

# MBTA Delay Dashboard

Yuandi Tang | CS 7250 Spring 2025, Northeastern Khoury College  
4/24/2025



# Project Background & Motivation

- MBTA (Boston subway) delays affect thousands of daily commuters.
- Delays vary by **line, station, and year**, but users lack accessible analysis tools.
- Goal: Understand and communicate the **patterns and impact** of subway delays.
- Target audience: Boston commuters, MBTA planners, civic tech enthusiasts.
- Aims to support **data-informed decisions** and awareness around transit inefficiencies.

o2



Yuandi Tang | CS 7250 Spring 2025,  
Northeastern Khoury College  
4/24/2025

# The Problem



- Public transportation delays are not easily understood or visualized by the public.
- Who is affected?:
  - Daily MBTA riders
  - City planners and transit authorities
- Problem: Existing public data is **raw and inaccessible** to most users.
- Need for an **interactive and visual dashboard** to explore delay patterns.

# Explanation of the Data

- Data Sources:

- grouped\_delay.csv: MBTA delay records aggregated by station, line, and year. from [MBTA DATA PORTAL](#)

- station\_location.csv: Latitude and longitude for MBTA subway stations. FROM [MASSMAPPER](#)

- Key Attributes: delay\_minutes, line, station, year,

- Major Transformations:

- Grouped raw records into summarized views
- Converted delay minutes to hours
- Unified route IDs (e.g., “Green-B”, “Red”) into consistent line labels

route_id	stop_name	year	delay_dura
Blue	Airport	2020	283428
Blue	Airport	2021	36345
	Airport	2022	5729123
Chiswick R GREEN	B - Boston C	-71.150951	42.3406181
Chestnut H GREEN	B - Boston C	-71.152655	42.3385778
Reservoir GREEN	D - Riverstad	-71.148379	42.3351399
Cleveland C GREEN	C - Cleveland	-71.149287	42.3361367
Englewood GREEN	C - Cleveland	-71.145498	42.3369848
Sutherland GREEN	B - Boston C	-71.14633	42.3415615
Boston Coll GREEN	B - Boston C	-71.167025	42.3401554
Chestnut H GREEN	D - Riverstad	-71.164609	42.3267418
South Street GREEN	B - Boston C	-71.157262	42.339454
Chelsea SILVER	SL3	-71.040393	42.3968022
Bellingham SILVER	SL3	-71.033467	42.3954024
Maverick BLUE	Bowdoin to	-71.03958	42.368993
Airport SILVER	SL3	-71.029806	42.3741526
Airport BLUE	Bowdoin to	-71.030225	42.3743627
Box District SILVER	SL3	-71.028466	42.3941167
Wood Island BLUE	Bowdoin to	-71.022912	42.3796474
Eastern Avenue SILVER	SL3	-71.023934	42.3884581
Airport Ter SILVER	SL1	-71.021029	42.369638
Airport Ter SILVER	SL1	-71.021154	42.3647256
Airport Ter SILVER	SL1	-71.018544	42.3613397
Airport Ter SILVER	SL1	-71.018368	42.3630752
Airport Ter SILVER	SL1	-71.017832	42.3672363
23 Dry Dock SILVER	SL2	-71.030395	42.3446863
27 Dry Dock SILVER	SL2	-71.02806	42.3447293
Black Falcon SILVER	SL2	-71.027215	42.3441714

- Which lines have the most delays?
- Which stations are most affected?
- How have delays changed over the years?
- Are certain subway lines more consistent than others?
- How do delays compare geographically?



## MBTA Delay Analysis Dashboard



Interactive filters for year, line, and station

### Charts:

Bar chart: Delay by line

Pie chart: Distribution by line

Bar chart: Top delayed stations

Trend chart: Delay over time

### Map:

Circles sized by delay hours

Colored by MBTA line

Click to zoom and tooltip for station info

Summary cards: Total delay, average, max, most-affected station/line

# Walkthrough of the Visualization

# Design Choices

- Used D3.js for dynamic data binding and interactivity.
- Mapbox GL for a smooth and intuitive map experience.
- Consistent MBTA brand colors for lines (Green, Red, Blue, etc.)
- Simplified data views for accessibility (e.g., minutes → hours).
- Tooltip & filter-based interaction for user engagement.

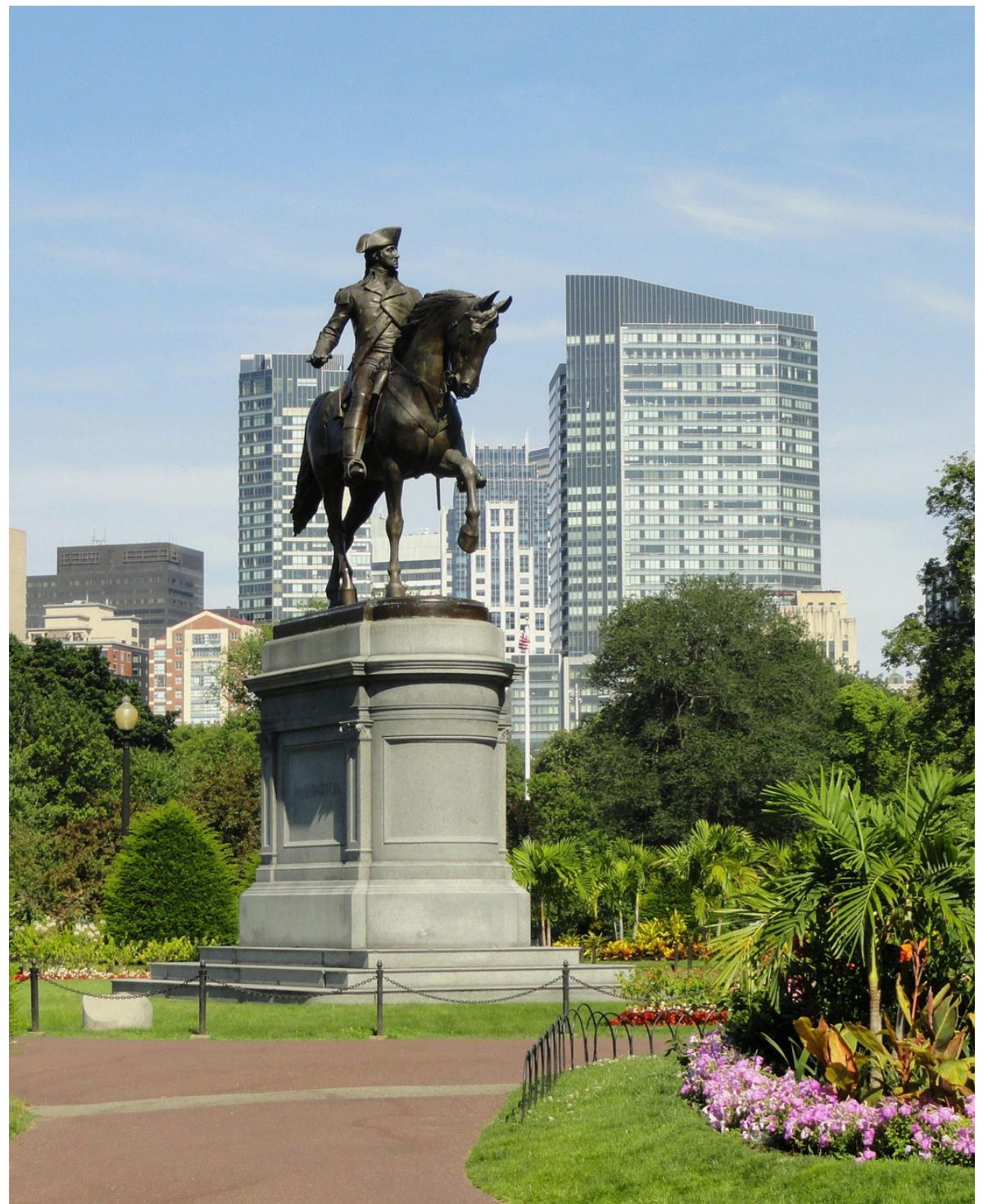


# What Did I Learn?

- How to create a **data-driven storytelling** dashboard from scratch.
- Hands-on experience with D3's **bar, pie, and stacked charts**.
- **Mapbox GL** integration and geo-visualization techniques.
- Importance of **UI/UX** in visual analytics — tooltips, filters, layout.

# Challenges & Solutions

- Challenge: Merging location and delay datasets — inconsistent station naming
  - Solution: Use various sources, apply Python in data cleaning
- Challenge: Scaling circles proportionally on the map
  - Solution: Used D3 scaleSqrt() to preserve perceptual accuracy
- Challenge: Managing complexity with many filter combinations
  - Solution: Modularized functions and state updates



Yuandi Tang | CS 7250 Spring 2025,  
Northeastern Khoury College  
4/24/2025

# Future Improvements

- Add **real-time delay data** via API (e.g., GTFS)
- Enable **data export** (CSV, image snapshots)
- Include **weekday vs. weekend or rush hour analysis**
- Build mobile-friendly responsive layout

# Thank you