

**CITY OF CAMBRIDGE**

# **VISUALIZING CAMBRIDGE DATA**

---

FEBRUARY 25, 2026

# What we'll cover

---

1. Core visualization concepts
2. Basic types of visualization (Socrata, Google)
3. Maps (Socrata, Google)
4. Dashboards (Tableau)
5. AI Tools (TBD)

Except for section 1, all of our examples will be real Cambridge Data available on the Open Data Portal

# 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

# 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**



## Why Visualize?

---

"Why should we be interested in visualization? Because **the human visual system is a pattern seeker** of enormous power and subtlety."

-- Colin Ware (leading researcher on visualization and perception)

"The best graphic operations are those carried out by the viewer themselves, **in a single instant of perception.**"

-- Jacques Bertin (the father of modern data graphics theory)

**"Graphics reveal data."**

-- Edward Tufte (statistician and author of *The Visual Display of Quantitative Information*)

# Anscombe's Quartet: Raw Data

Dataset I	Dataset II	Dataset III	Dataset IV
-----------	------------	-------------	------------

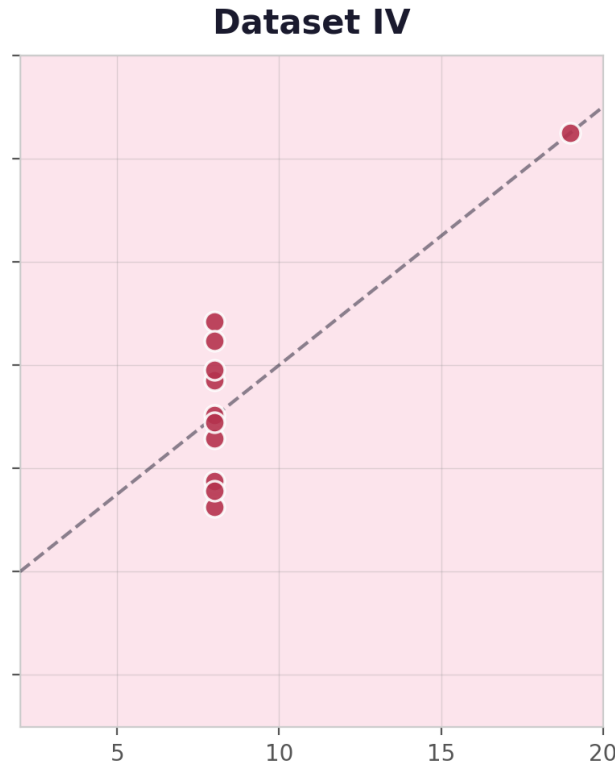
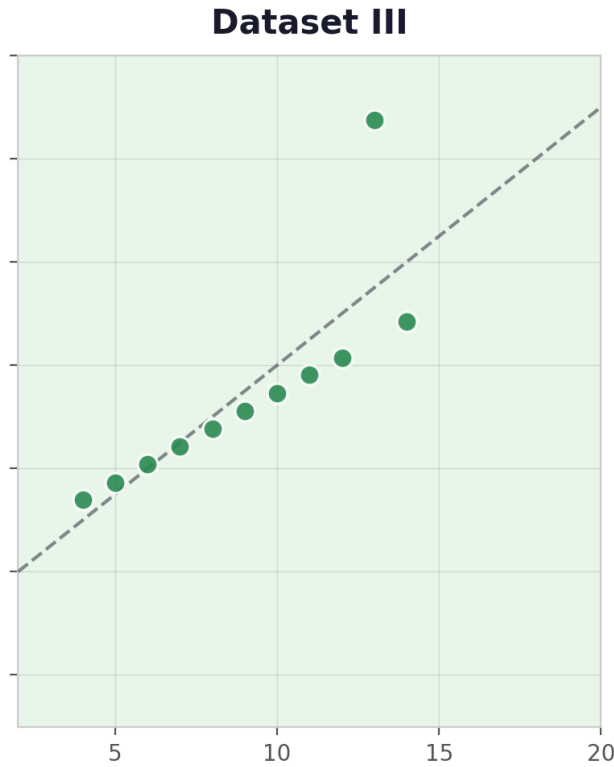
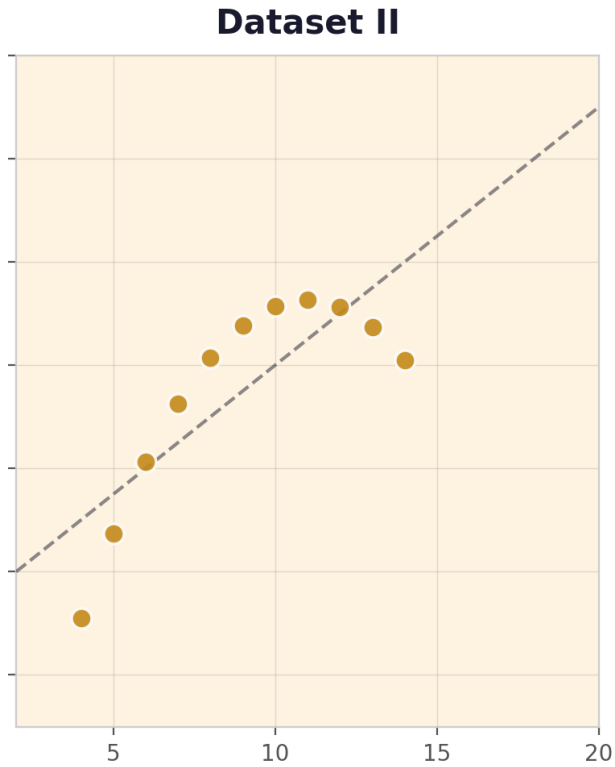
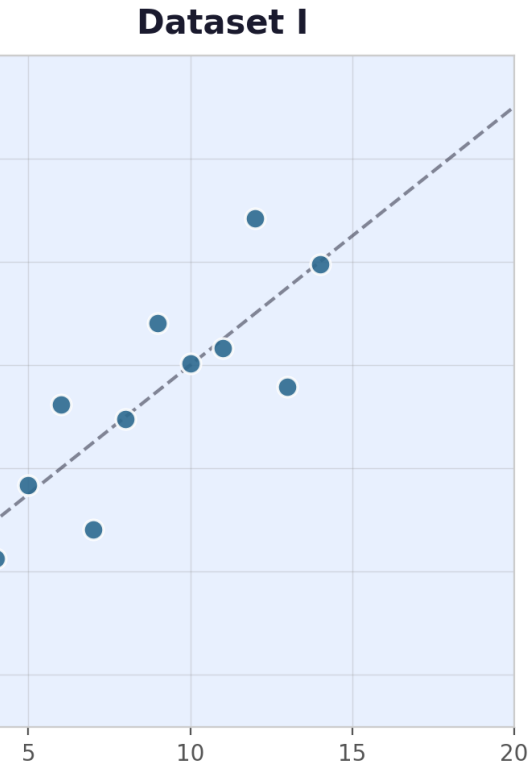
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

# Anscombe's Quartet: Summary Statistics

---

	Dataset I	Dataset II	Dataset III	Dataset IV
<b>Mean of x</b>	9.00	9.00	9.00	9.00
<b>Mean of y</b>	7.50	7.50	7.50	7.50
<b>Variance of x</b>	11.00	11.00	11.00	11.00
<b>Variance of y</b>	4.13	4.13	4.12	4.12
<b>Correlation</b>	0.816	0.816	0.816	0.817
<b>Regression (y =)</b>	$3.0 + 0.5x$	$3.0 + 0.5x$	$3.0 + 0.5x$	$3.0 + 0.5x$

# Anscombe's Quartet: Visualized



## Two Modes of Visualization: Exploratory vs. Explanatory

---

**Exploratory:** What is the story? Is there even a story?

Your role: Detective

"The greatest value of a picture is when it forces us to notice what we never expected to see."

— John Tukey, statistician who coined the term "Exploratory Data Analysis"

**Explanatory:** You know the story. Now make it unmissable.

Your role: Storyteller

"Clutter and confusion are not attributes of data — they are shortcomings of design."

— Edward Tufte



SOUTH ROW  
MARSH ROW

DUFOURS PLACE

BROAD STREET

PUMP

CAMBRIDGE ST

NEW STREET

BREWERY

HOPKINS

HUSBAND ST

WALDRFIELD COURT

STREET

EDWARD ST

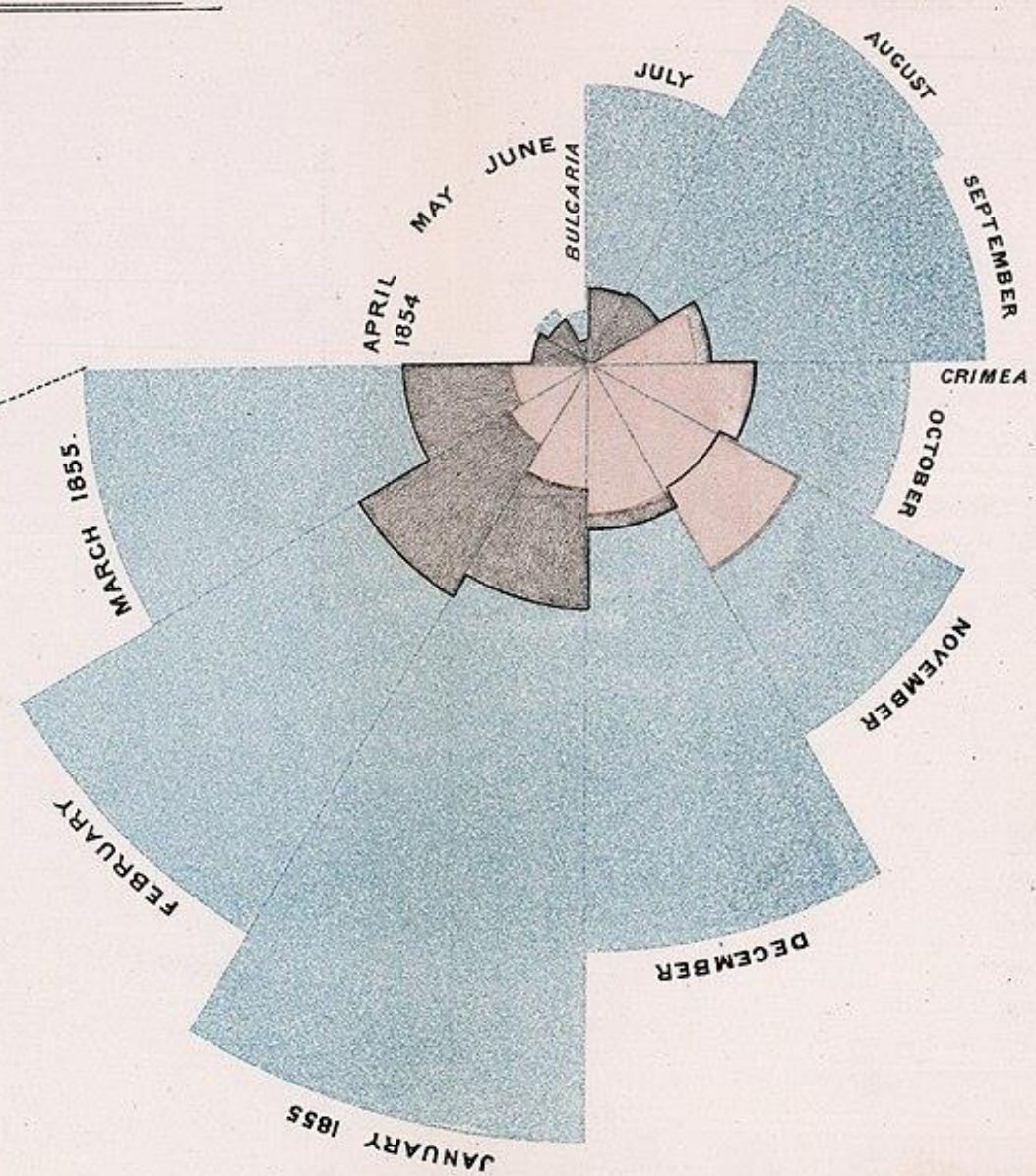
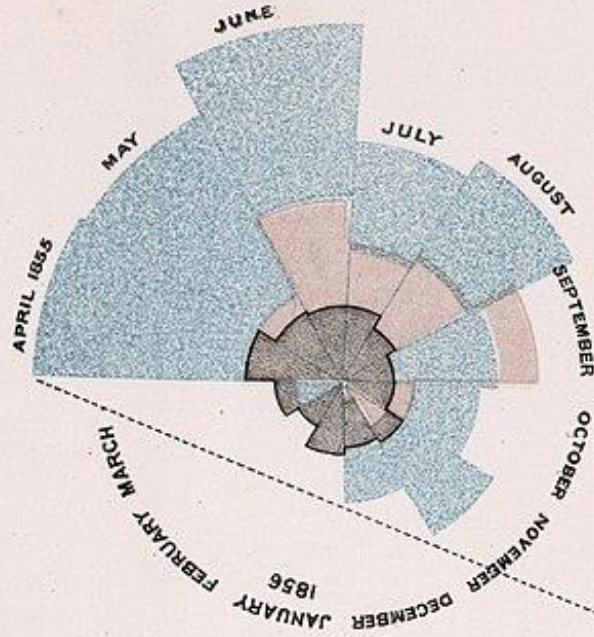
DUCK LANE

LITTLE

# DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.  
APRIL 1855 TO MARCH 1856.

1.  
APRIL 1854 TO MARCH 1855.



*The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.*

*The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.*

*The black line across the red triangle in Nov<sup>r</sup> 1854 marks the boundary of the deaths from all other causes during the month.*

*In October 1854, & April 1855, the black area coincides with the red; in January & February 1855, the blue coincides with the black.*

*The entire areas may be compared by following the blue, the red & the black lines enclosing them.*

# Putting it Together: The Workflow

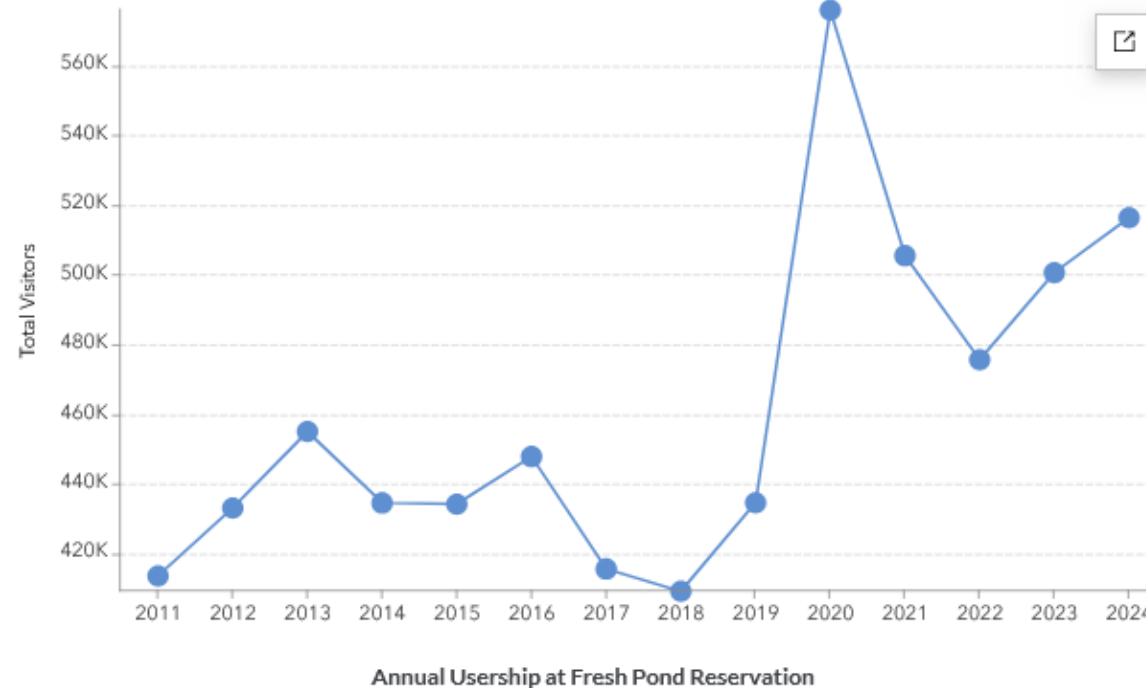
	Exploratory	Explanatory
<b>Audience</b>	You (or your team)	Someone else — public, leadership, council
<b>Title</b>	Descriptive: "311 Requests by Month"	A finding: "311 Requests Spike Every Spring"
<b>How many charts?</b>	Many — try everything	One or two — the best ones
<b>Color</b>	Defaults are fine	Highlight what matters, gray out the rest
<b>Clutter</b>	Tolerable — you need the detail	Remove ruthlessly — every pixel earns its place
<b>Labels &amp; annotations</b>	Minimal — you know the context	Add them — your audience doesn't know
<b>Polish</b>	Low — this is a sketch	High — this is a presentation
<b>Tool</b>	Whatever's fast — AI, Sheets, Socrata	Whatever looks good — Sheets, Tableau, polished AI
<b>Goal</b>	"What's in here?"	"Here's what I found."

# Line chart

## Annual Summary

CWD has observed the following annual trends:

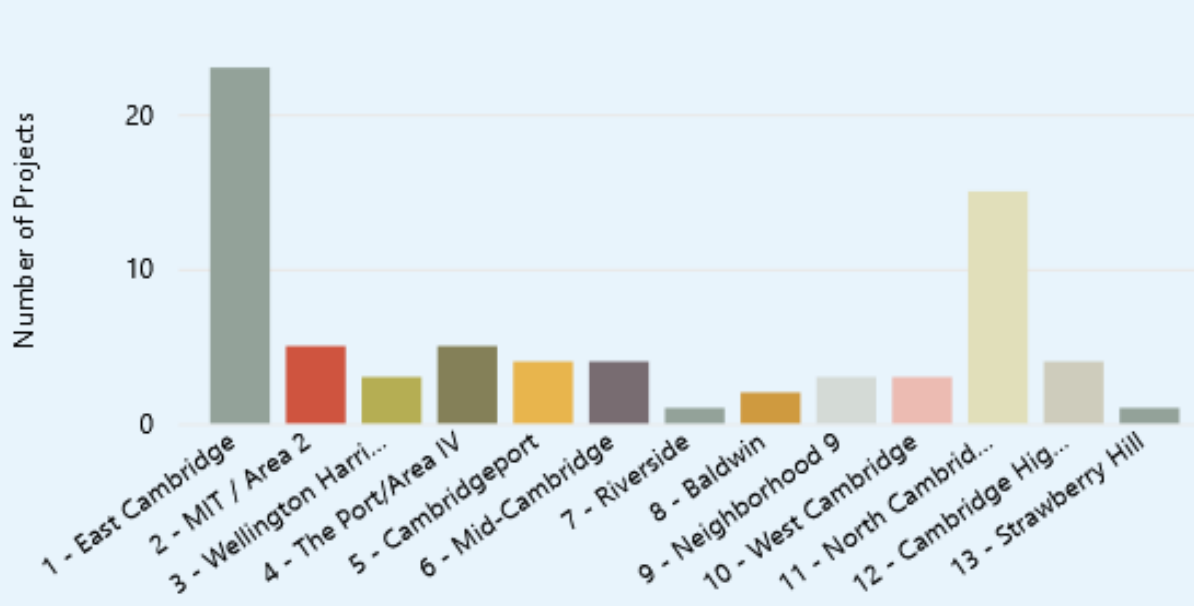
- **Before 2020**, an average of 431,400 people visited FPR each year.
- **During 2020**, usership jumped to over 576,300 people. This was the first year of the COVID-19 pandemic.
- **After 2020**, usership decreased. Despite this decrease, annual visitor totals remain higher than the pre-pandemic average.



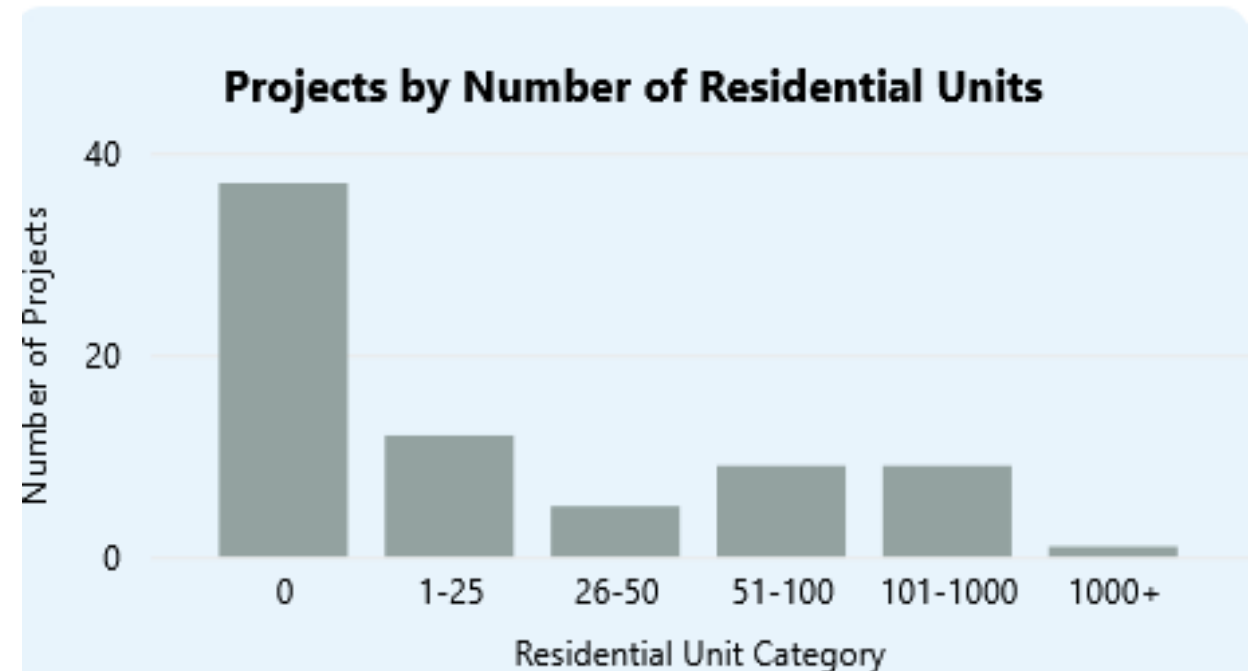
Fresh Pond Census Report  
<https://storymaps.arcgis.com/stories/29629c0df3ed47ffae5269752f79a0ec>

# Bar charts and histograms

### Projects by Neighborhood



### Projects by Number of Residential Units

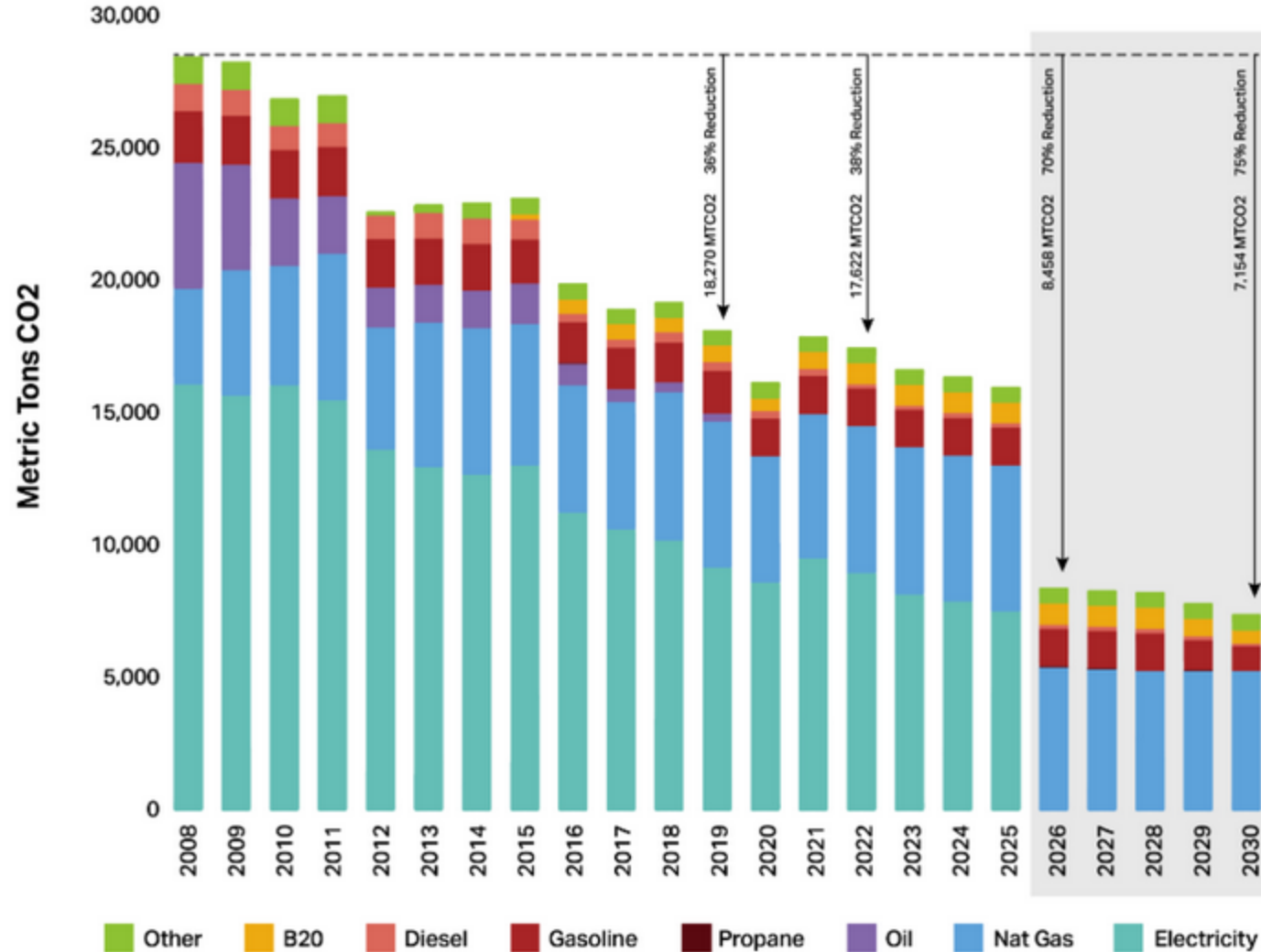


Cambridge Development Log

<https://app.powerbigov.us/view?r=eyJrIjoiNThlNjZjYzUtODEwZS00MzBiLWFlOWQzZjQyNTMzMzNIY2E5IiwidCI6ImMwNmE4YmU3LTg0NzktNGQ3My1iMzUxLTkzYmM5YmE4Mjk1YyJ9>

# Stacked bar chart

City of Cambridge Municipal Operations GHG Emissions  
Actual and Projected



Cambridge Sustainability Dashboard  
<https://sustainabilitydashboard.cambridgema.gov/category/greenhouse-gas-emissions>

# Scatter

## Yes, the bike crash rate has steadily declined since 2004 but remained stable since the COVID-19 Pandemic.

The citywide annual bike crash rate declined from about 21 to 12 bike crashes per million BMT in 2004 and 2024, respectively (see Figure 27). That's because estimated BMT has increased while reported bike crashes have decreased (see Figure 28). Estimated BMT increased 233% since 2004, growing from 3.7 million miles traveled to 12.8 million miles traveled in 2024. Reported bike crashes peaked in 2012, then dropped in the years leading up to the COVID-19 pandemic. Although reported bike crashes have increased since the COVID-19 pandemic, they remain below pre-pandemic peaks. This has resulted in more stable bike crash rates in recent years despite continued growth of the estimated BMT.

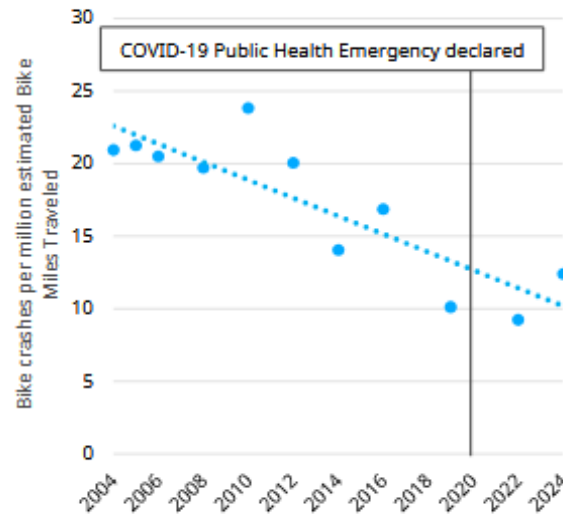


Figure 27: Citywide bike crash rate (2004-2024)

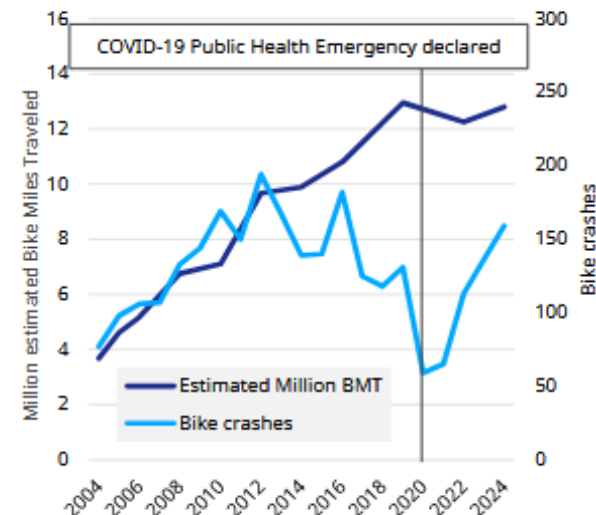


Figure 28: Reported bike crashes and estimated Bike Miles Traveled (2004-2024)

Biking in Cambridge: Data Report 2025

[https://www.cambridgema.gov/-/media/Files/CDD/Transportation/Bike/bikereports/Biking\\_in\\_Cambridge\\_Data\\_Report\\_2025\\_Final.pdf](https://www.cambridgema.gov/-/media/Files/CDD/Transportation/Bike/bikereports/Biking_in_Cambridge_Data_Report_2025_Final.pdf) (page 41)

# Stacked area chart

## How has Cambridge's bike network evolved over time?

### The bike network has grown steadily since 2004.

The City, State, and partners have greatly increased [bikeway](#) lane-miles over the past several decades. Total lane-miles of bikeways have more than doubled between 2004 and 2024, creating a more connected and accessible bike network (see Figure 21). As shown in the figure, biking has grown significantly at the same time. Readers can view the evolution of the bike network in the [Citywide Bicycle Count Data portal](#) by first selecting "Data Viewer" and then the "Map" tab.

### The bike network is becoming more comfortable for people of all ages and abilities.

Physically separated bikeways, like separated bike lanes and shared-use paths, appeal to people of all ages, abilities, and identities. The Massachusetts Institute of Technology pioneered separated bike lanes in 2003 with the reconstruction of Vassar St. Since 2017 with the completion of the Cambridge Street Bicycle Safety Demonstration Project, the City has invested in separated bike lanes and, as a result, this type of bikeway continues to make up an increasingly larger share of total bike network lane-miles (see Figure 21). At the same time, standard bike lanes are playing a smaller role in the bike network and their lane-miles peaked in 2019. That's because many separated bike lane projects, for example Hampshire St, replaced standard bike lanes.

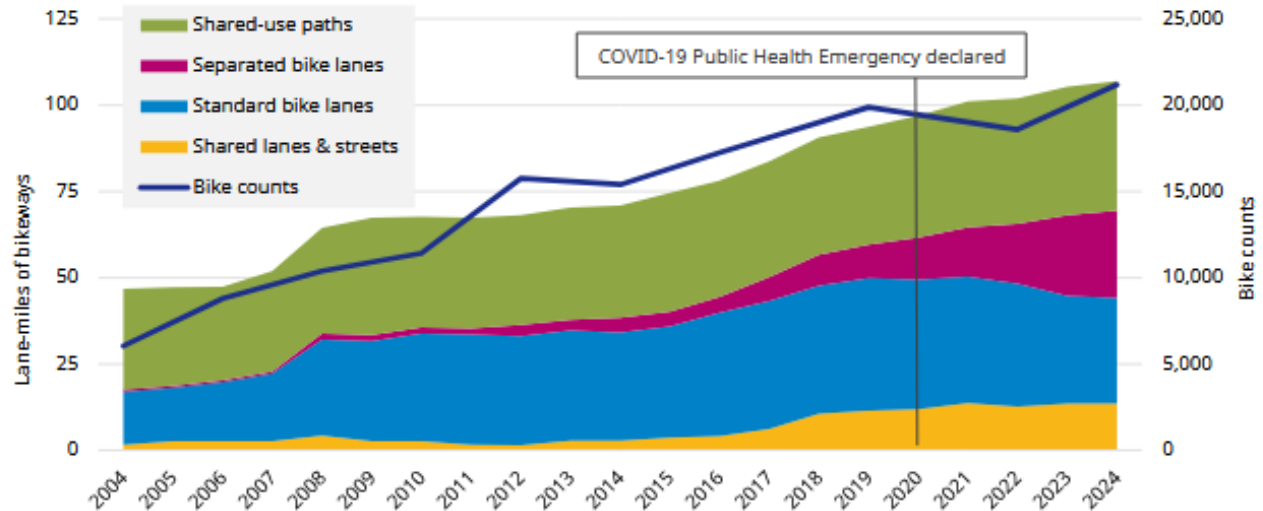


Figure 21: Bikeway lane-miles and bicyclists counted by the Citywide Bicycle Count Program during morning and evening commutes (2004–2024)

Biking in Cambridge: Data Report 2025

[https://www.cambridgema.gov/-/media/Files/CDD/Transportation/Bike/bikereports/Biking\\_in\\_Cambridge\\_Data\\_Report\\_2025\\_Final.pdf](https://www.cambridgema.gov/-/media/Files/CDD/Transportation/Bike/bikereports/Biking_in_Cambridge_Data_Report_2025_Final.pdf) (page 31)

# Geographic Data



## Crime Reports

File Number	Date of Report ↓	Crime Date Time	Crime	Reporting Area	Neighborhood	Location	Reporting Area Lat	Reporting Area Lon	Reporting Area Location
2026-00863	2026 Jan 31 09:04:00 PM	01/31/2026 19:00 -	Hit and Run	611	Mid-Cambridge	0 Quincy St, Cambridge, MA	42.3743	-71.1164	POINT (-71.11637311 42.37433565)
2026-00861	2026 Jan 31 07:52:00 PM	01/29/2026 22:00 -	Missing Person	803	Baldwin	0 Wendell St, Cambridge, MA	42.3808	-71.1181	POINT (-71.11812104 42.38077156)
2026-00858	2026 Jan 31 06:45:00 PM	01/31/2026 16:00 -	Mal. Dest. Property	1101	North Cambridge	0 White St, Cambridge, MA	42.3897	-71.1189	POINT (-71.11894512 42.38970057)
2026-00857	2026 Jan 31 06:01:00 PM	01/28/2026 19:00 -	Mal. Dest. Property	507	Cambridgeport	100 Pleasant St, Cambridge, M	42.3629	-71.1087	POINT (-71.10874348 42.36286482)
2026-00856	2026 Jan 31 04:35:00 PM	01/30/2026 23:30 -	Hit and Run	1014	West Cambridge	200 Lexington Ave, Cambridge,	42.3782	-71.142	POINT (-71.14200509 42.37823402)
2026-00855	2026 Jan 31 04:29:00 PM	01/30/2026 12:00 -	Hit and Run	1303	Strawberry Hill	100 Thingvalla Ave, Cambridge	42.378	-71.1537	POINT (-71.15372665 42.37795957)
2026-00852	2026 Jan 31 02:43:00 PM	01/29/2026 15:30 -	Hit and Run	506	Cambridgeport	200 Franklin St, Cambridge, MA	42.3618	-71.1057	POINT (-71.1057128 42.36182199)
2026-00851	2026 Jan 31 02:06:00 PM	01/31/2026 14:05	Medical	1101	North Cambridge	0 White St, Cambridge, MA	42.3897	-71.1189	POINT (-71.11894512 42.38970057)
2026-00850	2026 Jan 31 01:59:00 PM	01/31/2026 13:59	Accident	107	East Cambridge	200 Cambridge St, Cambridge,	42.3696	-71.0794	POINT (-71.07944904 42.36961719)
2026-00854	2026 Jan 31 01:31:00 PM	01/21/2026 12:00 -	Larceny from Residence	903	Peabody	0 Linnaean St, Cambridge, MA	42.3825	-71.1227	POINT (-71.12266538 42.38248494)
2026-00846	2026 Jan 31 12:04:00 PM	01/31/2026 12:04	Mal. Dest. Property	902	Peabody	100 Follen St, Cambridge, MA	42.3796	-71.122	POINT (-71.12199156 42.37957595)
2026-00845	2026 Jan 31 11:42:00 AM	01/31/2026 11:41	Shoplifting	1201	Highlands	200 Mowife Brook Pkwy, Camb	42.3014	-71.1465	POINT (-71.14651574 42.30141872)

Columns

RE

# Making Maps directly in Socrata

### Configure Visualization

Style by Value: None

Point Aggregation:  None,  Heat Map,  Region Map

Police Response Districts

Measure: (Count of Rows)

Color Palette: Sequential 2

Classification Method: Jenks

Number of Data Classes: 5

Midpoint: 0

Crime is

Search values

- (No value)
- Hit and Run
- Larceny from MV
- Shoplifting
- Larceny of Bicycle
- Forgery
- Mal. Dest. Property
- Warrant Arrest
- Larceny from Building

Remove Reset Apply

Preview of [Crime Reports](#)

File Nu...	Date of ...	Crime D...	Crime	Reporti...	Neighb...	Location	Reporti...	Reporti...	Reporti...
2009-01712	2009 Mar 0...	03/08/2009...	Shoplifting	1101	North Camb...	0 White St, ...	42.3897	-71.1189	POINT (-71...
2009-04219	2009 Jun 1...	06/11/2009...	Missing Per...	403	The Port	200 Harvard...	42.3648	-71.0946	POINT (-71...
2009-00005	2009 Nov 0...	11/00/2009...	Hit and Run	005	Dash...	01 Harvard...	42.3660	-71.1010	POINT (-71...

# Cambridge Geographic Information Services (GIS) Department & Resources

The screenshot displays the ArcGIS web application interface. On the left, a sidebar contains a layer management panel with the following items:

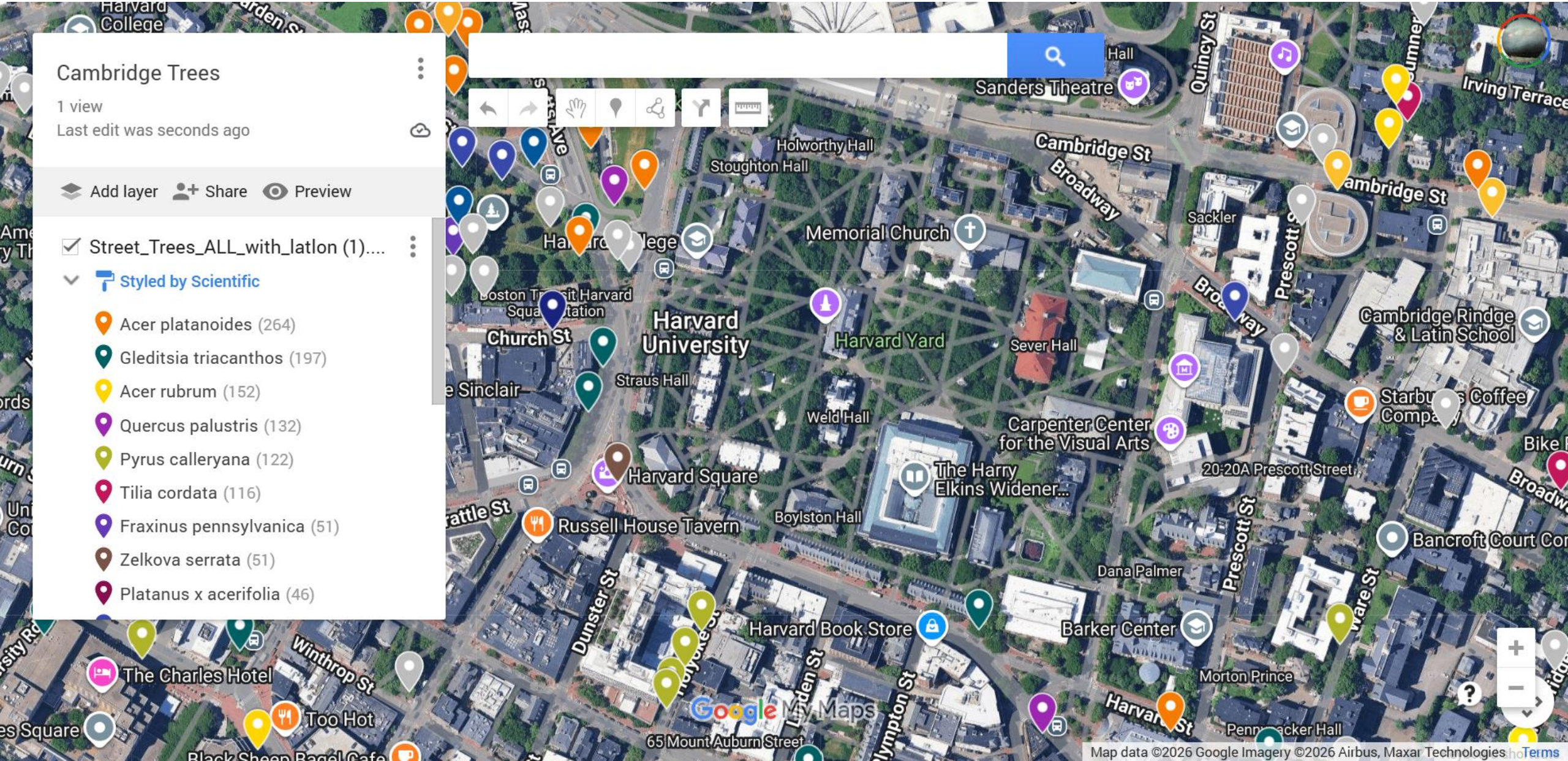
- City of Cambridge (dropdown)
- Search
- City of Cambridge (GIS icon) + Add
- Public Schools (Feature layer) with a thumbnail map
- City of Cambridge (GIS icon) + Add
- Trees (Feature layer) with a thumbnail map
- City of Cambridge (GIS icon) + Add

The main map area shows a street map of Cambridge with a large blue-shaded region representing a speed study area. The region is bounded by a dashed black line and contains numerous blue rectangular markers. Labels on the map include "Spy Pond", "Alewife", "Somerville", "Harvard", "Kendall/MIT", and "Boston College Campus GIS".

On the right side, a vertical toolbar contains icons for various map functions. Below the toolbar, a panel titled "Speed Studies" is open, showing a "Properties" section with a message: "Use the selector above to switch between layers in the map." Below this, there are sections for "Information", "Symbology", "Speed Studies", "Appearance", and "Blending".

At the bottom of the map, a footer reads: "Boston College Campus GIS, Massport, City of Cambridge, Esri, TomTom, Garmin, Saf... Powered by Esri".

# Google My Maps



Cambridge Trees

1 view  
Last edit was seconds ago

Add layer Share Preview

Street\_Trees\_ALL\_with\_latlon (1)...

▼ **Styled by Scientific**

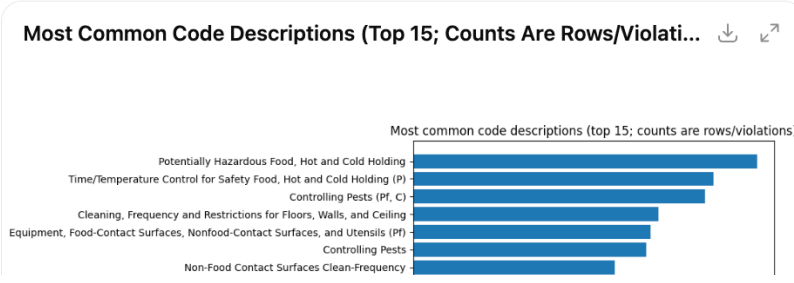
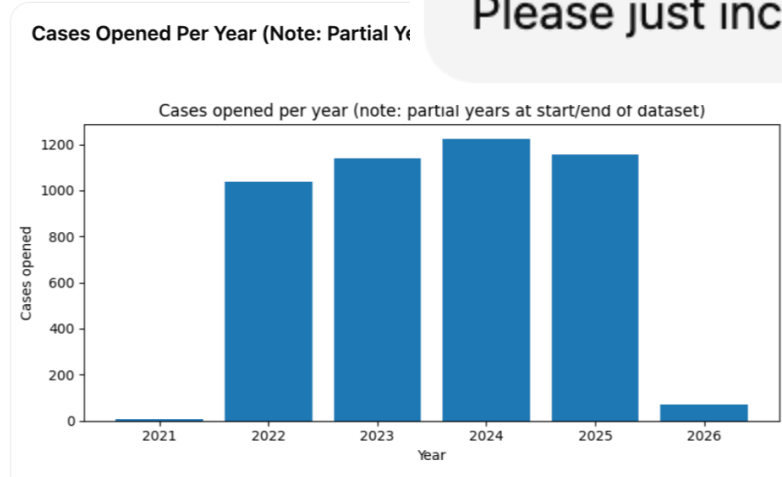
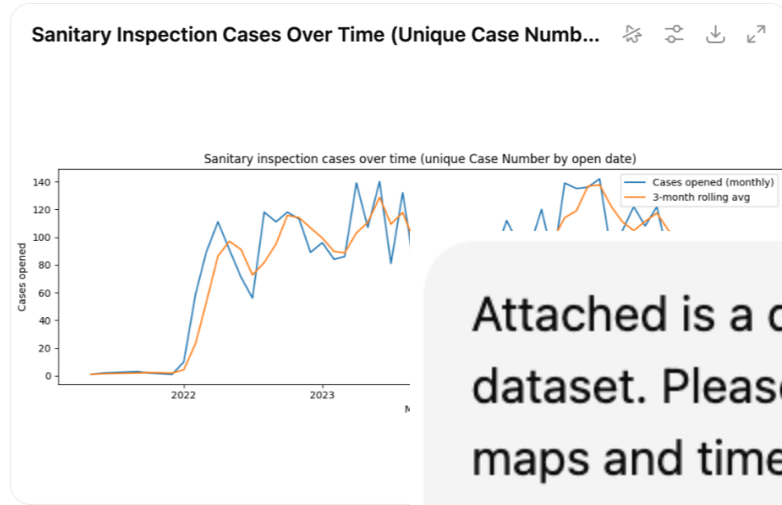
- Acer platanoides (264)
- Gleditsia triacanthos (197)
- Acer rubrum (152)
- Quercus palustris (132)
- Pyrus calleryana (122)
- Tilia cordata (116)
- Fraxinus pennsylvanica (51)
- Zelkova serrata (51)
- Platanus x acerifolia (46)

# Dashboards (Tableau)

---

# AI Chatbots: Exploratory visualization

Thought for 3m 16s >



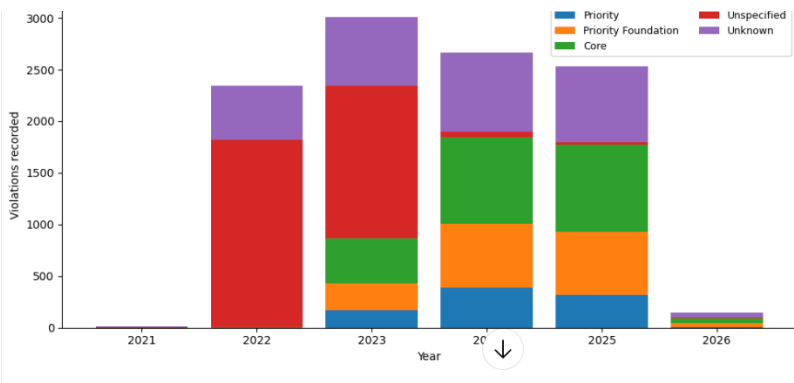
The pictures above are built from two "levels" hiding in your file:

- The dataset has 12,211 rows, but only 4,881 unique cases (Case Number). So most "trend" charts should be about cases, while "what kinds of problems" charts are about rows/violations.
- About 11.4% of cases have no lat/long, so the hotspot map is slightly incomplete (and missingness might not be random).

What jumps out (from the visuals):

The trend line shows a ramp-up into 2022, then a fairly steady inspection/case volume through 2023-2025 with visible month-to-month noise. The 2026 drop is almost certainly because the dump ends in February.

Attached is a dump of the Cambridge Sanitary Inspections open dataset. Please help me understand this data visually with charts, maps and timelines. What are the trends? Hotspots? Surprises? Please just include visuals directly in the response here.



The duration histogram shows most cases close fast (a heavy pile-up in the first few days/weeks), with a long tail. There are a handful of obvious data oddities too (8 cases have negative open→close duration, meaning dates are flipped or messy).

If you want to go one notch deeper visually next, the most revealing add-ons would be:

- a "hotspots over time" animation (monthly hex frames), and
- a "repeat locations" plot showing how concentrated repeat cases are (e.g., Lorenz curve / top X% of

August are the high months (with February and July)

very on-theme: lots of temperature/holding, pests, and pection greatest hits.

ostly for a data reason: a big chunk is ng heuristic (because many descriptions don't end with 'ptions are missing). That's a real limitation if you were

ite spatially. It's not a basemap, but the bright clusters rs and large complexes. Pair it with the high specific sites are driving those bright cells.

# AI within a spreadsheet (gemini in google sheets)

The image shows a Google Sheets interface with a bar chart and a Gemini AI analysis panel. The chart displays 'Total Sanitary Inspections by Case Closed Year (2022-2025)'. The Gemini panel provides an analysis of the data, including key takeaways and a question about the chart's data range.

**Total Sanitary Inspections by Case Closed Year (2022-2025)**

Year	Total Inspections
2021	5
2022	2380
2023	3920
2024	2858
2025	2858
2026	200

**Key Takeaways from Sanitary Inspection Data**

- High Rate of Non-Compliance, Focused on Critical Health Areas**

Approximately **77.06%** of all inspections recorded a violation, indicating that the majority of inspections result in the identification of a non-compliant

Can you get rid of 2021 and 2026 in the chart because they represent partial year data?

Gemini in Workspace can make mistakes. [Learn more](#)

# AI to “vibecode” an interactive data explorer

## Cambridge Sanitary Inspections Explorer

Interactive local explorer for sanitary inspection records: map, timelines, top code descriptions, streets/businesses, lag distributions, and a filterable table. This dataset appears to be *row-level* (often multiple rows per inspection case, e.g., one row per code/violation), so the app shows both row counts and unique case counts.

Tip: use the timeline range slider/zoom to focus on a period. Date filtering applies to the selected date field (issue/open/closed/corrected/submit).

Please create an interactive visual explorer for the attached data file. I want a tool that will help me understand through interactive maps, charts and timelines what is going on with sanitary inspections in cambridge.

Date field for filtering &

Issue date

Text search (business, a

e.g. Cambridge St, n

Lag metric histogram

Days to close

Date field: Issue date

Filtered rows

12,211

100.0% of 12,211 total ro

Reset filters

mappable records

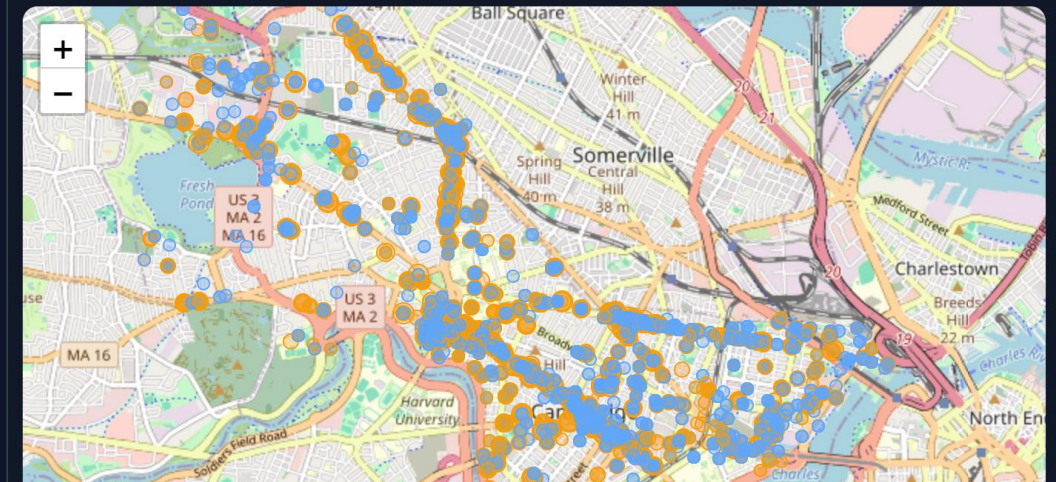
### Timeline

Unique cases by issue date (month); 0 records missing that date



### Map of cases

Case-level points linked to filtered rows (4,327 mappable cases)



# 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](mailto:OpenData@CambridgeMA.gov)**

<https://cambridgema.jotform.com/260555026278863>

