections have 4 separate data visualizations of Traffic congestion, traffic violation incidents, crash reportings, as well as pavement conditions in Montgomery County, MD.
- ❖ The image first shown displays all 4 data visualizations side by side, while the image below displays all 4 data visualizations overlaid on top of each other, with a “map selector” bar on the side that allows the user to swap between visualizations, allowing for new analyses to be generated.
- ❖ The design is very user friendly and easy to navigate, as you can click on different data points to gain more insight and zoom in/out or move around the map.
- ❖ You can utilize the different filters to your preference, for example some filters include arrest types, accident types, injury severity, etc.
- ❖ These visualizations are useful for determining correlations between traffic congestion in Montgomery county and other factors such as pavement conditions, etc.
- ❖ https://public.tableau.com/app/profile/zaz6705/viz/TrafficVisualizations_16838390538280/Dashboard4?publish=yes
- ❖ <https://public.tableau.com/app/profile/zaz6705/viz/TrafficVisualizationsOverlaid/Dashboard5?publish=yes>

Montgomery County Traffic Congestion



Montgomery County Traffic Violations



Arrest Type (Traffic ..

- ☒ (All)
- ☒ A - Marked ...
- ☒ B - Unmarke...
- ☒ E - Marked ...
- ☒ F - Unmarke...
- ☒ G - Marked ...
- ☒ H - Unmark...
- ☒ I - Marked ...

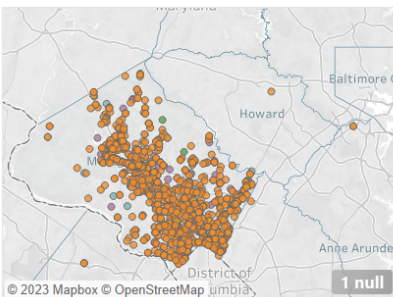
Accident (Traffic Vio..

- ☒ (All)
- ☒ No
- ☒ Yes

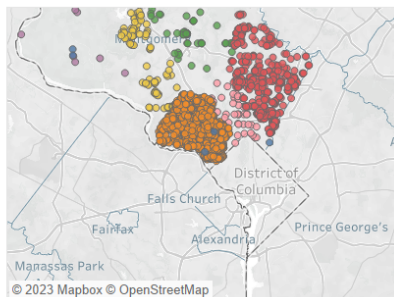
Alcohol (Traffic Viol..

- ☒ (All)
- ☒ No
- ☒ Yes

Montgomery County Crash Reportings



Montgomery County Pavement Conditions

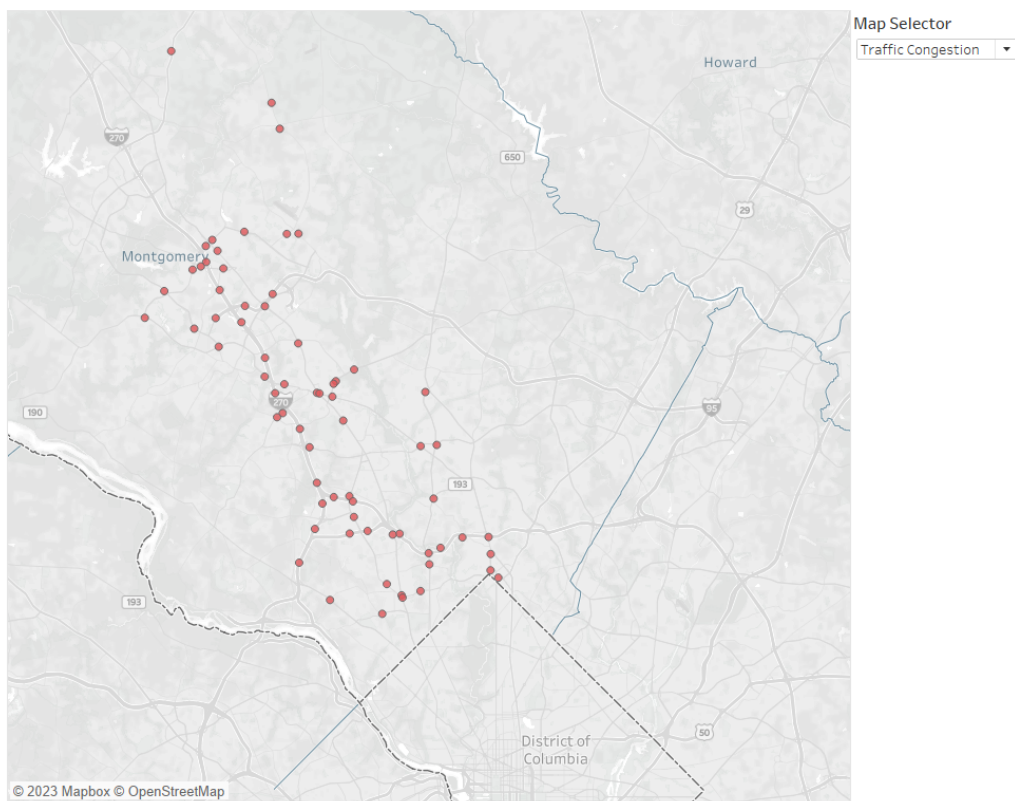


Arrest Type (Traffic ..

- ☒ G - Marked M..
- ☒ H - Unmarked..
- ☒ I - Marked Mo..
- ☒ J - Unmarked ..
- ☒ L - Motorcycle
- ☒ M - Marked (...)
- ☒ O - Foot Patrol
- ☒ Q - Marked La..
- ☒ R - Unmarked..

Injury Severity (Cra..

- ☒ Null
- ☒ NO APPARENT I..
- ☒ POSSIBLE INJURY
- ☒ SUSPECTED MIN..
- ☒ SUSPECTED SER..



Version History

Version	Date
1.0	April 2023
2.0	May 2023