

CITY OF CAMBRIDGE

OPEN DATA WORKSHOP

DECEMBER 10, 2025

Who are we?

Alix Epstein

- Systems Data Analyst
- College librarian
- Public health background
- Running, trivia, crossword puzzles

Reinhard Engels

- Data Analysis & Open Data Program Manager
- Also an ex-librarian!
- Bioinformatics visualization at the Broad Institute
- Library Technology at Harvard and Cambridge Public Library

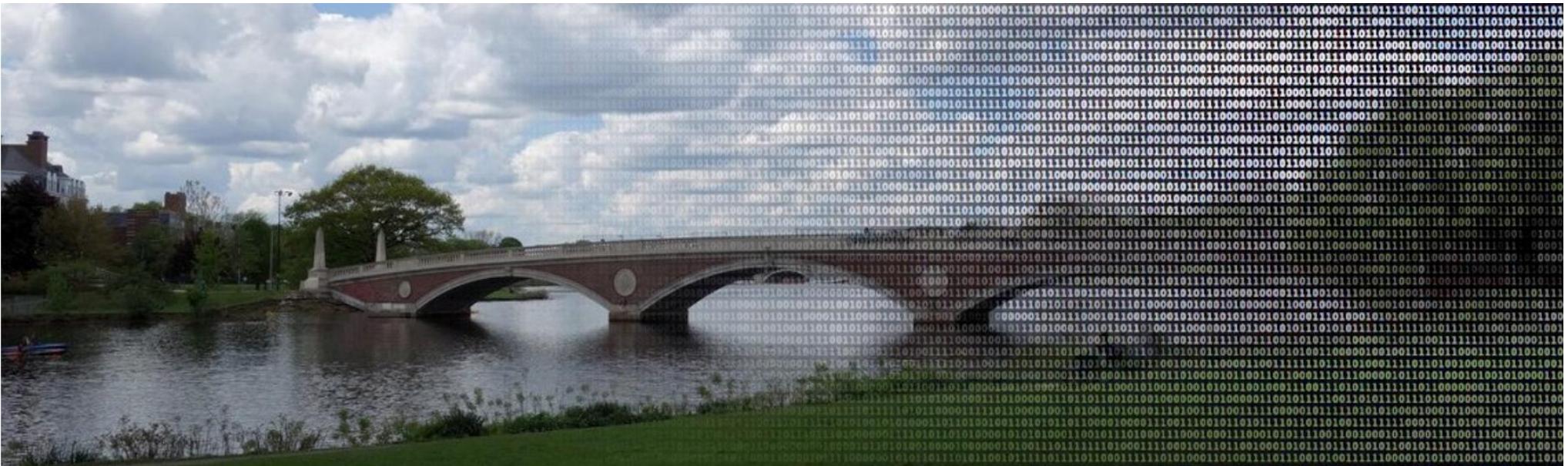
What is data?

- **Information in the form of facts and numbers, which can often be used to analyze something or make decisions**



What is open data?

- **The idea that some data should be free and accessible to everyone, to use however they want, without restrictions**



The Cambridge Open Data Program

- **A set of systems, technologies, and policies for maximizing the value of Cambridge's data by sharing it, enabling city staff and the public to use it in various ways, making sure businesses have access to it free of cost, and removing barriers between city departments in data sharing**
- **Fostering transparency, encouraging community engagement, stimulating innovation – while prioritizing privacy and data security**



The Data Portal

data.cambridgema.gov

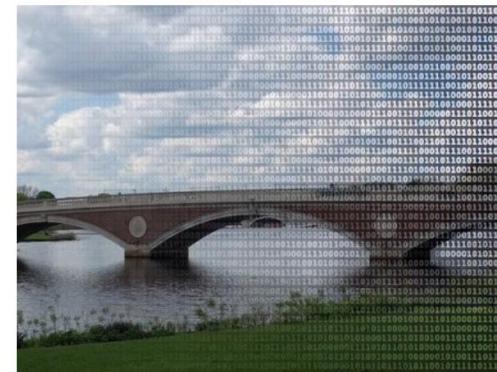
City of Cambridge

Open Data Program

Search

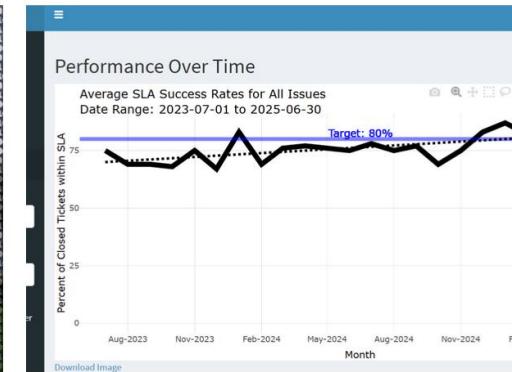
Services ▾ I Want To ▾ Departments ▾

Contact Us



Open Data Portal

Explore hundreds of free datasets from across the City.



Cambridge 311 Performance Dashboard...

It's the same tool our teams rely on every day to measure and improve how the City responds to SeeClickFix (311) service requests, and now we're giving you access ...



Cambridge Open Data Virtual Workshop...

Free, Hands-on course on Cambridge Open Data via zoom. All are welcome!

Gallery

Geographic Information Systems

Get Involved

Guides & Tutorials

Newsletter

Open Budget

Policy & Planning

Standards & Guidelines

The screenshot shows the City of Cambridge Open Data Portal interface. At the top, there is a blue header bar with the portal's name, a search bar, and various navigation icons. Below the header, a sidebar on the left contains filters for 'View Types' (Assessing, Budget/Finance, General Government, Geographic Information (GIS), Inspectional Services) and 'Tags' (Choose tag). The main content area displays 405 results for three datasets:

- Budget - Salaries- FY 2025** (Budget/Finance): Tags: No tags assigned. API Docs. Created: June 11, 2024. Views: 15.
- Outstanding Employee Awards** (General Government): Tags: No tags assigned. API Docs. Created: June 10, 2024. Views: 6.
- Certificates of Occupancy** (Inspectional Services): Tags: No tags assigned. API Docs. Created: May 31, 2024.

Use the search box or the categories or tags to look for a data set

What kinds of data are on the Open Data Portal? What formats?

"Data Asset"

- **Term for each individual result or item listed when you browse or search the portal**
- **Types of Data Asset:**
 - **Dataset**
 - **Chart**
 - **Files & Documents**
 - **Maps**
 - **Filtered Views**

What kinds of data are on the Open Data Portal? What subjects?

Category	Tag
<ul style="list-style-type: none">▪ City department producing or collecting the data▪ There can be only one for each data asset▪ Examples: Public Safety, Public Works, Traffic & Parking▪ Exceptions: there are currently a couple of non-department categories for historical reasons	<ul style="list-style-type: none">▪ Grab bag subject marker that can cut across departments or pinpoint something very specific within a department▪ Each data asset can have multiple tags▪ Examples: bicycles, permits, hvac

Metadata – data about the data

- **Description**
- **Date last updated**
- **Date created**
- **Update frequency**
- **Category**
- **Tags**
- **License**
- **Number of rows**
- **Number of columns**
- **Rows**
- **Columns**
- **Source of data**
- **Limitations**

Building Permits: Addition/Alteration Inspectional Services

List of permits related to the construction or the alteration of existing buildings in Cambridge. These permits cover construction that changes the structure, the footprint, or the volume of the building. This includes decks, porches, and additions, as well any reconstruction, renovation or repair to an existing structure. This dataset derives from...

[Read more](#)

Last Updated
June 9, 2024

Data Provided By
City of Cambridge Inspectional Services Department

Featured Content Using this Data

Choose to promote up to 3 assets that use this data, or choose to feature any relevant external resources. If left blank this section will be invisible.

[Manage Featured Content](#)

About this Dataset

Updated June 9, 2024	Civic Innovation
Data Last Updated June 9, 2024	Problem Statements Related to this Data What can the permits' application dates reveal about patterns of economic activity in the city?
Metadata Last Updated March 4, 2024	
Date Created October 30, 2018	Maintenance Plan
	Estimated Update Frequency Weekly
Views 6,172	Specific Limitations
Downloads 1,131	Limitations Only completed applications are listed in this dataset.
Data Provided by City of Cambridge Inspectional Services Department	Topics
Dataset Owner Steven Espinoza	Category Inspectional Services
	Tags construction, alteration, housing and development, permits
	Licensing and Attribution
	License Open Data Commons Public Domain Dedication and License
	Source Link https://www.cambridgema.gov/inspection.aspx

What's in this Dataset?

Rows 10.6K	Columns 86	Each row is a Permit
----------------------	----------------------	----------------------

Columns (86)

Column Name	Description	API Field Name	Data Type
# Record Number		id	Number

Looking at the data

Each row is an item (in this case, a permit)
Each column is an attribute of that item

Building Permits: Addition/Alteration

 Search X

Item ID	Number of Stories	Building Type	Gross Square Footage	Building Use	Owner-Occupied	Condo Attached	Building Construction Type	Off-Street Address	Proposed Date
1	6	25	155355	Commercial / Mixed Use	False	False	Steel/Concrete, Non-Combustible	120 Brattle St	2023-06-01
2	3	45	50550	Commercial / Mixed Use	False	False	Steel/Concrete, Non-Combustible	120 Brattle St	2023-06-01
3	4	60	125200	Commercial / Mixed Use	False	False	Steel/Concrete, Non-Combustible	120 Brattle St	2023-06-01
4	6	62	25844	Commercial / Mixed Use	False	False	Steel/Concrete, Non-Combustible	120 Brattle St	2023-06-01
5	2	0	0	Commercial / Mixed Use	False	False	Masonry / Wood (Type III)	120 Brattle St	2023-06-01
6	3	50	10000	Commercial / Mixed Use	False	False	Steel/Concrete, Non-Combustible	120 Brattle St	2023-06-01

Creating a visualization

Configure Visualization X

Data Selection ▼

Dimension ?
Select Column... ▼

Measure ?

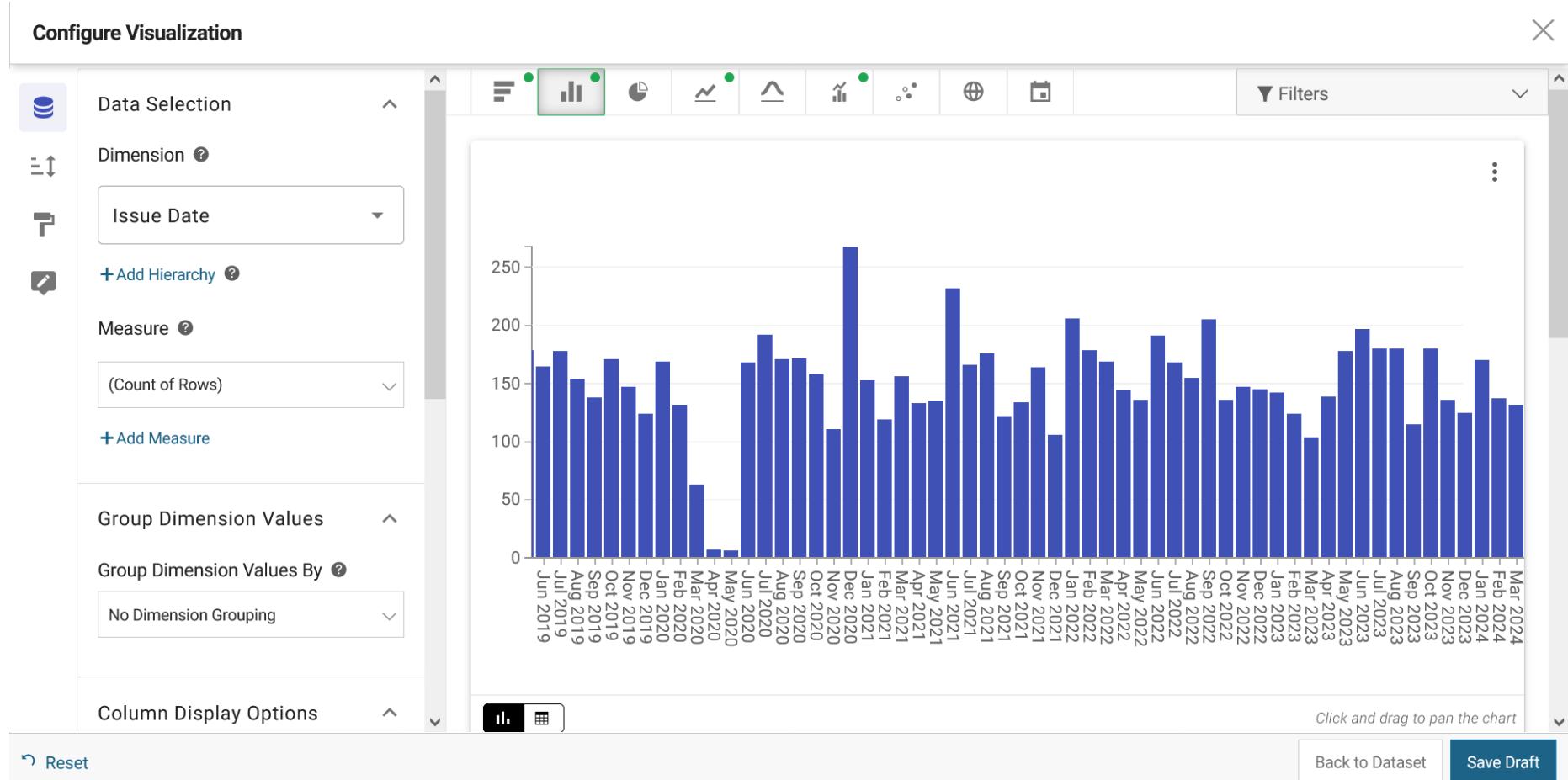
▼ Filters ▼

Select data or a chart type to get started

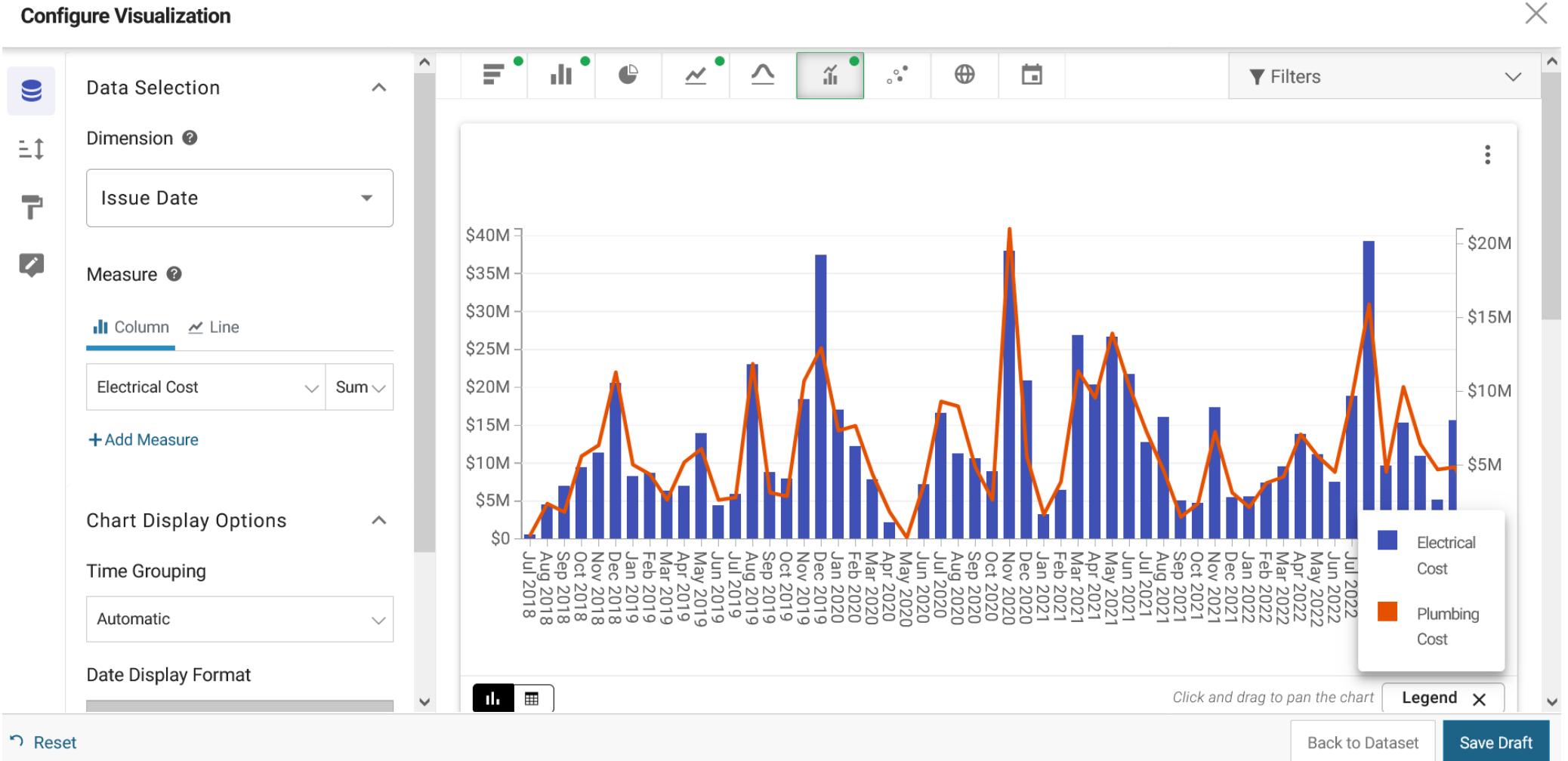
There are two ways to get started. Select a chart type to see recommended dimensions and measures or select a dimension or measure to see recommended chart types.

↻ Reset Back to Dataset Save Draft

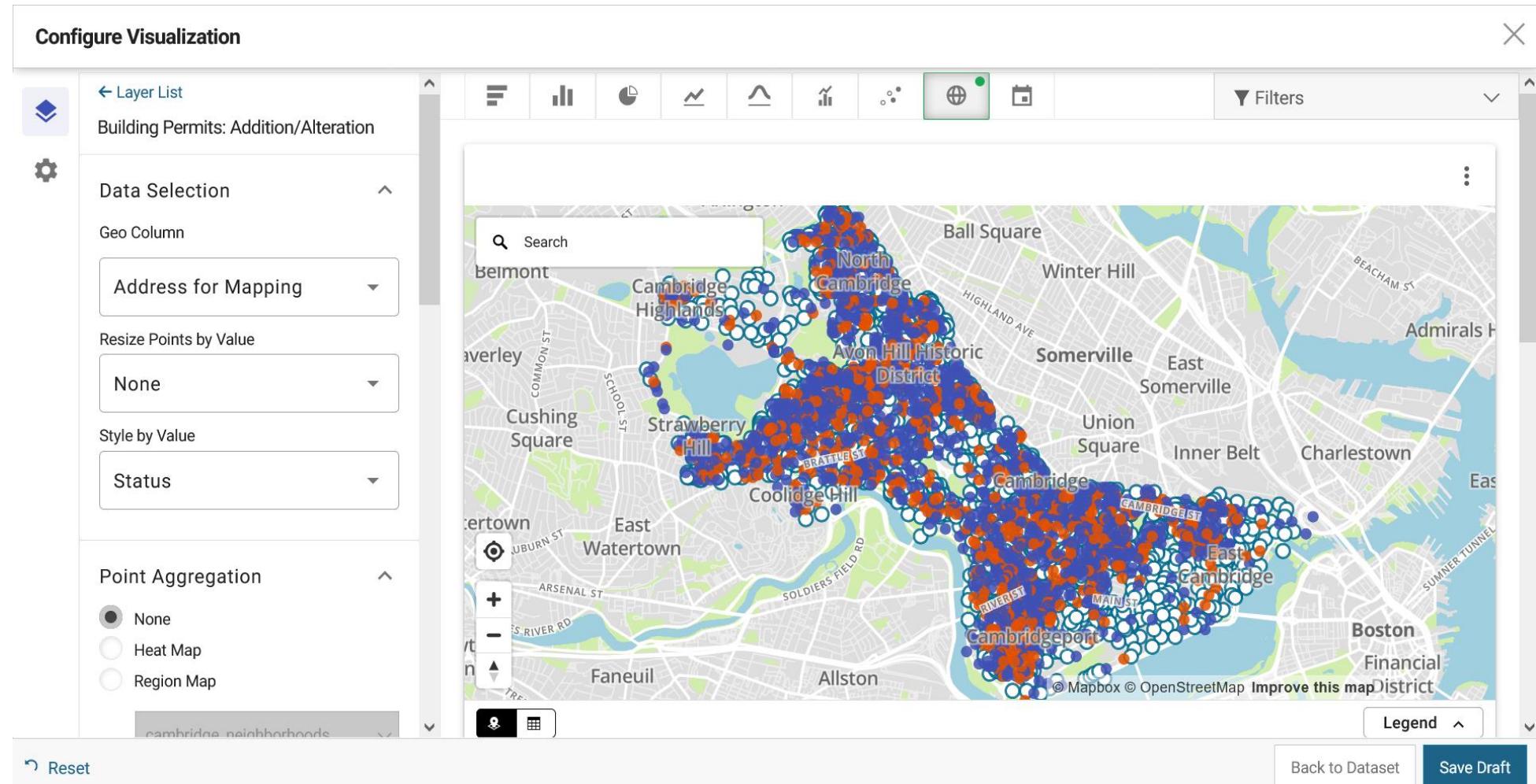
Bar chart



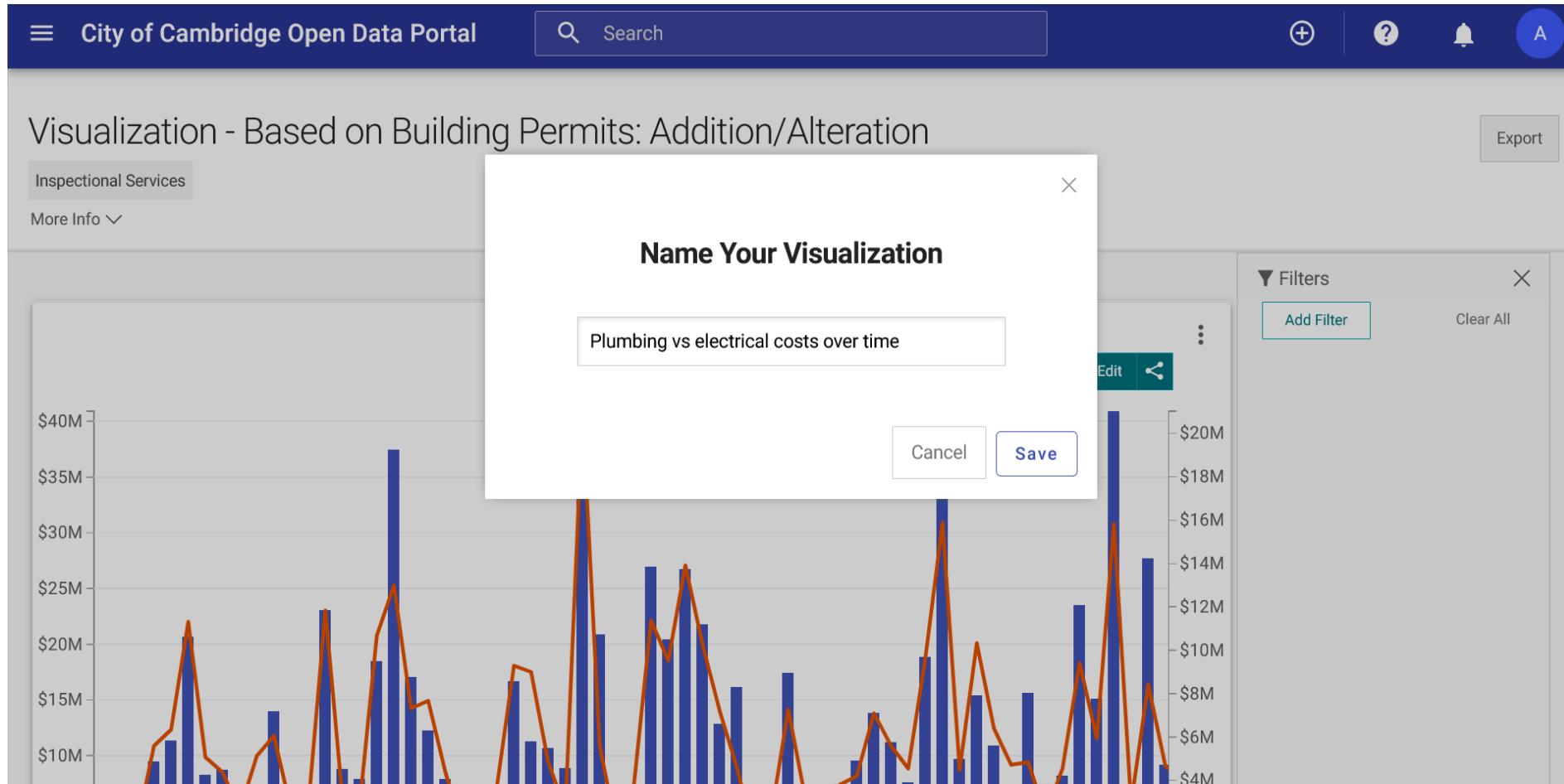
Combination chart



Map



Saving a visualization



Publish for private use or publicly

The screenshot shows the City of Cambridge Open Data Portal interface. At the top, there is a navigation bar with the portal's name, a search bar, and various user icons. Below the navigation bar, a chart titled "Plumbing vs electrical costs" is displayed. The chart has two data series: blue bars representing one metric and orange lines representing another. The y-axis is labeled with monetary values from \$15M to \$40M. A modal dialog box is overlaid on the chart, titled "Publish this asset". It asks the user to choose the audience for the published version. Two options are shown: "Private" (selected, indicated by a checked checkbox) and "Public". The "Private" option includes a lock icon and the text "Only individual collaborators and certain site roles can access". The "Public" option includes a globe icon and the text "The public can view". At the bottom of the dialog are "Cancel" and "Publish" buttons. To the right of the chart, there are "Save", "Publish...", "Export", and filter options ("Filters", "Add Filter", "Clear All").

Querying (filtering) the data

Building Permits: Addition/Alteration | Public | Save As | Edit | ⋮

← Back to Overview | Switch to Grid View | Refresh | Search | Export

...	# Build... buildin... ...	# Elect... electr... ...	# Plum... plum... ...	# Gas ... gas_... ...	# HVA... hvac... ...	# T... t... ...	# Final... final... ...	# Fire ... fire_p... ...	T Desc... detail... ...	T Size ... size_... ...	✓ Chan... floor... ...	✓ Chan... exter...
	\$3,119,225	\$1,063,837	\$197,130	\$0	\$973,730	\$5,700,292		\$346,370	Interior renov...	0	False	False	
	\$170,000	\$36,000	\$35,000	\$5,000	\$35,000	\$296,000		\$15,000	Frame new i...	0	False	True	
	\$28,000	\$0	\$0	\$0	\$0	\$28,000		\$0	Repair front ...	0	False	False	
	\$26,100	\$0	\$0	\$0	\$0	\$26,100		\$0	Demoing offi...	0	False	False	
	\$13,770	\$0	\$0	\$0	\$0	\$13,770		\$0	WEATHERIZ...	0	False	False	

< 1 of 107 > Showing rows 1-100 of 10628

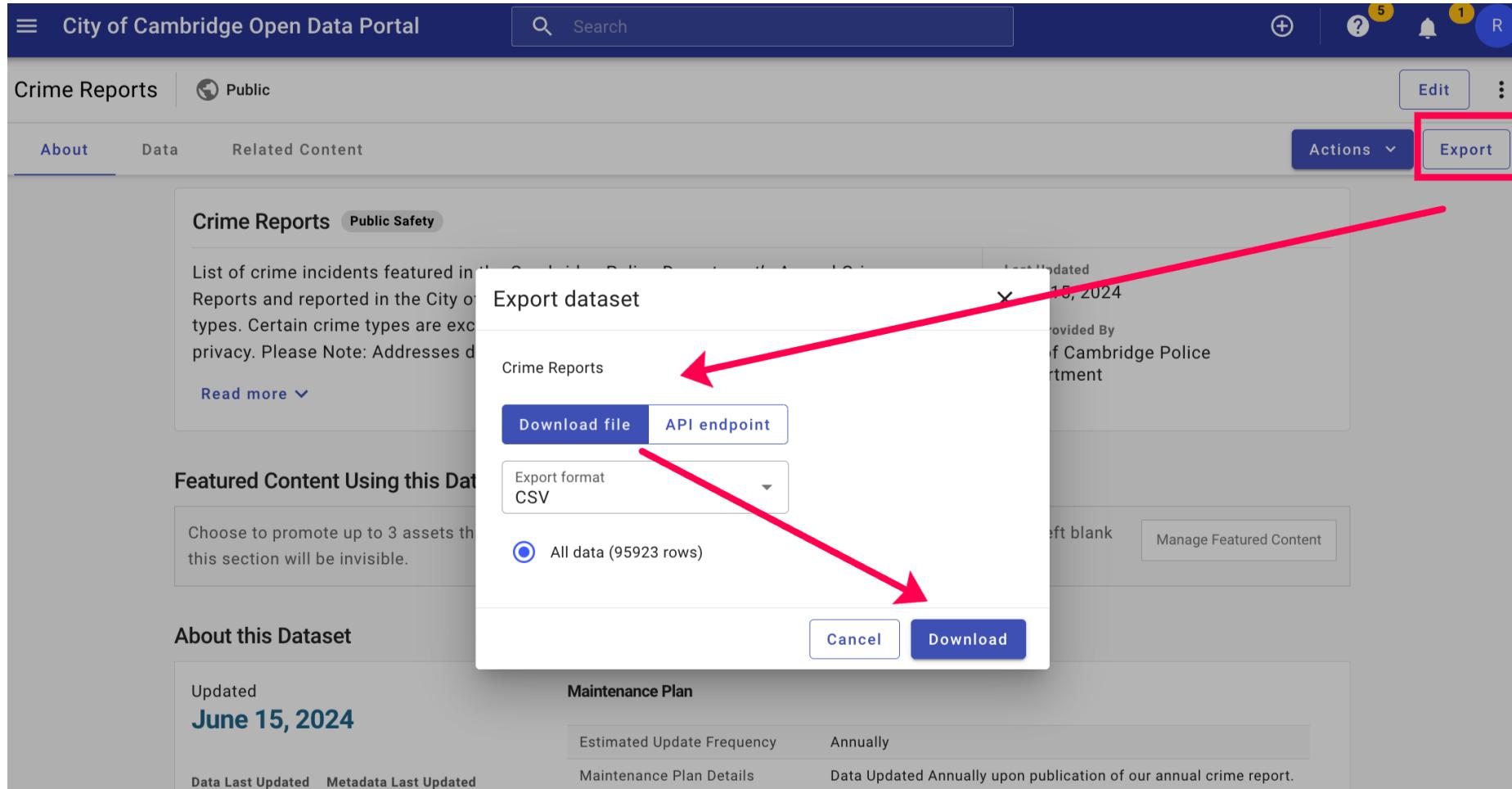
Filters | Clear all

Total Cost : is greater than or equal to 5000

+ AND

Apply

Exporting Data: use Cambridge data with other tools



oData: live connections for excel, Tableau, PowerBI)

Dogs of Cambridge General Government

This dataset displays the name, breed, and approximate location of dogs in Cambridge. It is based on dog license data collected by Cambridge's Animal Commission. All locations listed in this dataset have been obscured to protect privacy. Please see the limitations section below for more information.

Last Updated April 14, 2025

Data Provided By City of Cambridge Animal Commission

Featured Content Using this Data

- Dogs of Cambridge Data Story Public November 3, 2023 26.7K Views
- Top Dog Names in Cambridge Public April 14, 2025 1,276 Views

Based on the Dogs of Cambridge open dataset, which displays the name, breed, and approximate location of dogs in Cambridge.

About this Dataset

Updated April 14, 2025

Civic Innovation

Problem Statements Related to this Data Where should new dog parks be located?

Data Last Updated April 14, 2025 Metadata Last Updated September 23, 2024

Maintenance Plan

Estimated Update Frequency Annually

tableau public Create Learn

Register now for Tableau Conference 2025 →

Cambridge dog names by [Alexandra Epstein](#)

Dog name popularity

The chart is a bubble plot where the size of each blue circle represents the popularity of a dog name. Some names are labeled: Milo, Penny, Bella, Charlie, Lucy, Rosie, and Luna.

APIs: use Cambridge data programmatically

Export dataset X

Crime Reports

[Download file](#) [API endpoint](#)

Data format [JSON](#)

All data (95923 rows)

⚠ Default API limit exceeded
Our API has a default limit of providing 1,000 rows. [Learn more](#) about how you can modify the default limit.

API endpoint <https://data.cambridgema.gov/resource/xuad-73uj.j>

[API documentation](#) [Developer portal](#)

[Cancel](#) [Copy to clipboard](#)

JSON Raw Data Headers

Save Copy Collapse All Expand All (slow) Filter JSON

0:

```
file_number: "2024-04574"
date_of_report: "2024-05-31T22:04:00.000"
crime_date_time: "05/31/2024 22:00 - 22:03"
crime: "Aggravated Assault"
reporting_area: "1102"
neighborhood: "North Cambridge"
location: "0 CAMERON AVE, Cambridge, MA"
```

1:

```
file_number: "2024-04572"
date_of_report: "2024-05-31T22:02:00.000"
crime_date_time: "05/31/2024 18:00 - 21:00"
crime: "Larceny of Bicycle"
reporting_area: "701"
neighborhood: "Riverside"
location: "0 PLEASANT ST, Cambridge, MA"
```

2:

```
file_number: "2024-04573"
date_of_report: "2024-05-31T21:33:00.000"
crime_date_time: "05/31/2024 21:32 - 22:21"
crime: "Missing Person"
reporting_area: "504"
neighborhood: "Cambridgeport"
location: "200 FRANKLIN ST, Cambridge, MA"
```



<https://github.com/Chicago/RSocrata>



<https://github.com/socrata/socrata-py>

AI Tools: analyze Cambridge data with ChatGPT

 Police_Department_Crash_Data_-_Updated_20... CSV

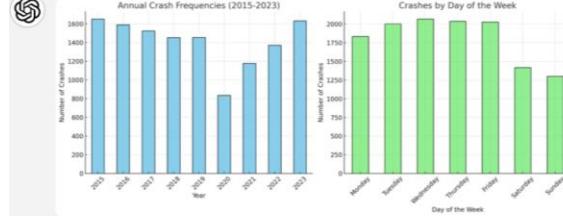
Analyze this Cambridge Police Department crash data. Do you notice any interesting or surprising trends? Please create charts to help me understand visually. Please exclude data from 2024 since it is incomplete.

The Cambridge Police Department crash data includes a vast array of columns, totaling 159, covering a wide range of details for each incident. Notable columns include:

- **Date Time:** The date and time of the incident.
- **Day of Week:** The day of the week on which the incident occurred.
- **Object 1 & Object 2:** Descriptions of the objects involved in the crash.
- **Street Number & Street Name:** The location of the incident.
- **May involve cyclist, May involve pedestrian, etc.:** Indicators of whether the crash involved cyclists, pedestrians, etc.
- Plus many more detailed attributes regarding the circumstances of each crash, involved parties, and outcomes.

Given the extensive nature of this dataset, a productive approach would be to identify key trends over time, such as annual crash frequencies, the involvement of cyclists or pedestrians, or crashes by day of the week. Since we should exclude data from 2024 due to its incompleteness, I will first filter the data accordingly and then proceed with analyzing trends and visualizing the results. [1]

Given these insights, I will now create visualizations for the annual crash frequencies and crashes by day of the week to provide a clearer understanding of these trends. The visualization for crashes involving cyclists and pedestrians will not be provided due to the unclear nature of the relevant data in the dataset. [2]



The visualizations offer a clear depiction of the identified trends in the Cambridge Police Department crash data:

1. **Annual Crash Frequencies (2015-2023):** This chart shows the number of crashes per year, highlighting the significant dip in 2020. Following this dip, there's a noticeable rebound in crash numbers, approaching pre-pandemic levels by 2023.
2. **Crashes by Day of the Week:** The frequency of crashes is higher on weekdays, with a peak on Wednesday, followed closely by Thursday and Friday. The number of crashes decreases notably on weekends, with Sunday having the fewest crashes.

These trends suggest a strong correlation between traffic patterns and the occurrence of crashes, with factors such as the COVID-19 pandemic affecting annual variations and typical weekly routines influencing daily fluctuations. [3]

AI Tools: "chat up" an interactive dashboard

This is great. Can you create a dashboard showing some of these visualizations with filter controls that make sense?

Stopped thinking

Looks like we had a network issue. Can you try again to create a dashboard showing some of these visualizations with filter controls that make sense? Can you put it in the canvas?

Thought for 44s

Rodent Sightings Explorer (csv-powered Dashboard)

All set—I dropped an interactive dashboard into the canvas. It's CSV-powered and fast:

- Upload your latest "Rodent Sightings" CSV (the one you shared works).
- Use the left-side filters (date range, status, type, city, hour window, days of week, keyword search).
- The right-side charts update instantly: monthly trend, seasonality, day-of-week, hour-of-day, plus a map-free lat/lon heatmap and a "top hotspots" table (lat/lon rounded to ~100–150m).
- Hit "Download filtered CSV" to export any slice.

Nerd notes, because we like those:

- The app auto-detects sensible column names (e.g., `ticket_created_date_time`, `ticket_status`, `issue_type`, `lat`, `lng`).
- Spatial density uses a binned SVG heatmap—no external basemap, so it's privacy-friendly and works offline. Use the "bins" slider to change resolution.
- KPI row shows total requests for your current filters and a quick "last 12 vs. prior 12 months" comparison.

If you want, I can add: neighborhood/ZIP grouping, a base-map toggle, or a "compare years" mode (aligned by month) to see whether 2025 is diverging from earlier seasonal patterns.

Do you like this personality?

Ask anything

ChatGPT can make mistakes. Check important info.

Rodent Sightings Explorer

Filters

Total requests (current filters) **4,626** Last 12 months **493** vs prior 12 months **-4.6%**

Monthly requests over time

Seasonality (by calendar month)

By day of week

By hour of day

Spatial density (lat/lon heatmap)

Top grid hotspots (lat, lon rounded to 3 decimals)

Rank	Grid	Requests
1	42.396,-71.130	47
2	42.368,-71.097	44
3	42.367,-71.096	40
4	42.365,-71.095	38
5	42.372,-71.098	37
6	42.363,-71.113	35
7	42.362,-71.105	34
8	42.369,-71.108	32
9	42.368,-71.096	32
10	42.375,-71.110	30
11	42.365,-71.093	30
12	42.368,-71.099	26

Binned heatmap (no base map). Darker = more requests. Bounds: lat 42.354–42.403, lon -71.159–71.069.

Thank you!

- You will receive an email link to these slides and a feedback form
- Feedback form: do it now! QR code and link
 - Anonymous
 - What you thought of this class
 - Suggestions
- Email us: OpenData@CambridgeMA.gov

