How to Build a Large Scale Data Visualization

Mike Barry - Twitter
Brian Card - ViaSat

Project History In Brief

Project History In Brief

February 2014 - Collected Data
March - June - Built Project
June - Published
January 2015 - NEASIST

Press

—“Beautiful Work!” –Mike Bostock

—“Insanely Awesome” – Roberto Scalese of Boston.com

—“Beautifully crafted exploration… one of those projects you simply dream of having in your portfolio” – Andy Kirk of Visualizing Data

—Mentions by Edward Tufte, The Guardian, CNN Money, Flowing Data, FiveThirtyEight, The Atlantic and others

Total Cost: $0

Total Cost: $0

Project Management

data Visualization Tools

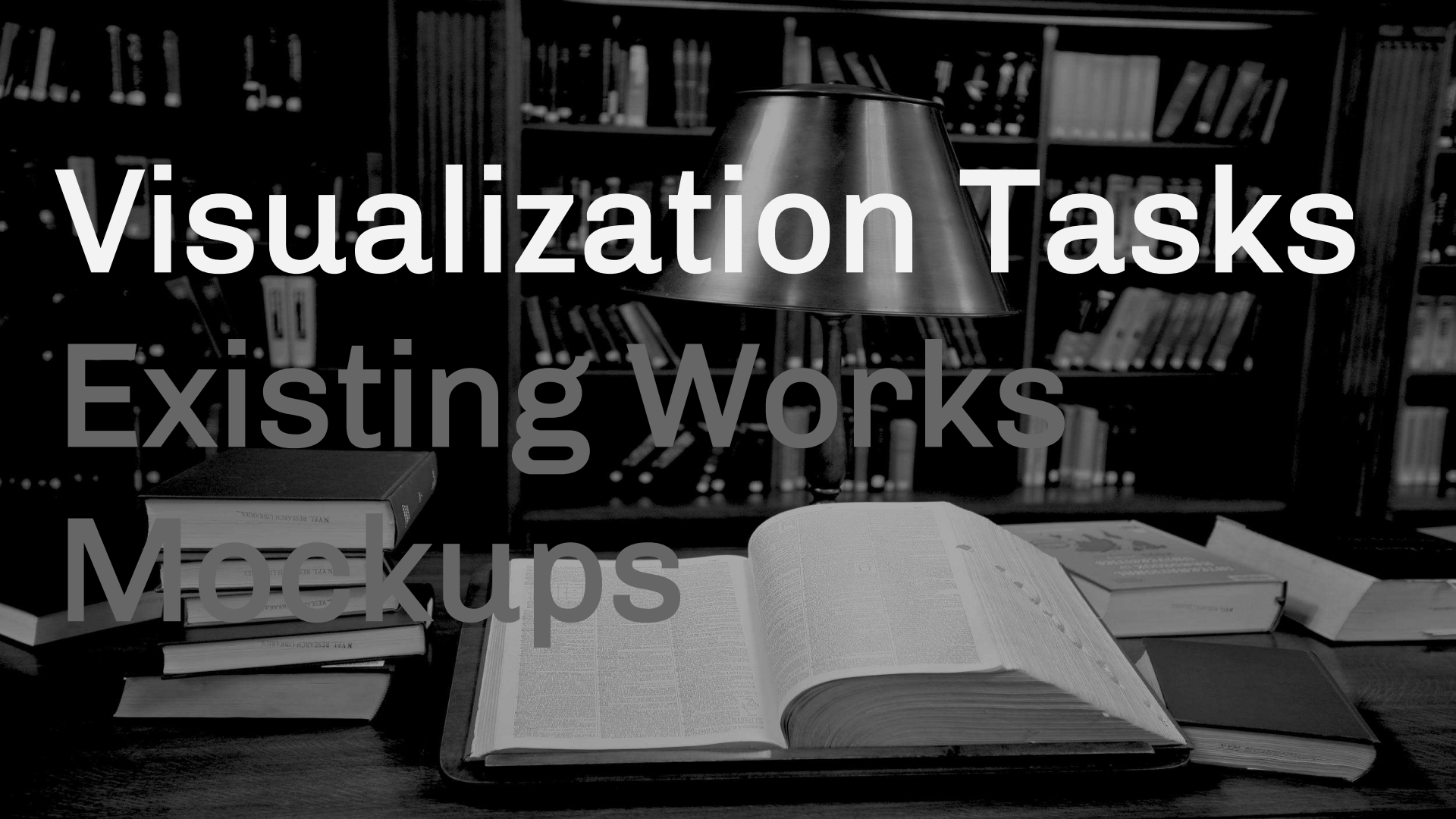
Website Publishing Code Hosting

Presentation Tools

How Did We Do It?


Research


Visualization Tasks
Existing Works
Mockups


Visualization Tasks
Existing Works
Mockups


What's Your Goal?


To Visualize The
Train System!
QUINCY ADAMS


QUINCY ADAMS

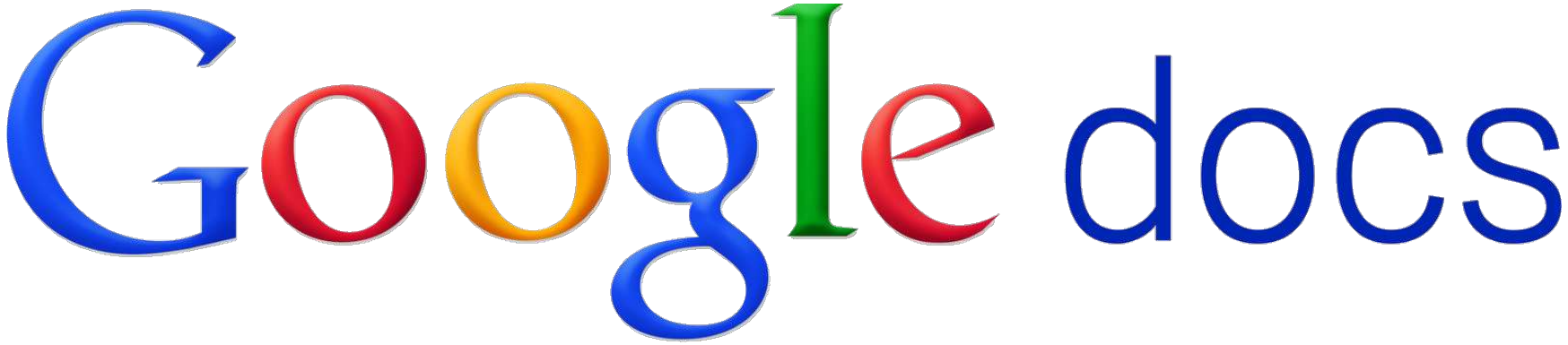
What's Important To People

- Congestion and Delay
- Snowstorms
- My Commute

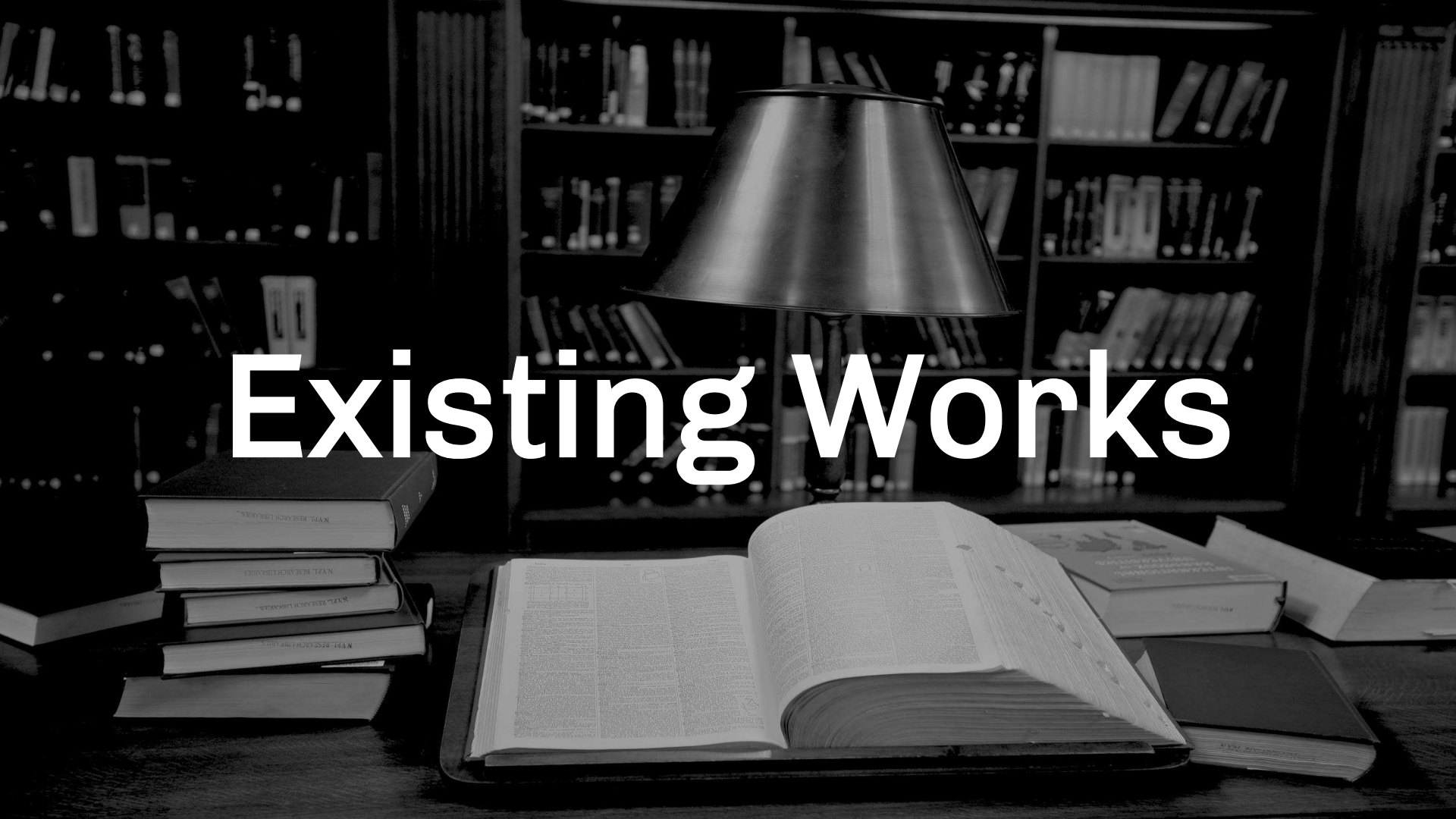
Have Ideas To Throw Away

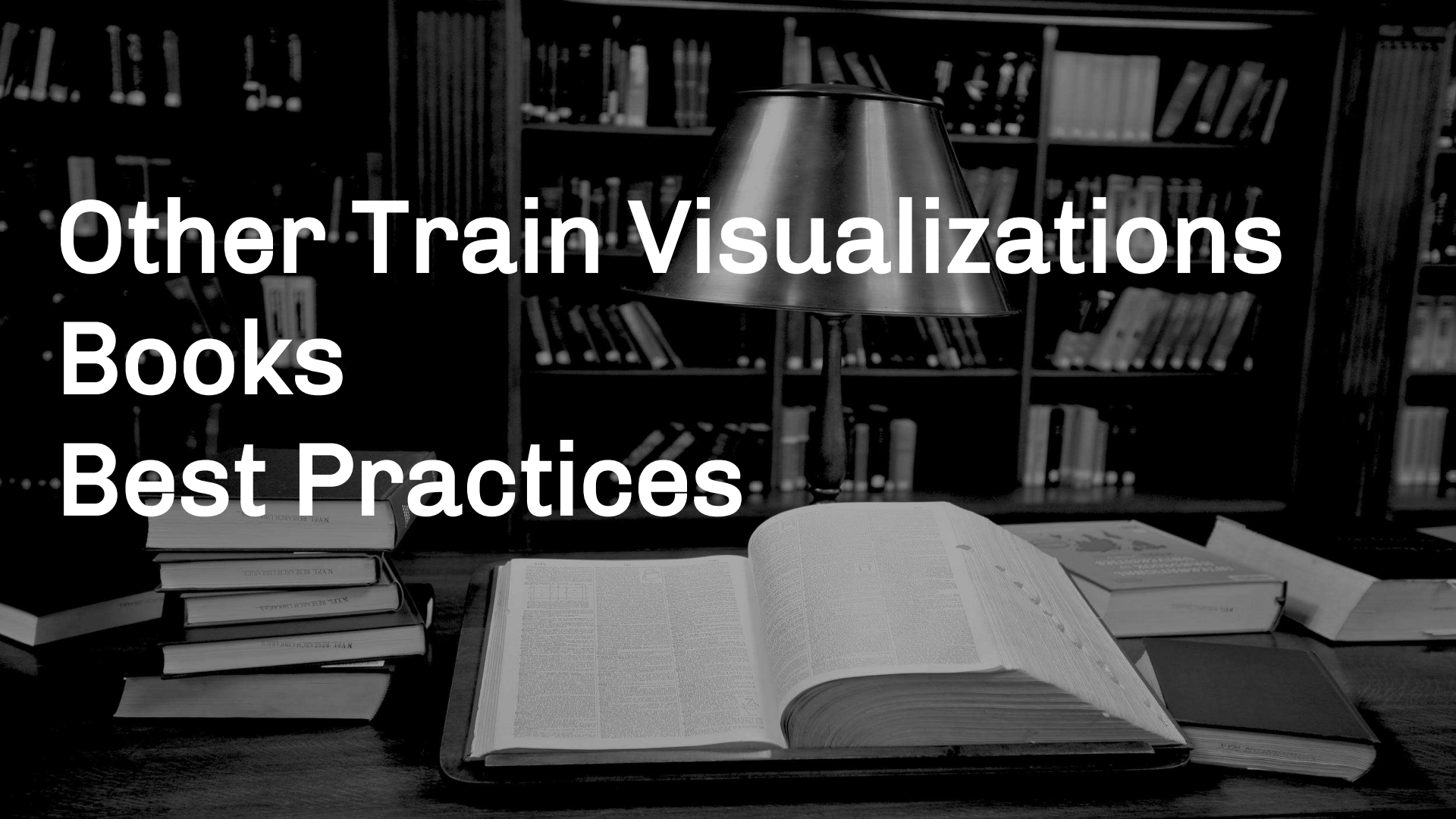
Organize Everything!


Gmail
by Google


Google docs


Trello

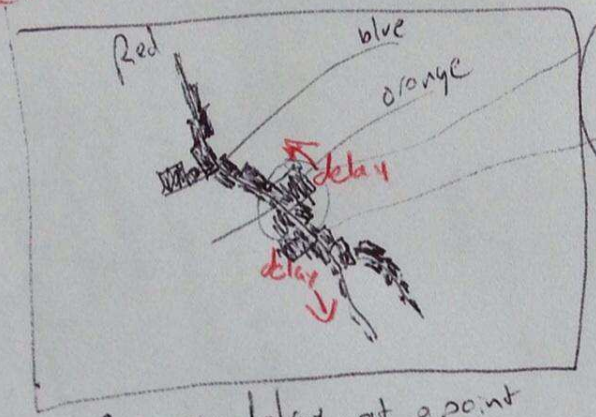

Existing Works


Other Train Visualizations
Books
Best Practices

Mockups

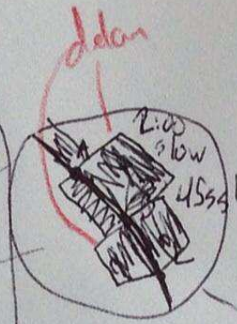

Pen + Phone +
Computer

Use real locations so shape is familiar



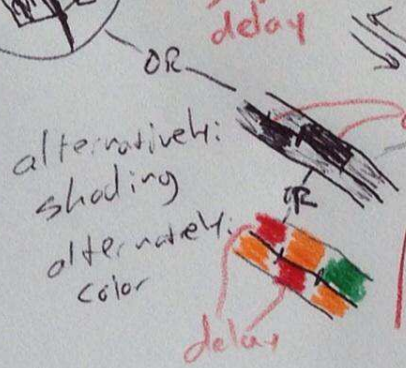
Current delay at a point in time

Overview at one point in time

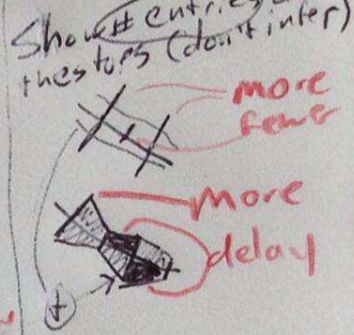


bar chart off each side between stops based on delay

OR

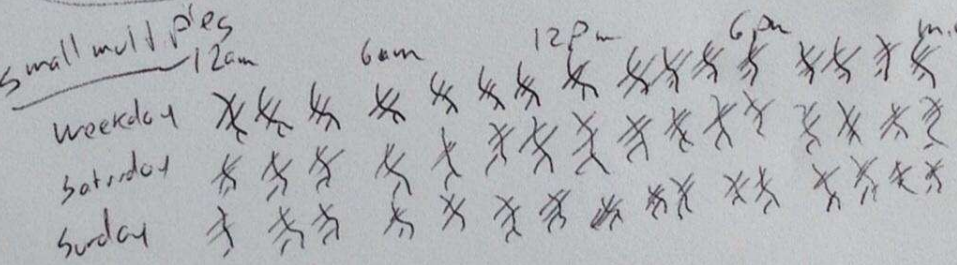


but this doesn't let us use red/orange/blue for lines



Use these as building block for:

small multiples



each one shows delay at that time

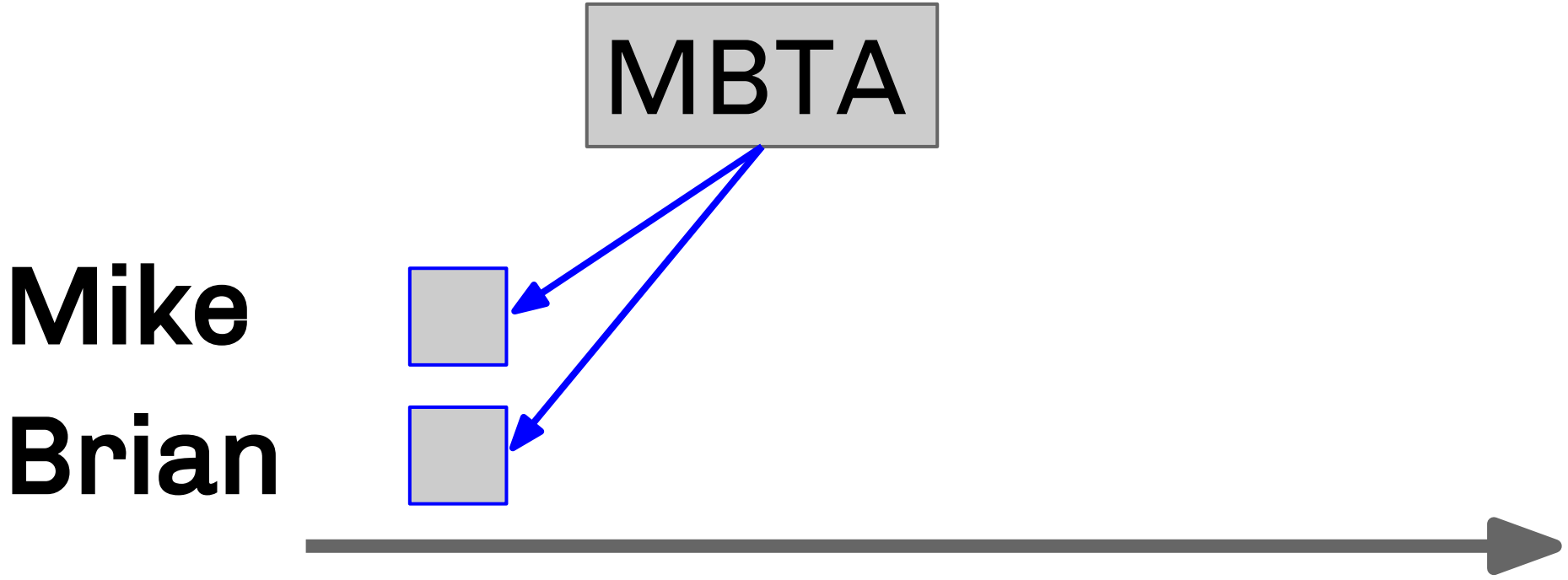
Data Collection And Prototypes

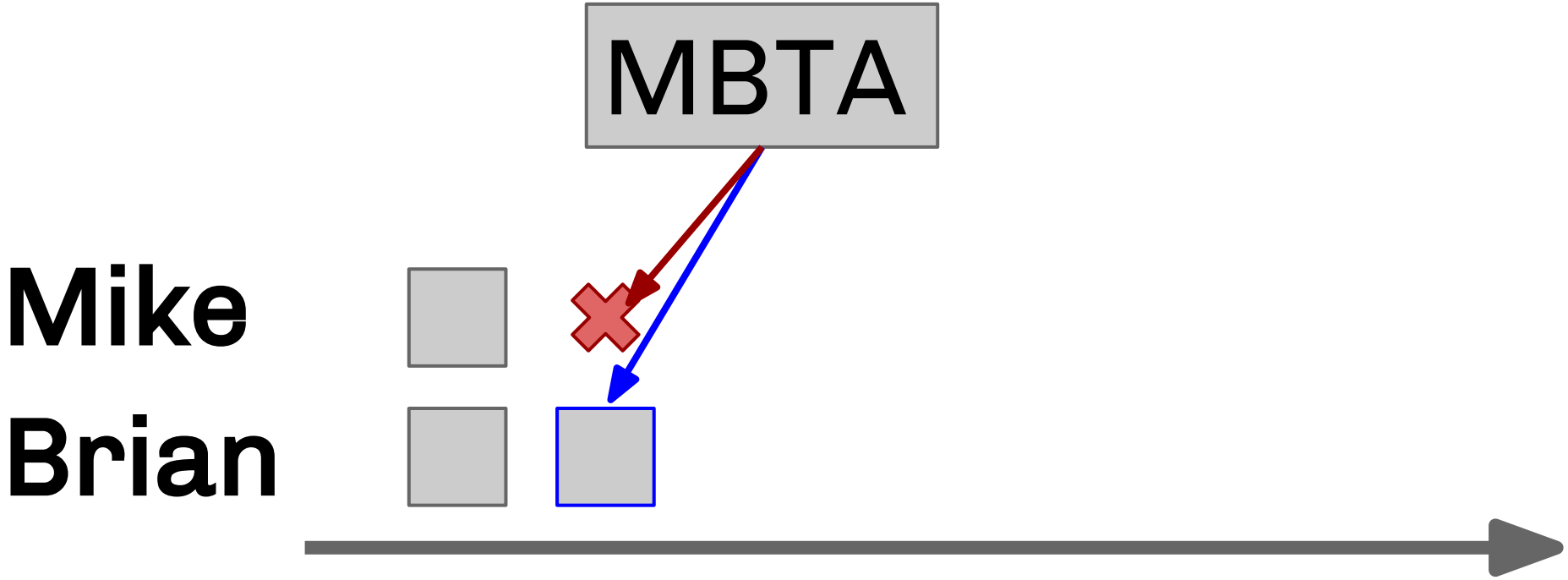
MBTA Web API

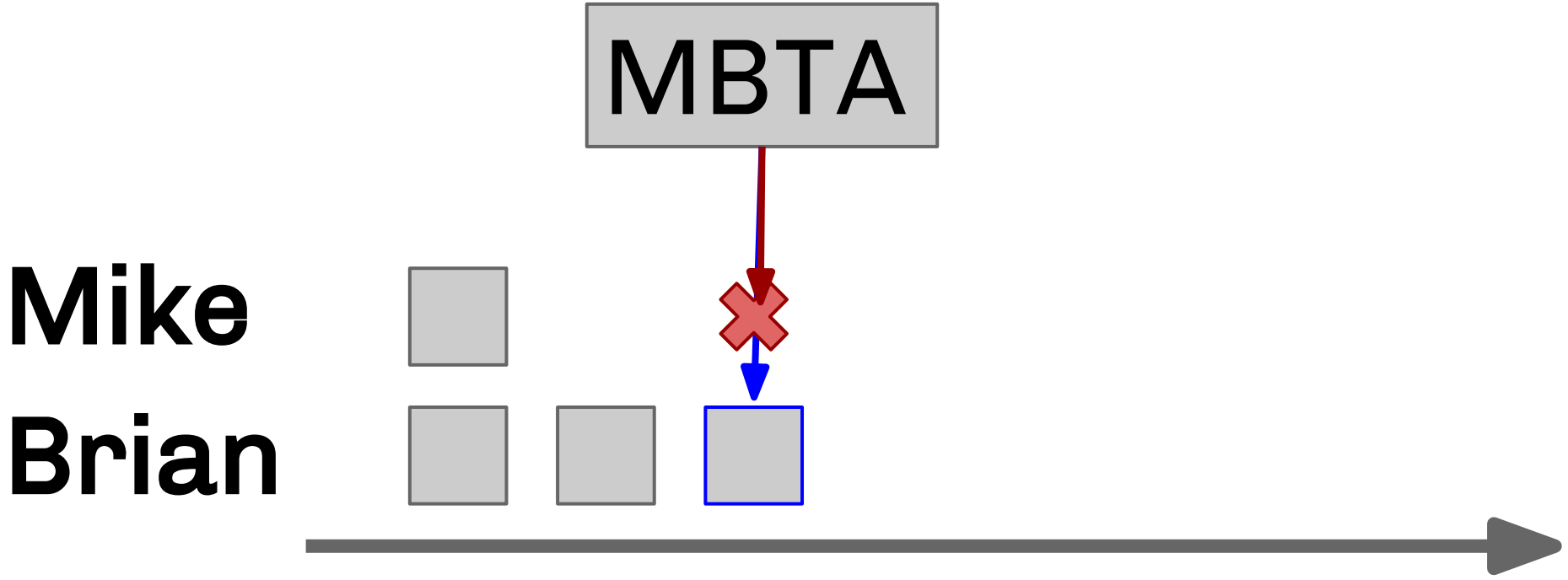
MBTA

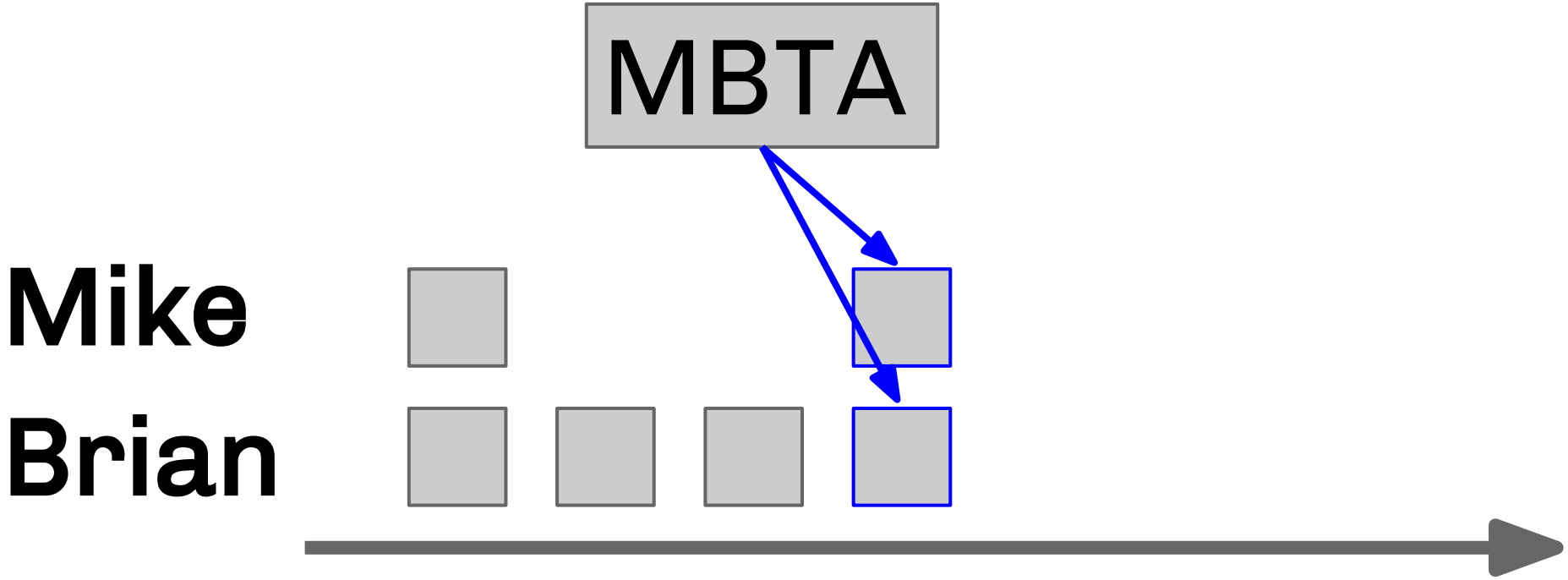
Mike

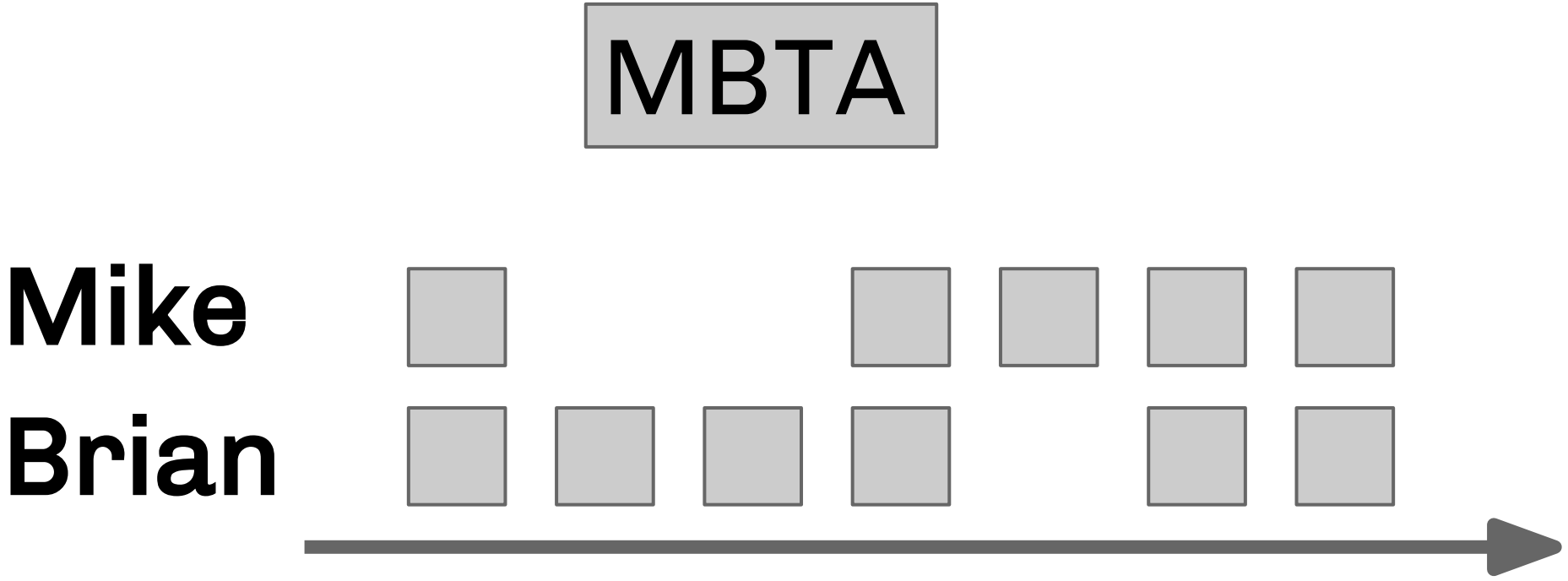
Brian

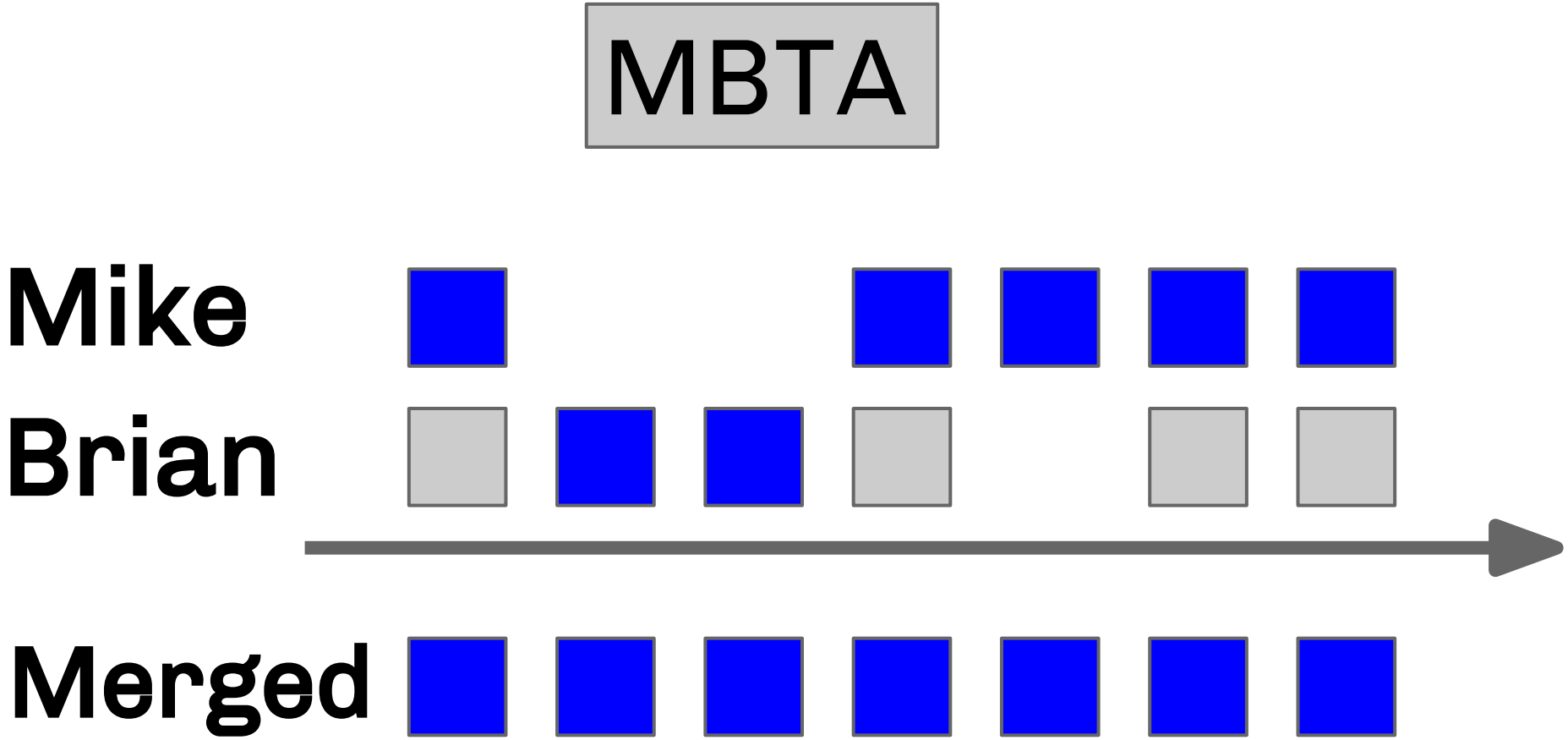

MBTA
Mike
Brian


MBTA
Mike
Brian

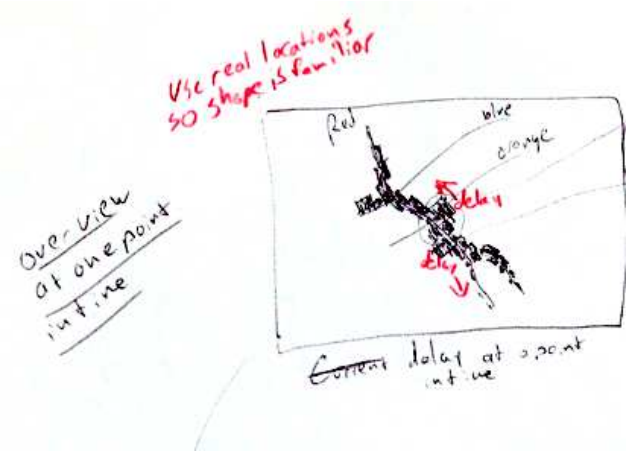

MBTA
Mike
Brian

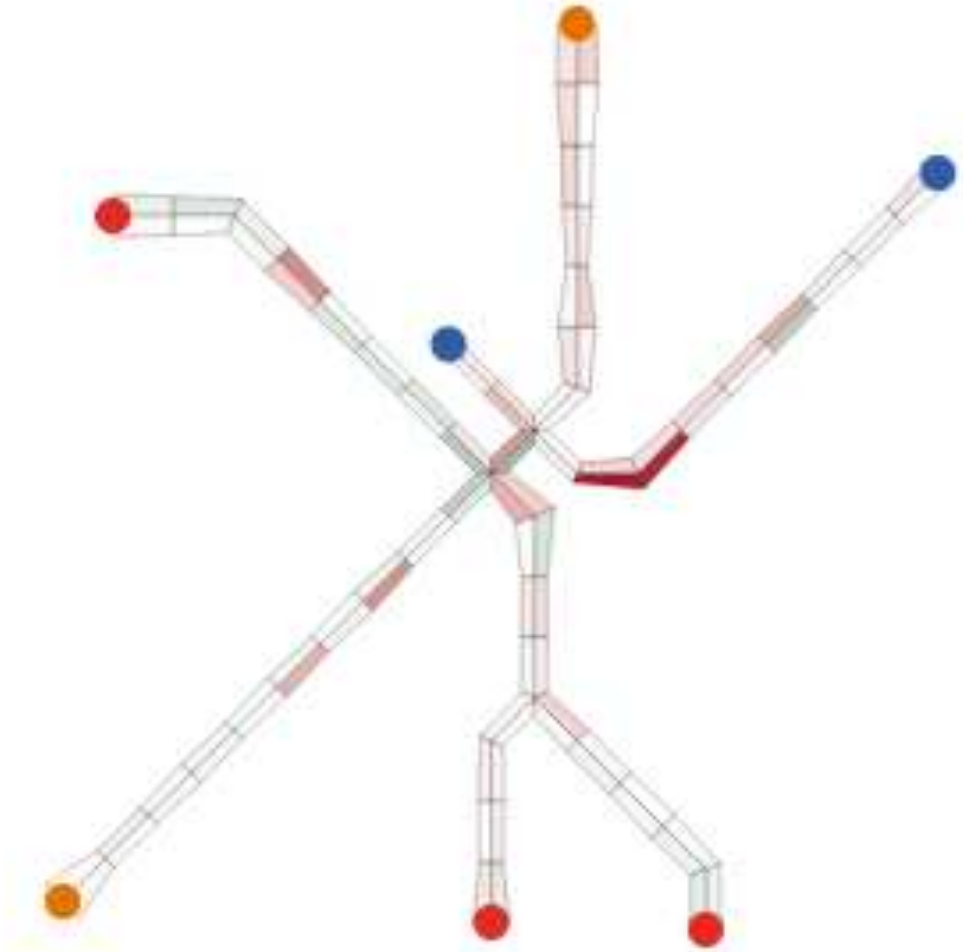

MBTA
Mike
Brian


MBTA
Mike
Brian

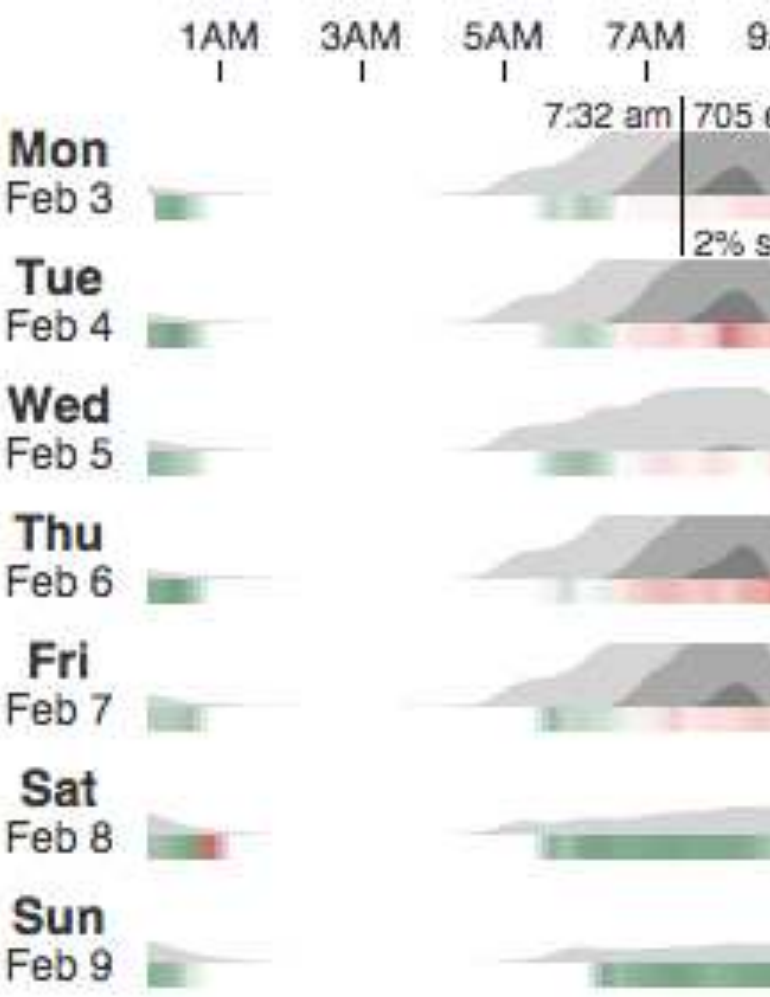

MBTA
Mike
Brian
Merged


?

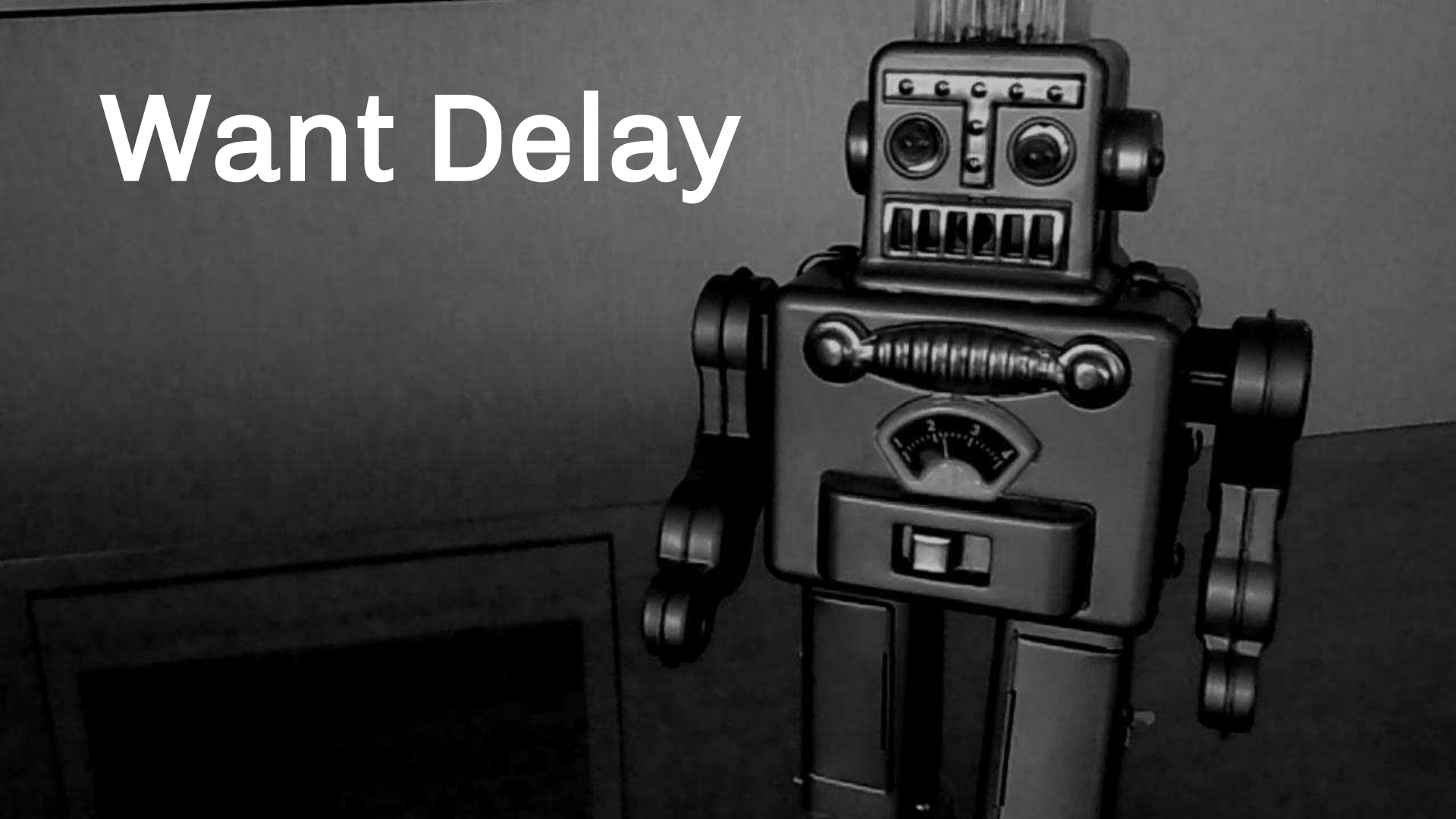

Use real locations
so shape is familiar
Red
blue
orange
delay
delay
Over-view
at one point
in time
delay at a point
in time



7:32 am on Mon Feb 3

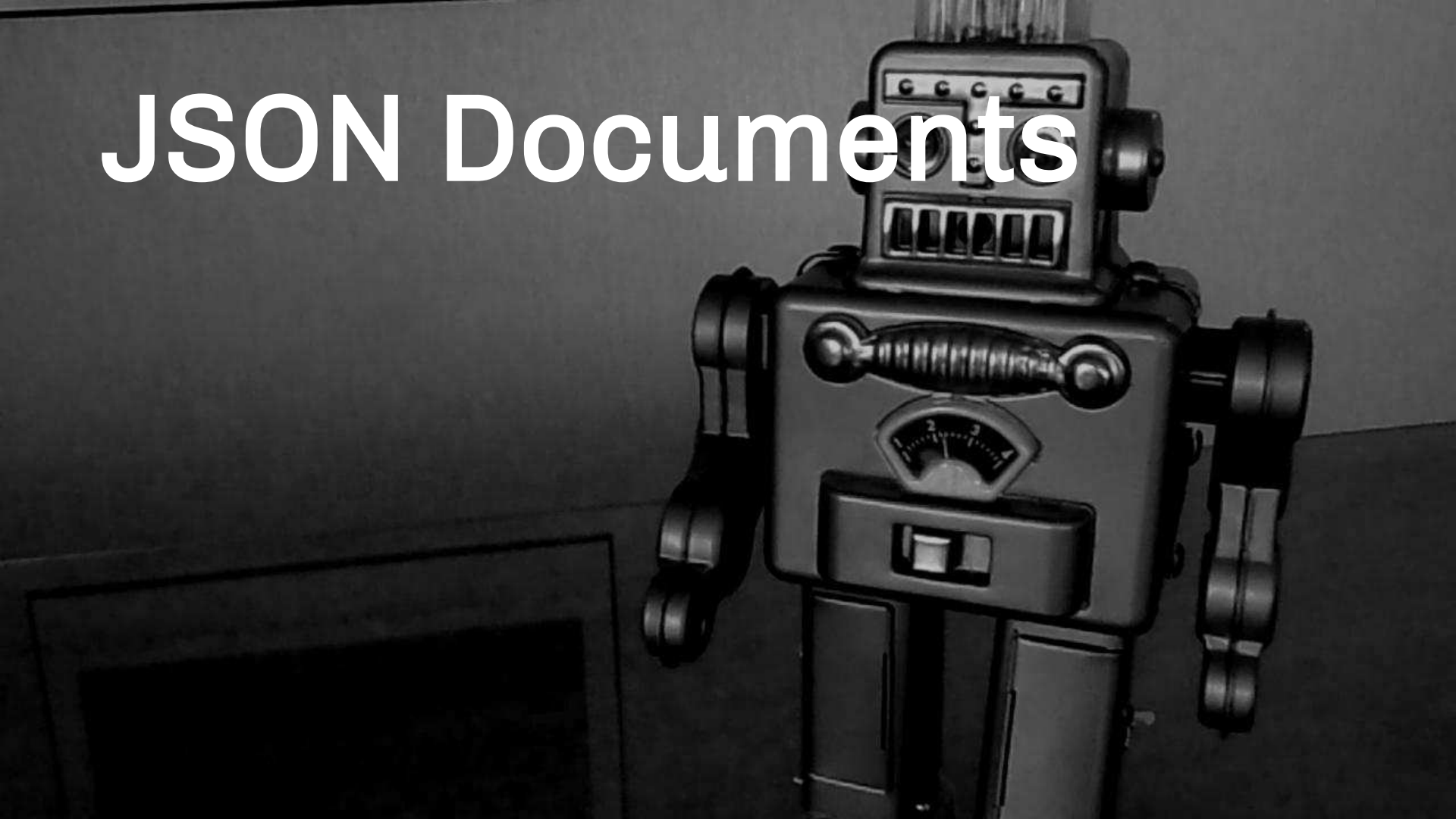


Want Delay

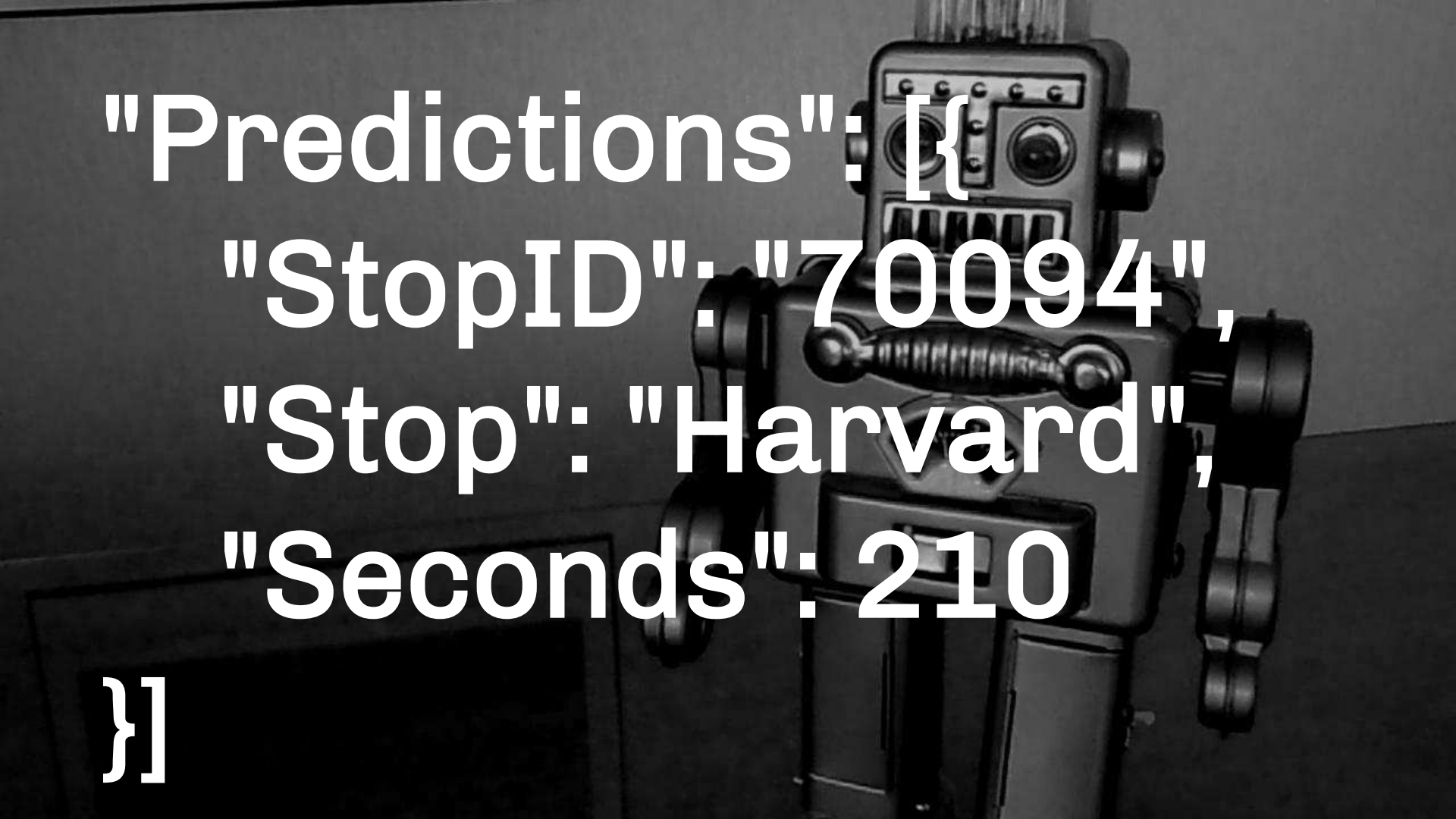



Want Delay
Need Transit Time

JSON Documents



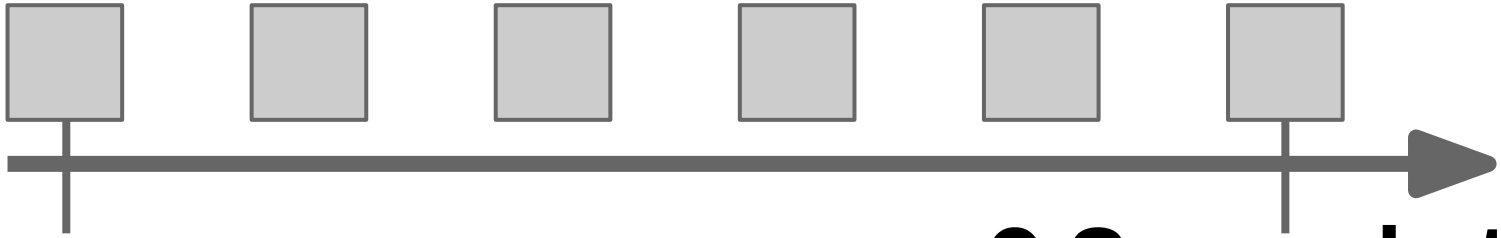

```
{"TripList": {
        "CurrentTime": 1342032950,
        "Line": "Red",
        "Trips": [{
                "TripID": "R982ECC1E",
                "Destination": "Alewife",
                "Predictions": [{
                    "StopID": "70094",
                    "Stop": "Harvard",
                    "Seconds": 210
                }]
            }]
}}
```


"Predictions": [{
"StopID": "70094",
"Stop": "Harvard",
"Seconds": 210
}]


"Predictions": [{
"StopID": "70094",
"Stop": "Harvard",
"Seconds": 210
}]

Time 0

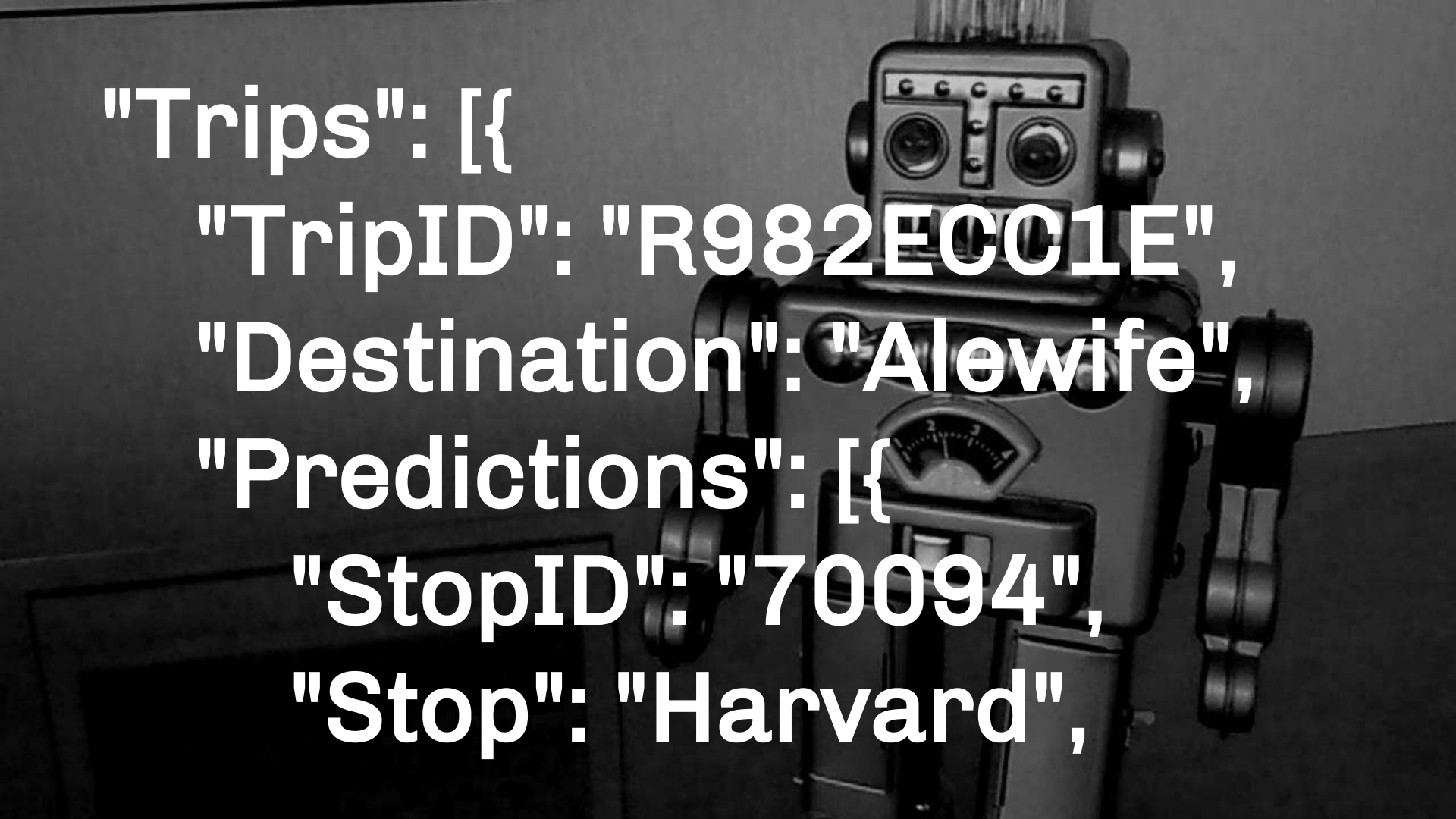
Time 240

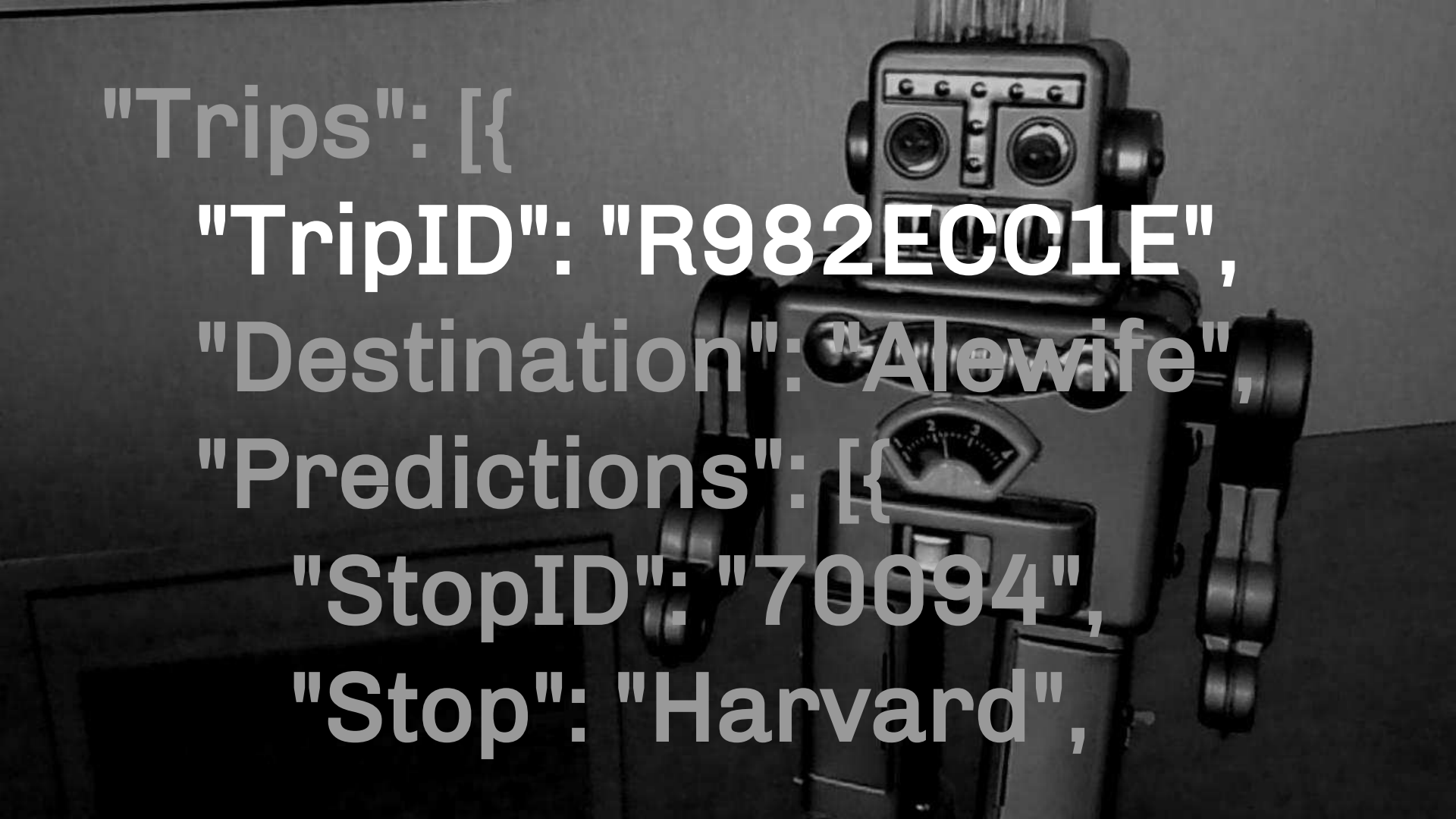


210 Seconds to Harvard

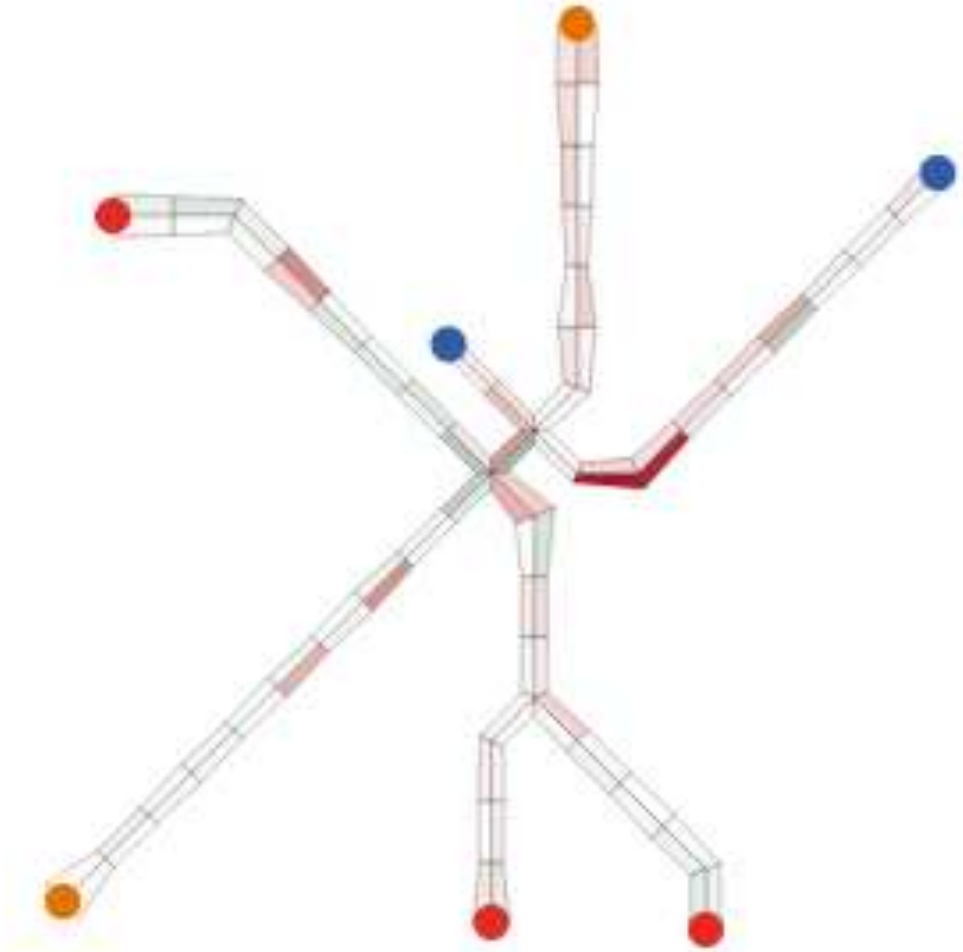
0 Seconds to Harvard

240 seconds to get from Central to Harvard

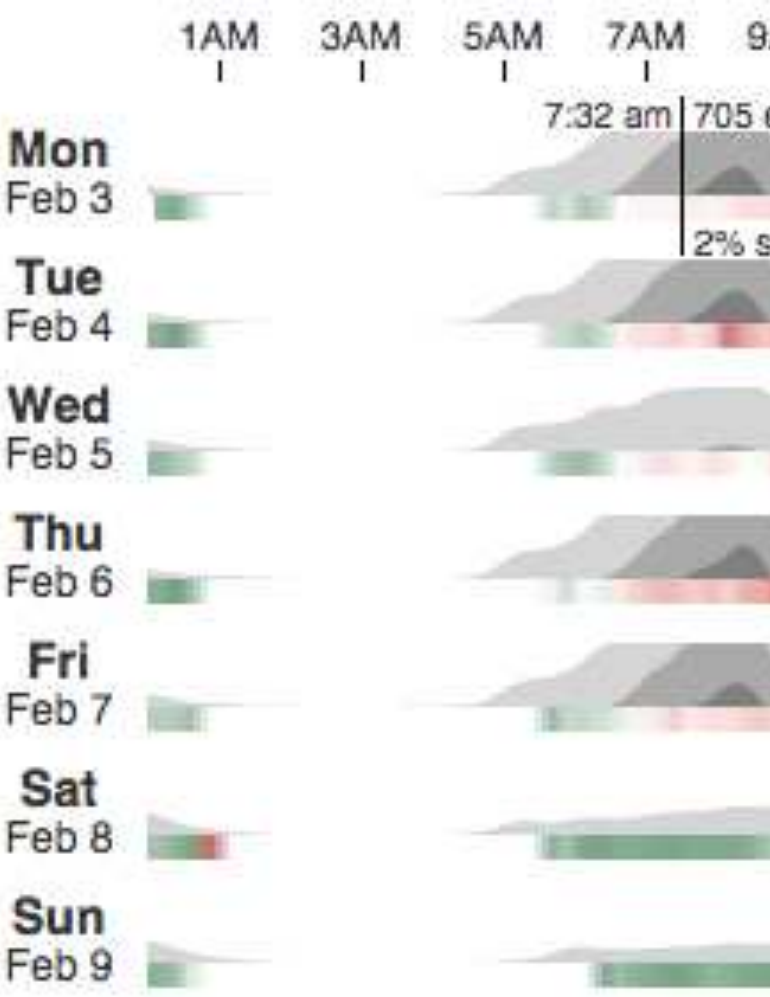

"Trips": [{
"TripID": "R982ECC1E",
"Destination": "Alewife",
"Predictions": [{
"StopID": "70094",
"Stop": "Harvard",

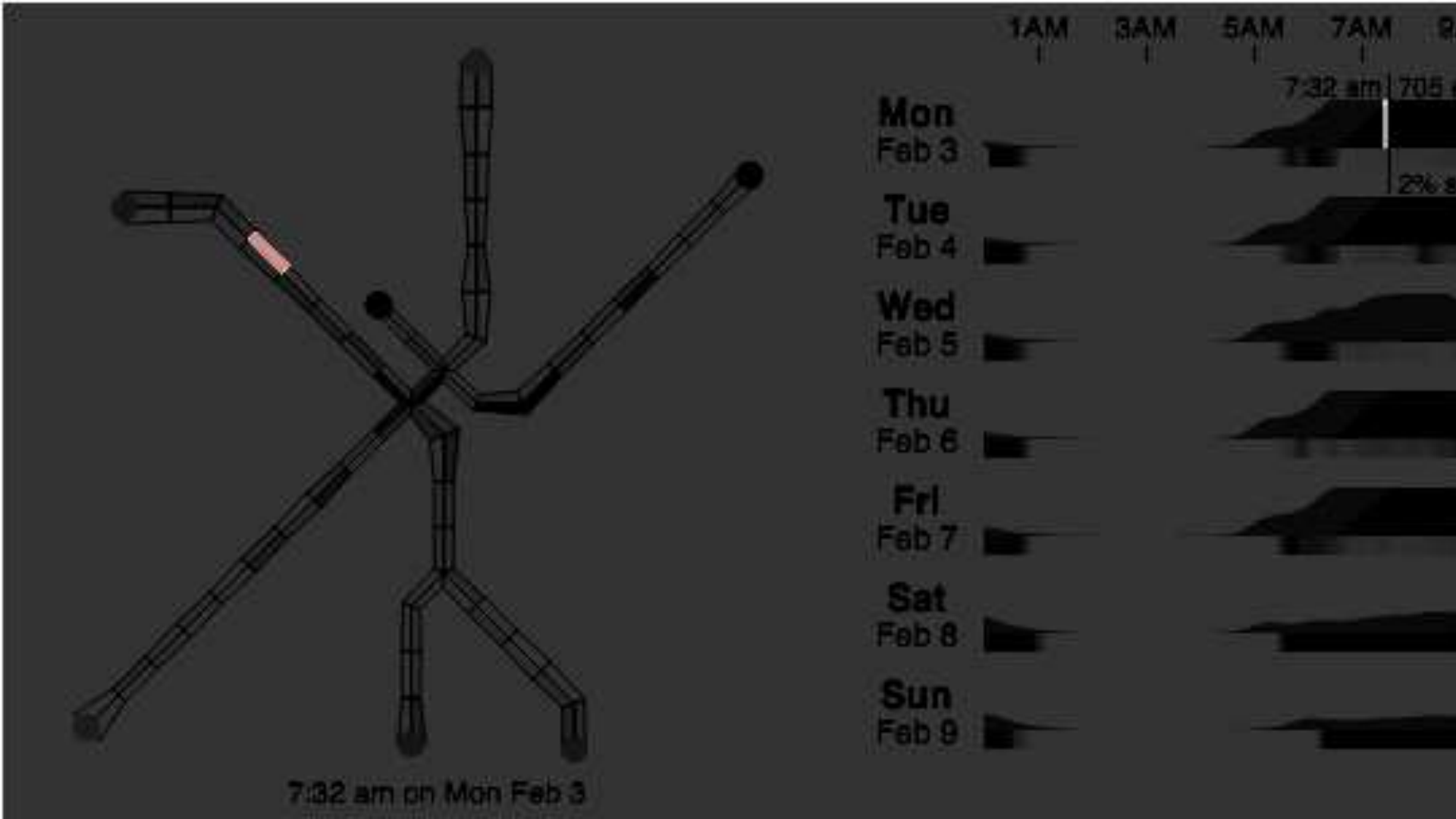

"Trips": [{
"TripID": "R982ECC1E",
"Destination": "Alewife",
"Predictions": [{
"StopID": "70094",
"Stop": "Harvard",

Trip	Time
R982ECC1E	240 sec
R98338169	220 sec
R98338126	300 sec

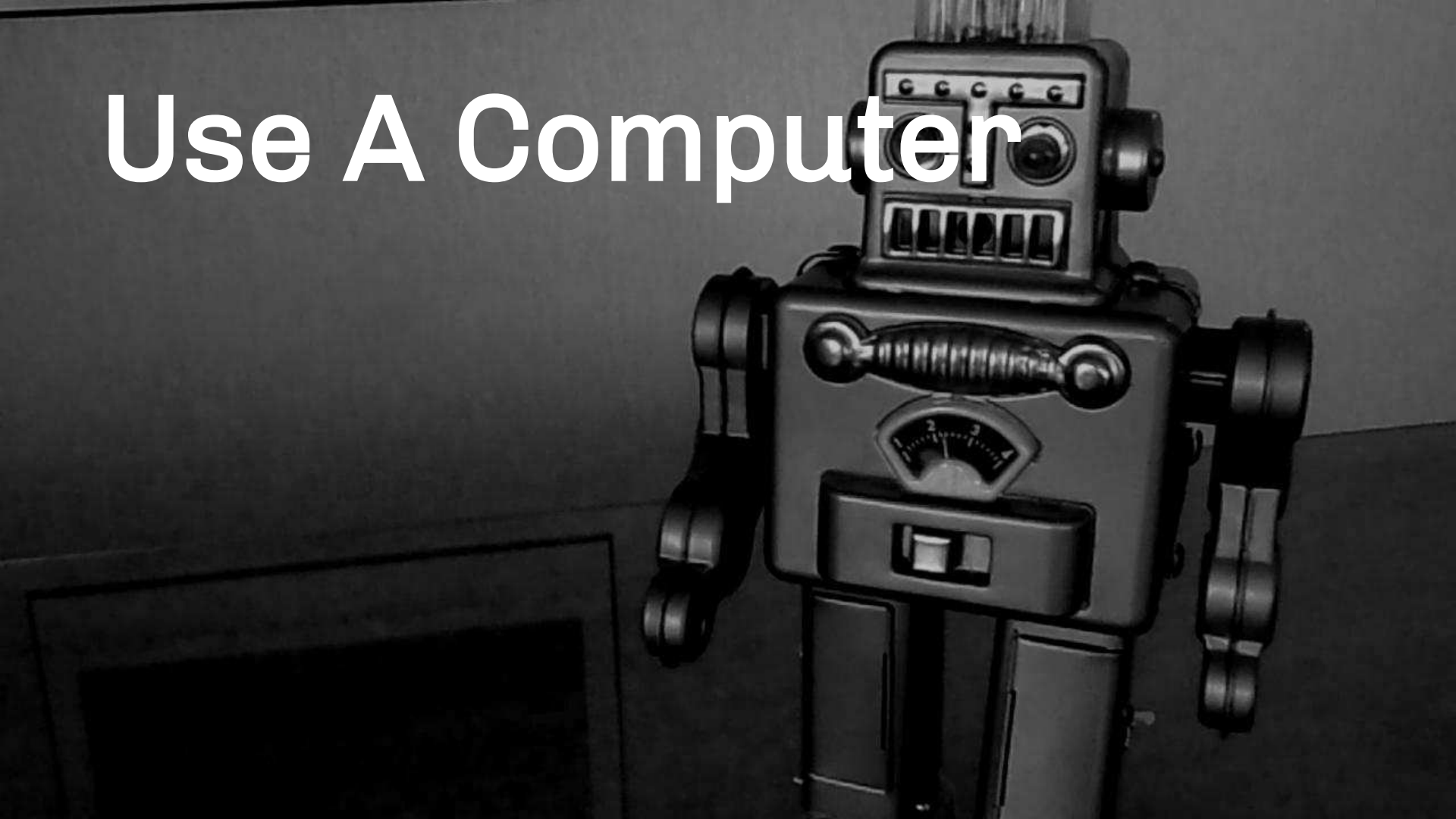


7:32 am on Mon Feb 3




1AM
3AM
5AM
7AM
Mon
Feb 3
Tue
Feb 4
Wed
Feb 5
Thu
Feb 6
Fri
Feb 7
Sat
Feb 8
Sun
Feb 9
7:32 am on Mon Feb 3

Use A Computer

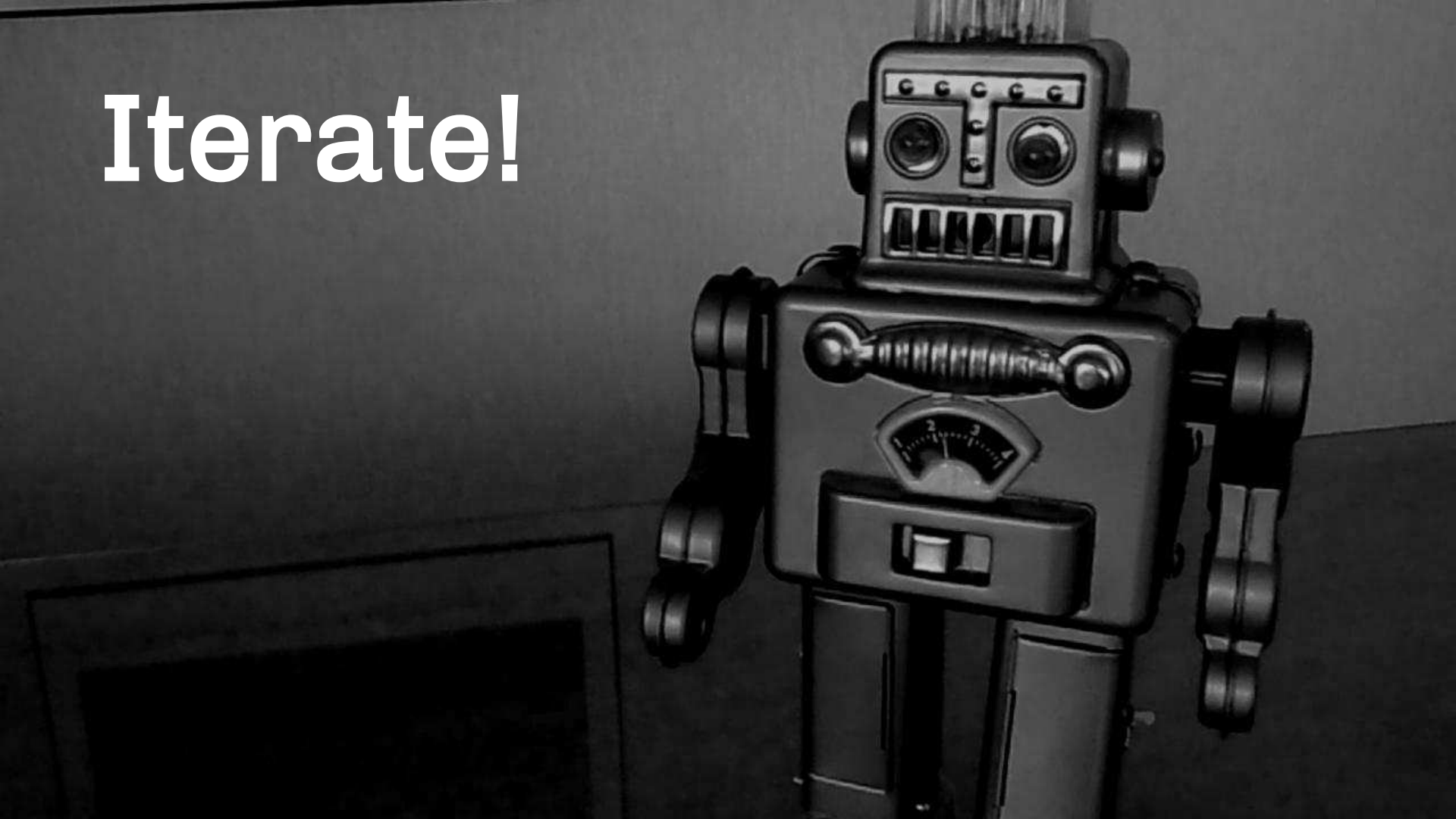


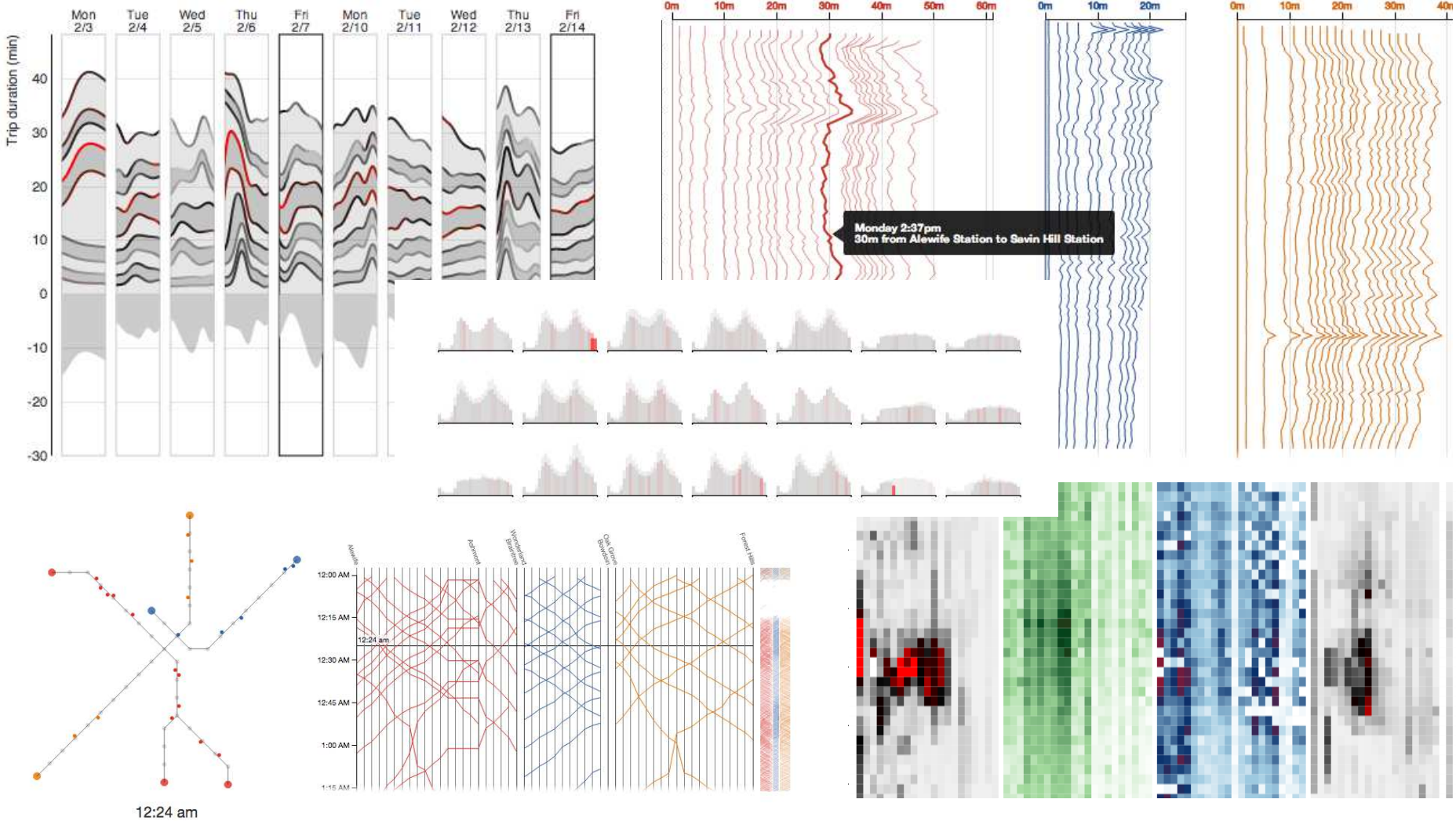

node
JS
TM


Data-Driven Documents

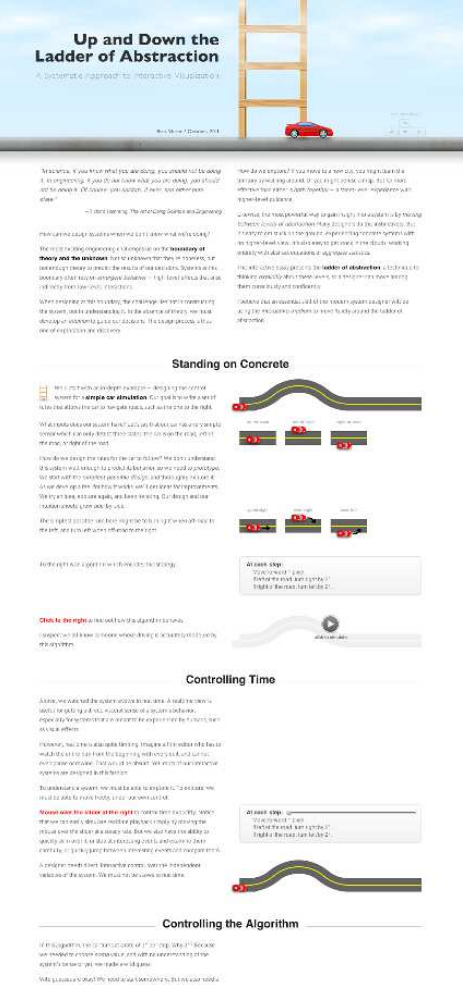

Atlassian
Bitbucket

Iterate!

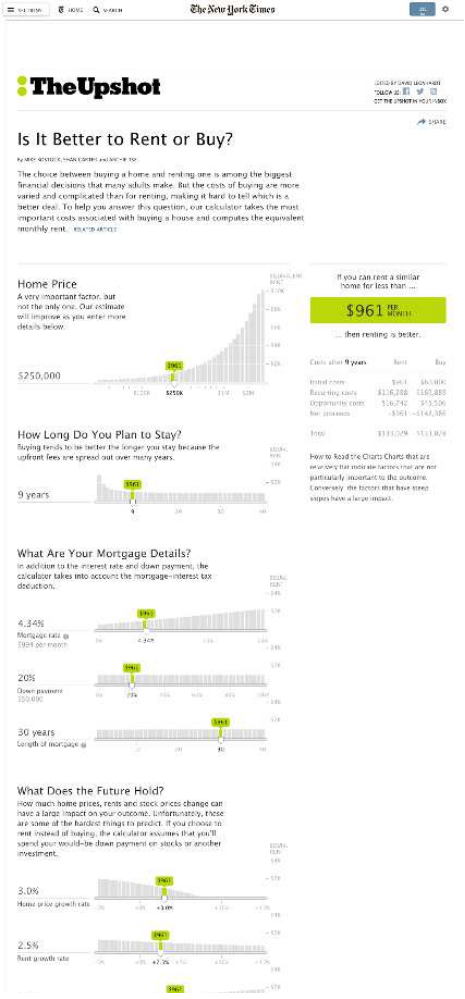



Mon 2/3
Tue 2/4
Wed 2/5
Thu 2/6
Fri 2/7
Mon 2/10
Tue 2/11
Wed 2/12
Thu 2/13
Fri 2/14
Trip duration (min)
40
30
20
10
0
-10
-20
-30
0m
10m
20m
30m
40m
50m
60m
Monday 2:37pm
30m from Alewife Station to Savin Hill Station
12:24 am
12:00 AM
12:15 AM
12:30 AM
12:45 AM
1:00 AM
Alewife
Davis
Porter
Harvard
Central
Kendall/MIT
Charles/MGH
Park Street
Downtown Crossing
Forest Hills

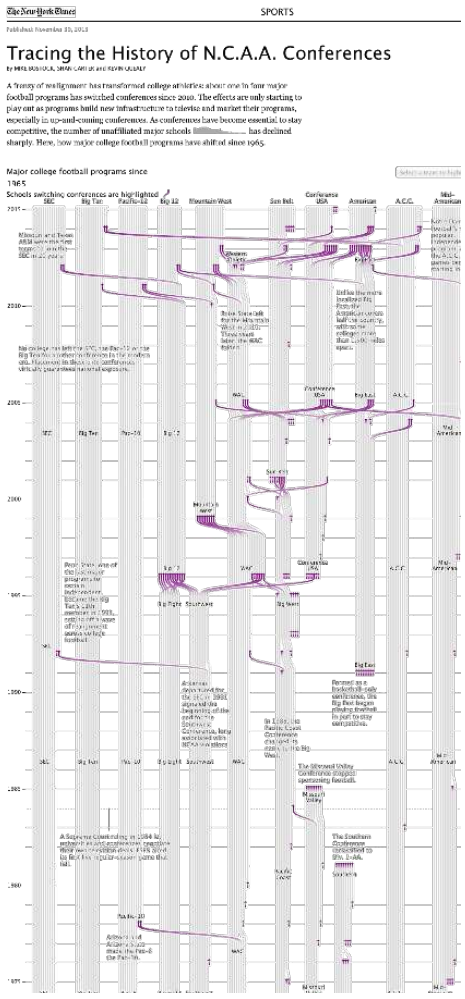
Putting It All Together



Victor, Bret. “Up and Down the Ladder of Abstraction.” October 2011. worrydream.com/LadderOfAbstraction



Bostock, Mike et al. “Is It Better to Rent or Buy.” May 2014. www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html



Bostock, Mike et al. “Tracing the History of N.C.A.A. Conferences.” November 2013. www.nytimes.com/newsgraphics/2013/11/30/football-conferences/

Bostock, Mike. “Visualizing Algorithms.” June 2014. bost.ocks.org/mike/algorithms

Outline

File Edit View Insert Format Tools Table Add-ons Help Last edit was on April 19, 2014

Comments Share

100% Normal text Arial 11 More

<< insert lined-up marey here >>

The People

In a typical weekday, over 400,000 people enter a station along the red, orange, or blue lines. On Saturdays and holidays that number drops to 230,000 and on Sundays it drops to 160,000. The busiest day was Friday February 7 when 470,187 people entered the system.

This shows the total number of entries and exits for the red, orange, and blue line for every hour over the entire month. You can see weekends and holidays, as well as daily peaks around rush hour. Our exit data is unreliable since not all stations measure require that people exit through a turnstile.

<< Render a single row of the turnstile heatmap for sum of all stations >>

The busiest stations are all along the Red Line. Harvard topped the list, followed closely by South Station, and then Downtown Crossing. The graphic below shows turnstile entries over time across all stations.

<< Render the turnstile heatmap, sortable by position along the line or entries/decreasing >>

How People and Trains Affect Each Other

When you look back at the Marey diagram, the slope of each line tells you how fast a train is going and the time it takes to get between stations. When all of the start and stop times are lined up you can see a drastic variation in the time it takes to get between stops throughout the day. If you have ever ridden the subway during rush hour then you have experienced what the steep lines in the Marey diagram feel like first-hand.

What causes these delays? It's hard to know for sure, but what we can determine is that delays usually happen when more people are riding the subway. The next visualization shows this correlation.

<<figure out a graphic here that shows correllation between crowds and delay>>

But how do these crowds and delay typically affect you on your commute? Choose the two stations you typically commute between on a line

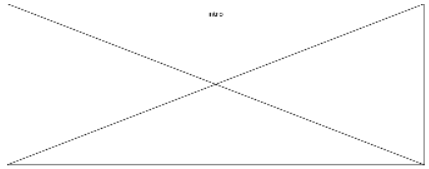
Red Line (laid out vertically)	Orange Line	Blue Line

Brian Card
10:13 PM Apr 17, ... Resolve

Would be interesting to color code your Mary diagram that has all of the starting points at the same position or something so you could see how the time of day affects the delay.

Michael Barry
5:55 AM Apr 18, 2014

Yea, we can do that, although we would probably want to select a small number of time buckets. What about:
- early
- morning commute
- mid-day
- evening commute
- night time



A Month in the Life of the MBTA

An exploration of ridership and the trains of Boston

Michael Barry & Brian Card

Boston's Massachusetts Bay Transit Authority (MBTA) operates the 4th busiest subway system in the U.S. after New York, Washington, and Chicago. If you live in or around the city you have probably ridden on it. You may remember subway rides that were slow, cars that were crowded, or everything going smoothly. When you get off the train, however, you lose sight of the subway until you need to ride it again. It's hard to know if you left at a different time if the trains would be less crowded or what the experience is like on the other side of the system.

The MBTA publishes a substantial amount of subway data. They provide the full schedule in General Transit Feed Specification (GTFS) format which powers Google's transit directions. They also publish realtime train locations for the Red, Orange, and Blue lines (but not Green or Silver). We captured this realtime data for the entire month of February, 2014. Also, working with the MBTA, we were able to acquire per-minute entry and exit counts at each station measured at the turnstiles used for payment.

We attempt to present this information to help people in Boston better understand the trains, how people use the trains, and how the people and trains interact with each other.

The Trains

In a typical weekday, trains make approximately 1150 trips on the red, orange, and blue lines starting at 5AM and continuing through 1AM the next morning. On Saturdays trains make 870 trips and on Sundays they make 760.

The visualization below shows all of these trips that trains took on the red, orange, and blue lines on February TBD, 2014. Each vertical line represents a station, and time extends from top to bottom. Steeper lines indicate slower trains. This visualization was first used by Etienne-Jules Marey and is typically called a "Marey Diagram"

Average Subway Trips per Day

	Weekdays	Saturdays	Sundays
Red	450	320	280
Orange	350	260	230
Blue	300	220	240
Total	1150	870	750

To better compare these individual trips, we line up the starting points and you can see the range of fastest to slowest trips, as well as variation. (describe a couple of notable trips, where hovering over them dims all of the others)

The People

In a typical weekday, over 400,000 people enter a station along the red, orange, or blue lines. On Saturdays and holidays that number drops to 230,000 and on Sundays it drops to 160,000. The busiest day was Friday February 7 when 470,187 people entered the system.

This shows the total number of entries and exits for the red, orange, and blue line for every hour over the entire month. You can see weekends and holidays, as well as daily peaks around rush hour. Our exit data is unreliable since not all stations measure require that people exit through a turnstile.

The busiest stations are all along the Red Line. Harvard topped the list, followed closely by South Station, and then Downtown Crossing. The graphic below shows turnstile entries over time across all stations.

How People and Trains Affect Each Other

When you look back at the Marey diagram, the slope of each line tells you how fast a train is going and the time it takes to get between stations. When all of the start and stop times are lined up you can see a drastic variation in the time it takes to get between stops throughout the day. If you have ever ridden the subway during rush hour then you have experienced what the steep lines in the Marey diagram feel like first-hand.

What causes these delays? It's hard to know for sure, but what we can determine is that delays usually happen when more people are riding the subway. The next visualization shows this correlation.

But how do these crowds and delay typically affect you on your commute? Choose the two stations you typically commute between on a line

Summary

Through publicly available data, we have the tools to understand the subway system better than we ever have before. We have seen how the system operates on a daily basis, how people use the system and how that affects the trains, and how the trains and people affect you.

Through our analysis and the realtime data feeds, you can also take these insights with you whenever you ride the train. Bookmark mbta.meteor.com on your mobile phone phone and you can check it any time to see up-to-the-minute congestion and delay information.


10 Days Left
MONTINARI
MILANO
QUARTZ

```
commit b2fb7010ca3222ba7a987b5f85ddaf2c7a604192
Author: Mike Barry
Date:   Fri Apr 18 08:13:13 2014 -0400

    Initial commit of final project
```

```
.
|____app
| |____index.html
|____bower.json
|____Gruntfile.js
|____package.json
|____README.md
```

```
commit b2fb7010ca3222ba7a987b5f85ddaf2c7a604192
Author: Mike Barry
Date:   Fri Apr 18 08:13:13 2014 -0400

    Initial commit of final project
```

```
commit c63f5989df5b014abacafc1bb92b53e3bc8ebb54
Author: Mike Barry
Date:   Tue Apr 22 06:47:40 2014 -0400

    try out MBTA theme
```

```
commit e0d609dc509724aa201429fe5e42c0f82ebf9ca2
Author: Mike Barry
Date:   Sun Apr 27 17:20:32 2014 -0400

    fix math :-(
```

```
.
|____app
| |____index.html
|____bower.json
|____Gruntfile.js
|____package.json
|____README.md
```

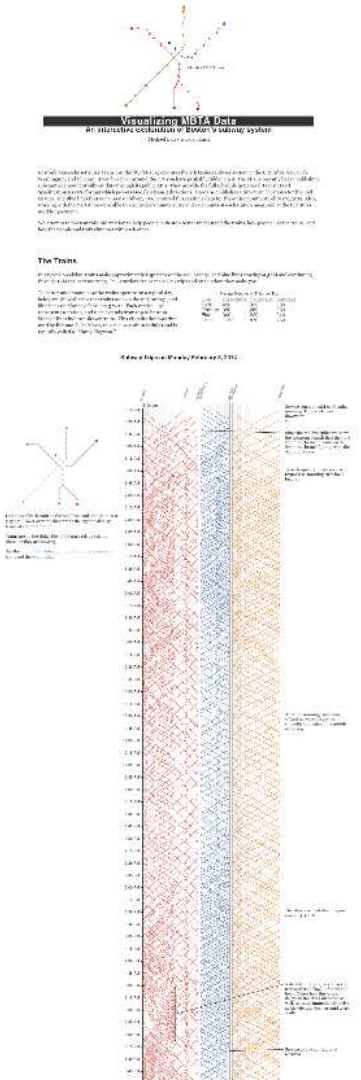
```
commit b2fb7010ca3222ba7a987b5f85ddaf2c7a604192
Author: Mike Barry
Date:   Fri Apr 18 08:13:13 2014 -0400

    Initial commit of final project
```

```
commit c63f5989df5b014abacafc1bb92b53e3bc8ebb54
Author: Mike Barry
Date:   Tue Apr 22 06:47:40 2014 -0400

    try out MBTA theme
```

```
commit e0d609dc509724aa201429fe5e42c0f82ebf9ca2
Author: Mike Barry
Date:   Sun Apr 27 17:20:32 2014 -0400

    fix math :-(
```

```
.
|____app
| |____index.html
|____bower.json
|____Gruntfile.js
|____package.json
|____README.md
```

```
.
|____app
| |____data
| | |____average-actual-delays.json
| | |____delay.json
| | |____interesting-turnstile-data.json
| | |____marey-header.json
| | |____marey-trips.json
| | |____spider.json
| | |____station-network.json
| | |____turnstile-gtfs-mapping.json
| | |____turnstile-heatmap.json
| |____ie.png
| |____index.html
| |____media
| |____preview.png
| |____scripts
| | |____dataloader.js
| | |____delay.js
| | |____files.js
| | |____first.js
| | |____header.js
| | |____horizon.js
| | |____marey.js
| | |____turnstile.js
| |____styles
| | |____d3tip.less
| | |____delay.less
| | |____header.less
| | |____main.less
| | |____marey.less
| | |____signs.less
| | |____table.less
| | |____theme.less
| | |____turnstile.less
|____bower.json
|____data-generators
| |____generate-delay.js
| |____generate-marey.js
| |____generate-turnstile-heatmap.js
| |____update-file-sizes.js
|____generate-data.sh
|____Gruntfile.js
|____package.json
|____README.md
|____tools
  |____gen-screenshots.js
```



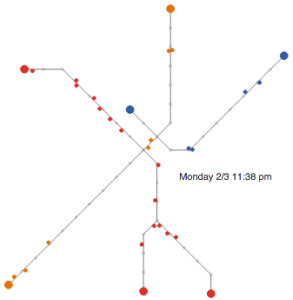
Michael Barry 4/29/14

to Matt, Brian

Professor Ward,

Here is the final paper and project for CS525D for Mike Barry and Brian Card.

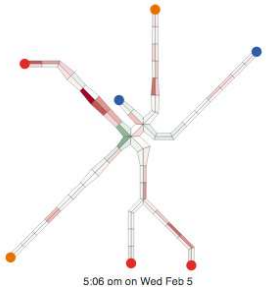

If You Find
Something That
Works, Run with It

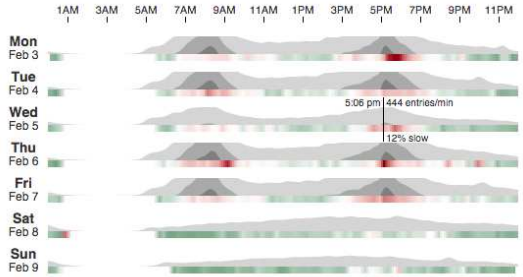


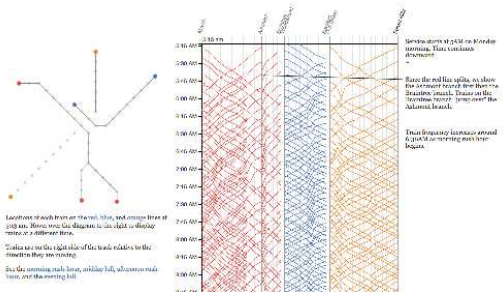
Visualizing MBTA Data

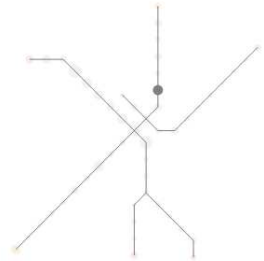
An interactive exploration of Boston's subway system

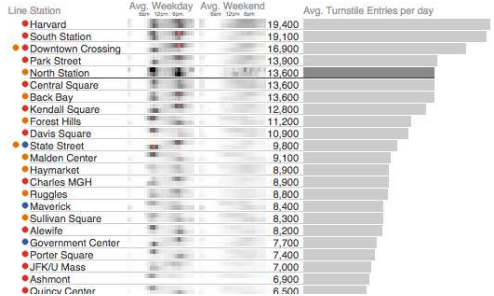
Michael Barry & Brian Card



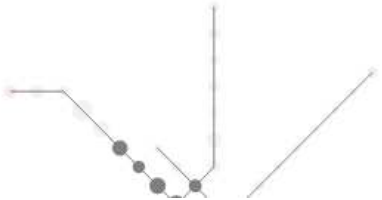


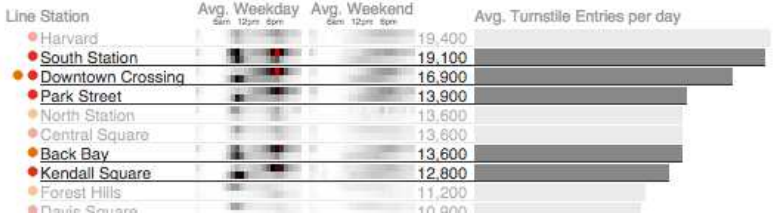


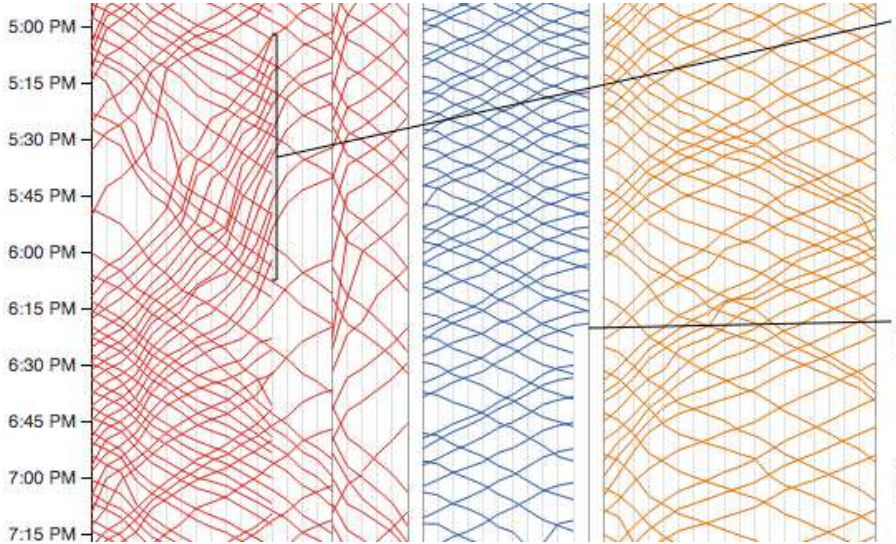




When entries and exits are broken down by station, you can see the busiest stations are all along the Red Line. Harvard topped the list, followed close by South Station , and then Downtown Crossing . Next to each station are heatmaps showing entrances and exits to each station per-hour for weekdays and weekends/holidays. You can see that some stations are work stations since their exits peak in the morning and entrances peak in the afternoon and that some stations are home stations since their entrances peak in the morning and exits peak in the afternoon. Some stations are just busy all the time. Hover over a stop to see where it on the map on the left. Click a stop to show a detailed heatmap below.







A disabled train causes delays on trains after (below) it for over an hour. Notice how this causes delays in the other direction as well, as trains immediately arrive at Alewife then turn around to go south.

Service to Bowdoin stops at 6:20PM

Normal service resumes for the evening starting around 7PM

Harvard is the busiest station and has a const

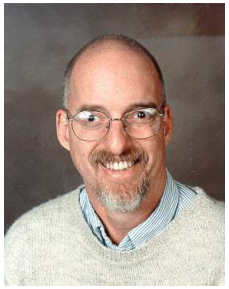
In the heart of Boston's Financial District, Sou
the end of the commuter rail and people who
the commuter

Get Feedback

Done?

A Few More Things...

- Respond to feedback
- Cross-browser and mobile testing
- Your commute
- Web hosting
- Marketing



WPI

Size shows turnstile entries on average day

500 10,000 19,400 people per day

Line width shows turnstile entries at a station

0 50 100 people per minute

Color shows delay

20% faster on time 40% slower than normal

Gray bars show entries to all stations

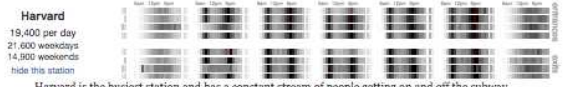
0 390 780 1060 people per minute

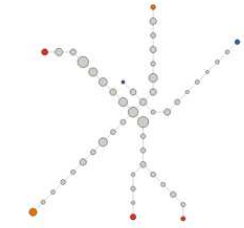
Color shows average entrances/exits

0 200 400 600 870 people per minute

BeehiveMedia

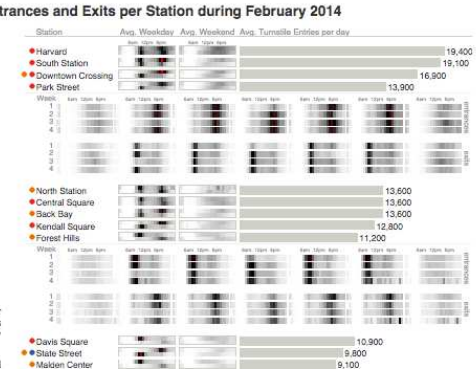
Click a station above to show a detailed breakdown of entrances/exits per hour during the month of February below.

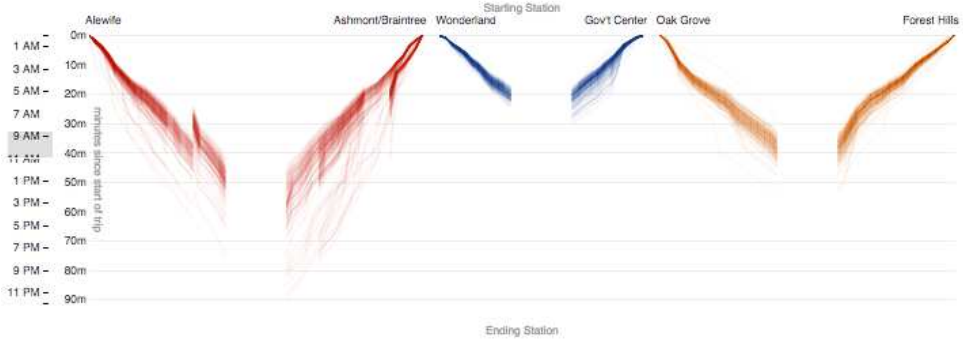


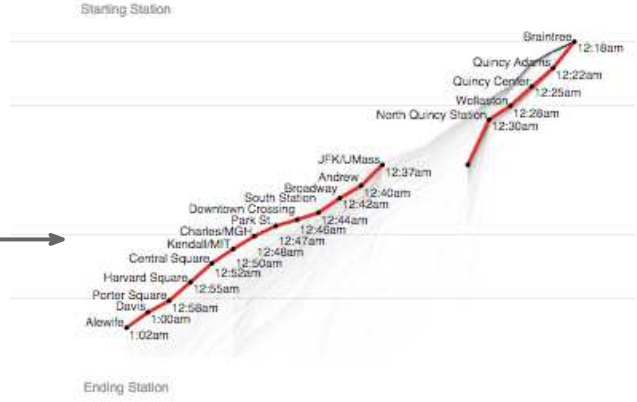


Each circle above and row in the table represent a station, hover over one to highlight the other. Next to each station are heatmaps showing entrances and exits to each station per-hour for weekdays and weekends/holidays.

Notice work stations with exit peaks in the morning and

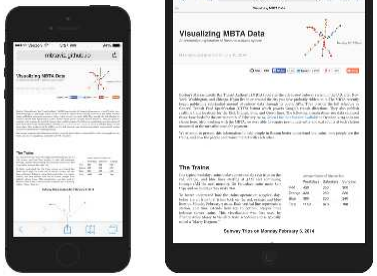


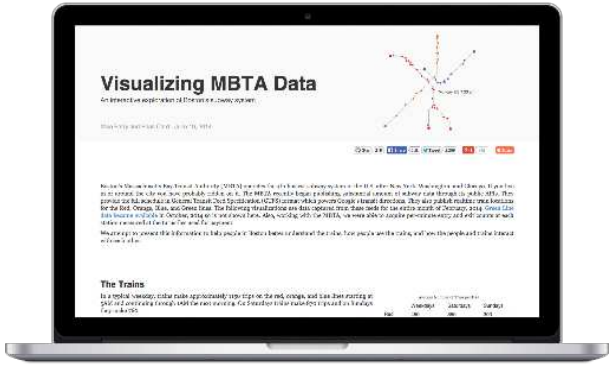


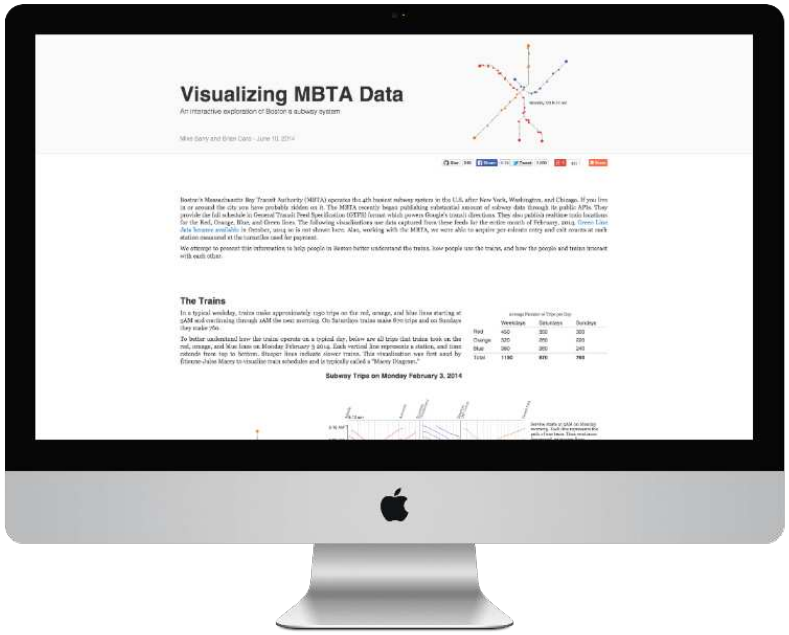


+ My Dad

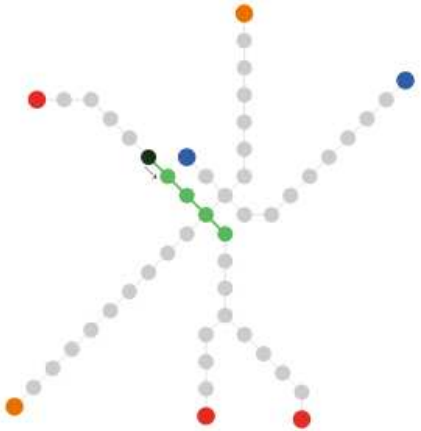
BeehiveMedia





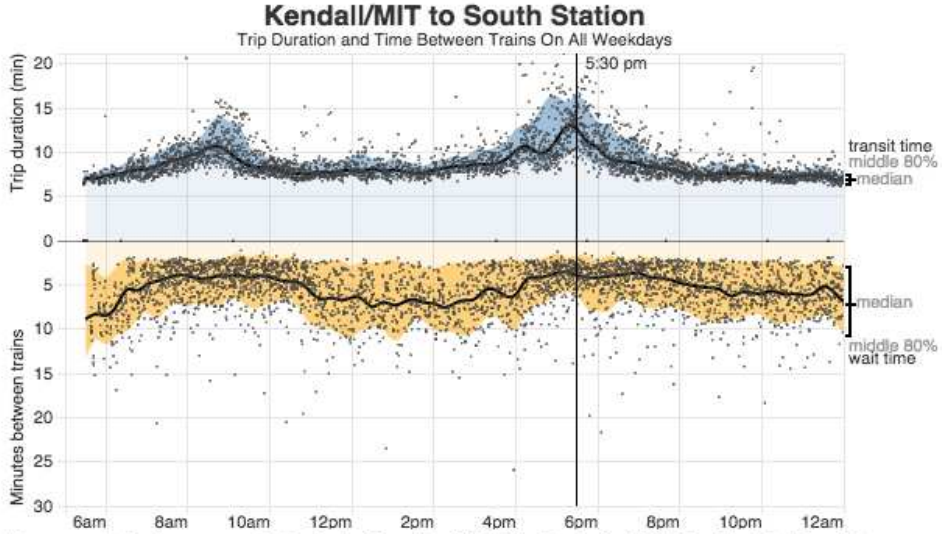


But How Long Is My Commute?

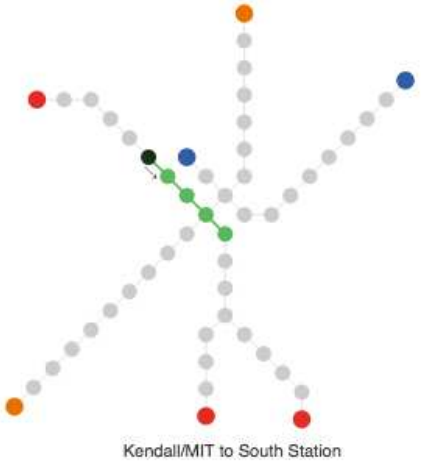


Kendall/MIT to South Station

Drag from a starting station to an ending station to see how long the trip takes over time in the chart.

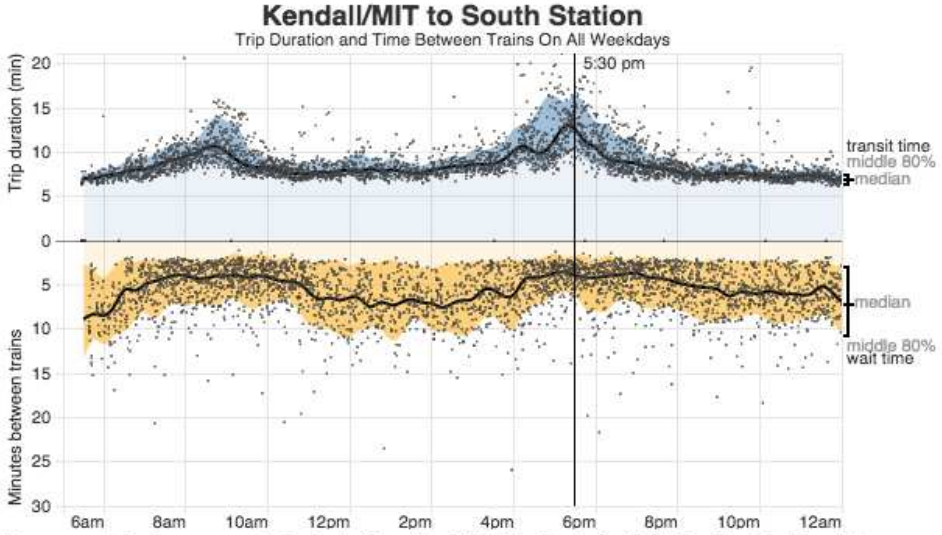


At 5:30 pm trains leave every 1 to 6 minutes from Kendall/MIT going to South Station. The trip takes between 10 and 17 minutes. The shortest time from when you walk into Kendall/MIT until you walk out of South Station is 10 minutes but it can be as long as 23 minutes. Usually it takes about 15 minutes including wait and transit time.



Kendall/MIT to South Station

Drag from a starting station to an ending station to see how long the trip takes over time in the chart.



At 5:30 pm trains leave every 1 to 6 minutes from Kendall/MIT going to South Station. The trip takes between 10 and 17 minutes. The shortest time from when you walk into Kendall/MIT until you walk out of South Station is 10 minutes but it can be as long as 23 minutes. Usually it takes about 15 minutes including wait and transit time.

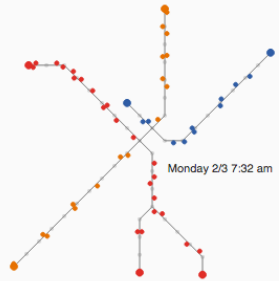
GitHub Pages

`git push origin master` → `http://mbtaviz.github.io`

Visualizing MBTA Data

An interactive exploration of Boston's subway system

Mike Barry and Brian Card - June 10, 2014



Star 219 Share 6.2k Tweet 2,300 +1 480

AddThis

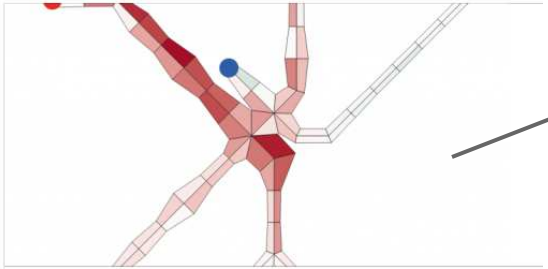
Google Analytics

Right now

1

active users on site

When shared, this is what will be included



Visualizing MBTA Data

An interactive exploration of ridership, congestion, and delay on Boston's subway system.

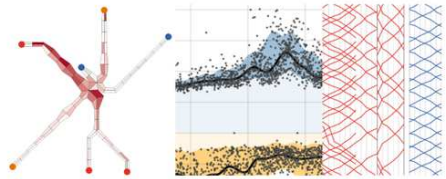
MBTAVIZ.GITHUB.IO

```
<!-- Open-graph metadata -->
<meta property="og:title" content="Visualizing
<meta property="og:description" content="An in
<meta property="og:type" content="website"/>
<meta property="og:url" content="http://mbtavi
<meta property="og:image" content="http://mbta
<meta property="fb:admins" content="9318920"/>

<!-- Twitter card metadata -->
<meta name="twitter:card" content="summary_lar
<meta name="twitter:site" content="@msb5014">
<meta name="twitter:title" content="Visualizing
<meta name="twitter:description" content="An i
<meta name="twitter:creator" content="@msb5014
<meta name="twitter:image:src" content="http:/
```

Mike Barry @msb5014

The card for your website will look a little something like this!

Mike Barry



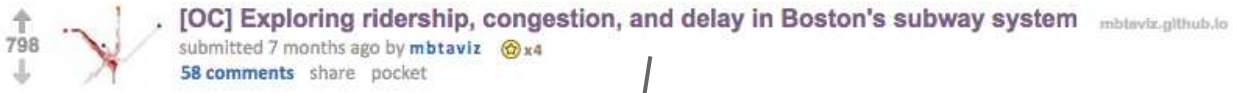
Visualizing MBTA Data

By **Mike Barry** @msb5014

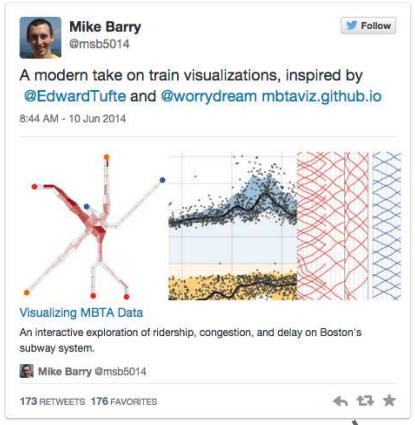
An interactive exploration of ridership, congestion, and delay on Boston's subway system.

View on web

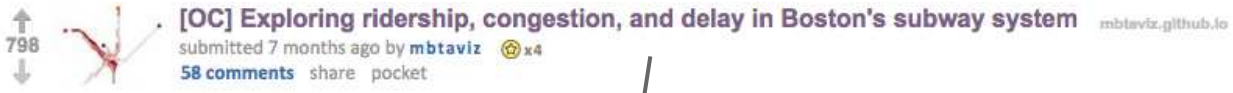

June 10
2014


798
[OC] Exploring ridership, congestion, and delay in Boston's subway system
mbtaviz.github.io
submitted 7 months ago by mbtaviz
x4
58 comments share pocket

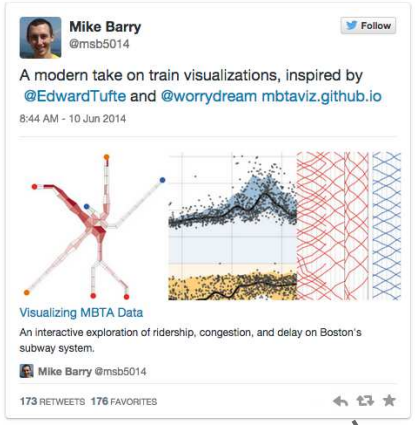

Hacker News new | comments | show | ask | jobs | submit
An interactive exploration of Boston's subway system (mbtaviz.github.io)
311 points by mbtaviz 214 days ago | comments | save to pocket


Mike Barry
@msb5014
Follow
A modern take on train visualizations, inspired by @EdwardTufte and @worrydream mbtaviz.github.io
8:44 AM - 10 Jun 2014
Visualizing MBTA Data
An interactive exploration of ridership, congestion, and delay on Boston's subway system.
Mike Barry @msb5014
173 RETWEETS 176 FAVORITES


9:00 AM
12:00 PM
3:00 PM
6:00 PM
9:00 PM


798
[OC] Exploring ridership, congestion, and delay in Boston's subway system
mbtaviz.github.io
submitted 7 months ago by mbtaviz
x4
58 comments share pocket


Hacker News new | comments | show | ask | jobs | submit
An interactive exploration of Boston's subway system (mbtaviz.github.io)
311 points by mbtaviz 214 days ago | comments | save to pocket


Mike Barry
@msb5014
Follow
A modern take on train visualizations, inspired by @EdwardTufte and @worrydream mbtaviz.github.io
8:44 AM - 10 Jun 2014
Visualizing MBTA Data
An interactive exploration of ridership, congestion, and delay on Boston's subway system.
Mike Barry @msb5014
173 RETWEETS 176 FAVORITES


Pat Barry @palbarry · Jun 10
@msb5014 Very impressive!


9:00 AM
12:00 PM
3:00 PM
6:00 PM
9:00 PM

798

[OC] Exploring ridership, congestion, and delay in Boston's subway system mbtaviz.github.io

submitted 7 months ago by **mbtaviz** x4

58 comments share pocket

Hacker News new | comments | show | ask | jobs | submit

An interactive exploration of Boston's subway system (mbtaviz.github.io)

311 points by mbtaviz 214 days ago | comments | save to pocket

Mike Barry
@msb5014
Follow

A modern take on train visualizations, inspired by @EdwardTufte and @worrydream mbtaviz.github.io

8:44 AM - 10 Jun 2014

Visualizing MBTA Data

An interactive exploration of ridership, congestion, and delay on Boston's subway system.

Mike Barry @msb5014

173 RETWEETS 176 FAVORITES

Mike Bostock @mbostock ·

@msb5014 Beautiful work!

View conversation

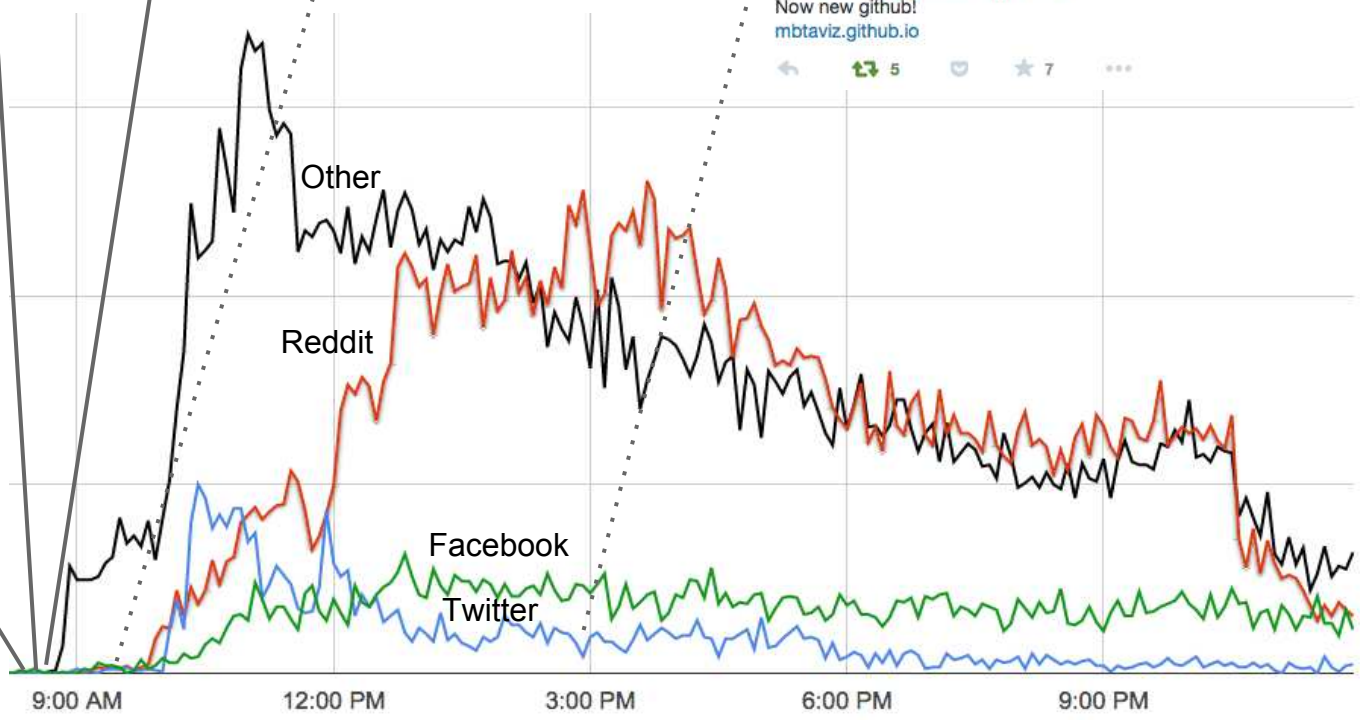
Edward Tufte @EdwardTufte · Jun 10

Graphical timetables once done by hand, big collection+ET NJTransit project: edwardtufte.com/bboard/q-and-a…

Now new github!

mbtaviz.github.io

5 7



Visualize All the Things

- All the free tools you need are at your fingertips
- Focus on answering questions
- Learn from the best
- Find your tools and stick with them

Questions?
Thank you!

Mike Barry @msb5014
Brian Card @bmcard

Backround images adapted from these sources

STS-135 Atlantis rollout 1 By NASA/Kim Shiflett [Public domain], via Wikimedia Commons.

Montinari Milano By André Karwath aka Aka (Own work) [CC BY-SA 2.5 (http://creativecommons.org/licenses/by-sa/2.5)], via Wikimedia Commons.

Open Book Policy (5914169915). By Alex Proimos from Sydney, Australia (Open Book Policy Uploaded by russavia) [CC BY 2.0 (http://creativecommons.org/licenses/by/2.0)], via Wikimedia Commons

MBTA Red Line train departing Quincy Adams station. By Ben Schumin (Own work) [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons

Repro Smoking Spaceman Robot – Ha Ha Toy – Silver - In Action!! By D J Shin (Own work) [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0) or GFDL (http://www.gnu.org/copyleft/fdl.html)], via Wikimedia Commons.

Stipula fountain pen By Power_of_Words_by_Antonio_Litterio.jpg: Antonio Litterio derivative work: InverseHypercube (Power_of_Words_by_Antonio_Litterio.jpg) [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons.