



**NJ TRANSIT**  
The Way To Go.

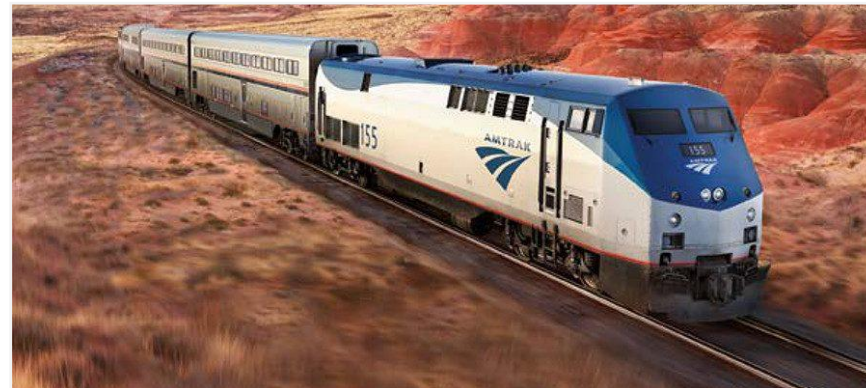
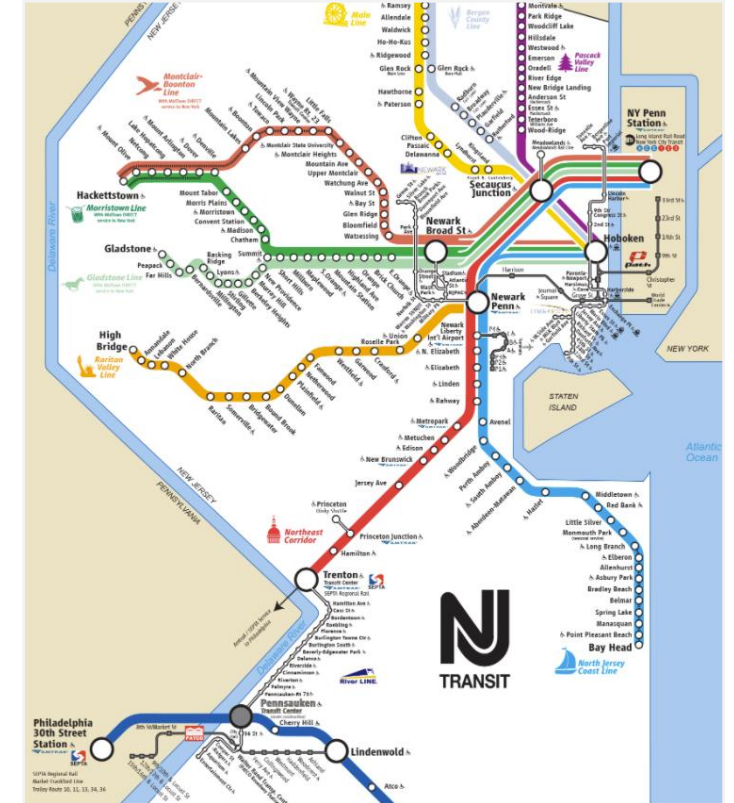
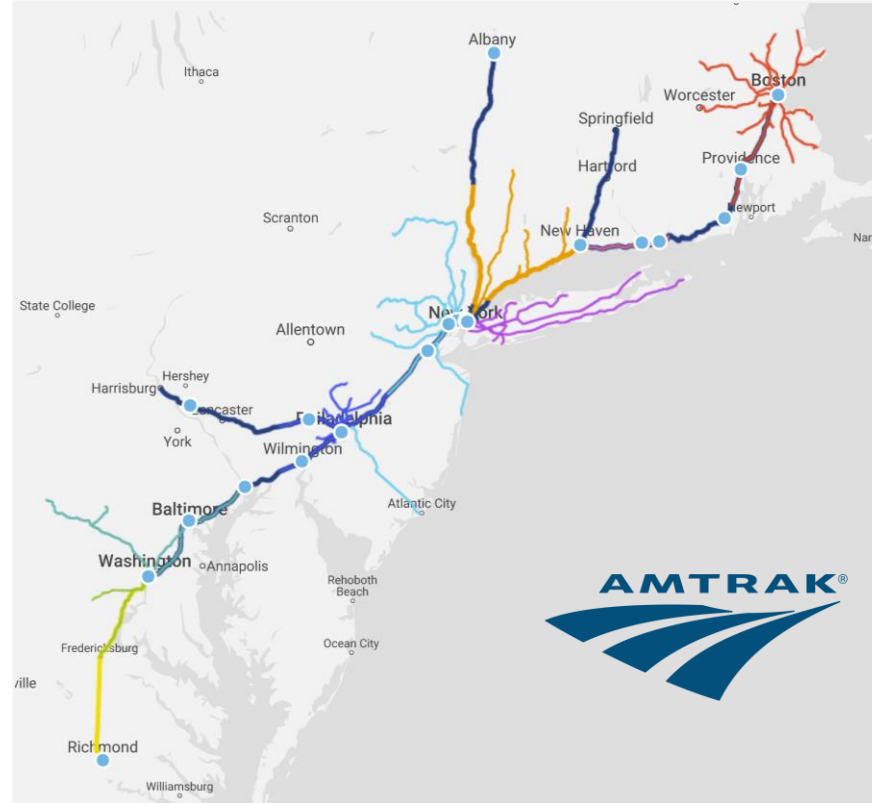


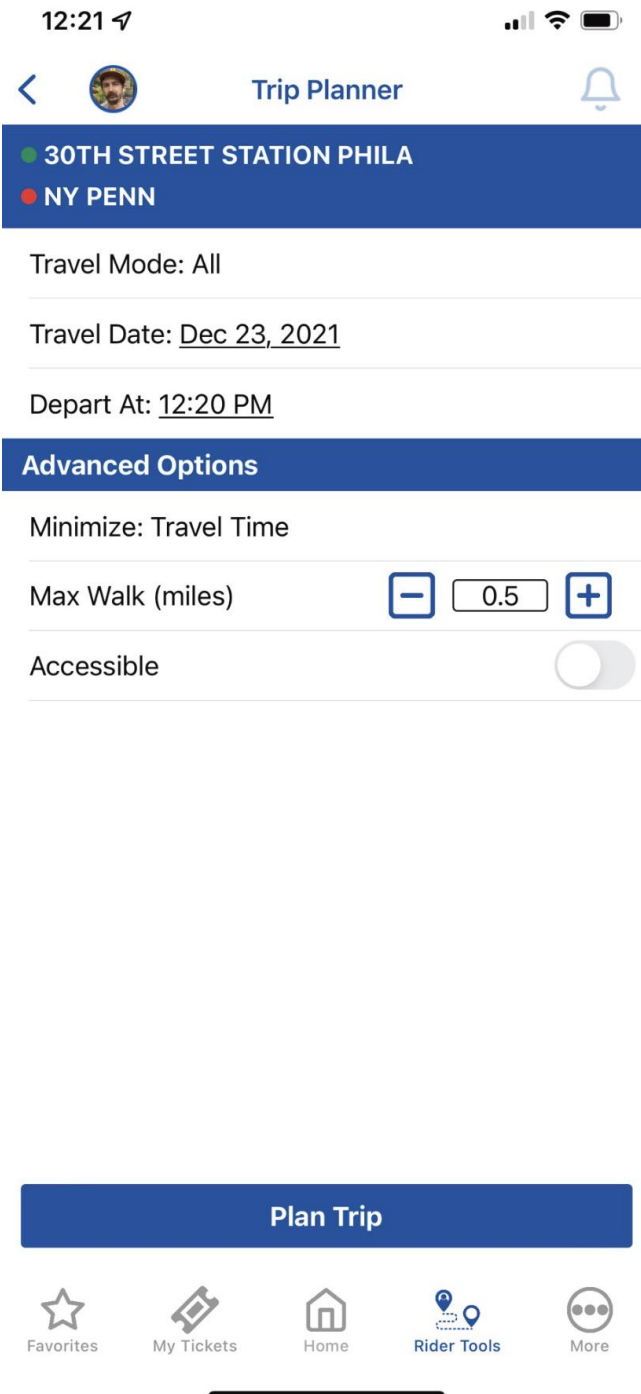
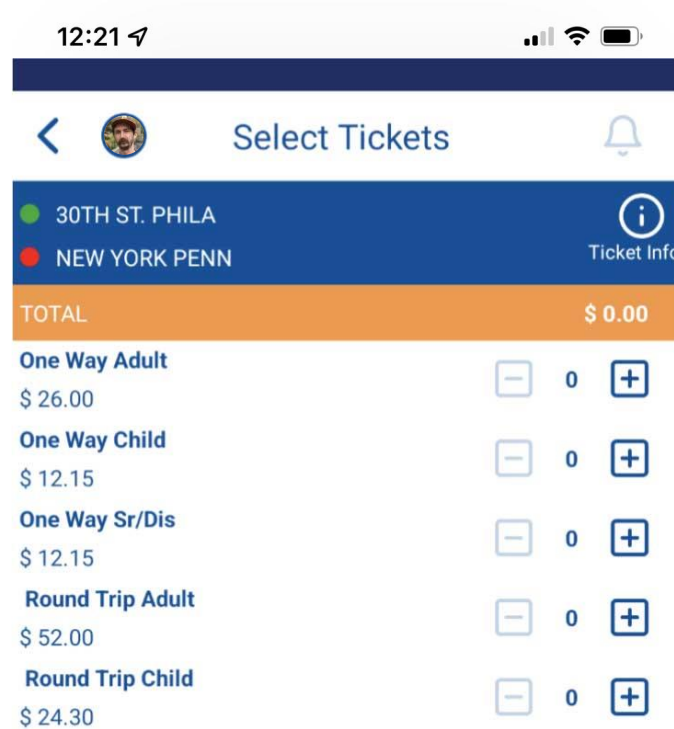
# A Forecasting Tool for New Jersey Train Delay

Jasmine Siyu Wu and Zoe Yoo

# Scope

- Initial idea to develop delay estimates for Amtrak and NJ online ticketing
- Similar geographies, but data constraints
  - No scheduling times in Amtrak data



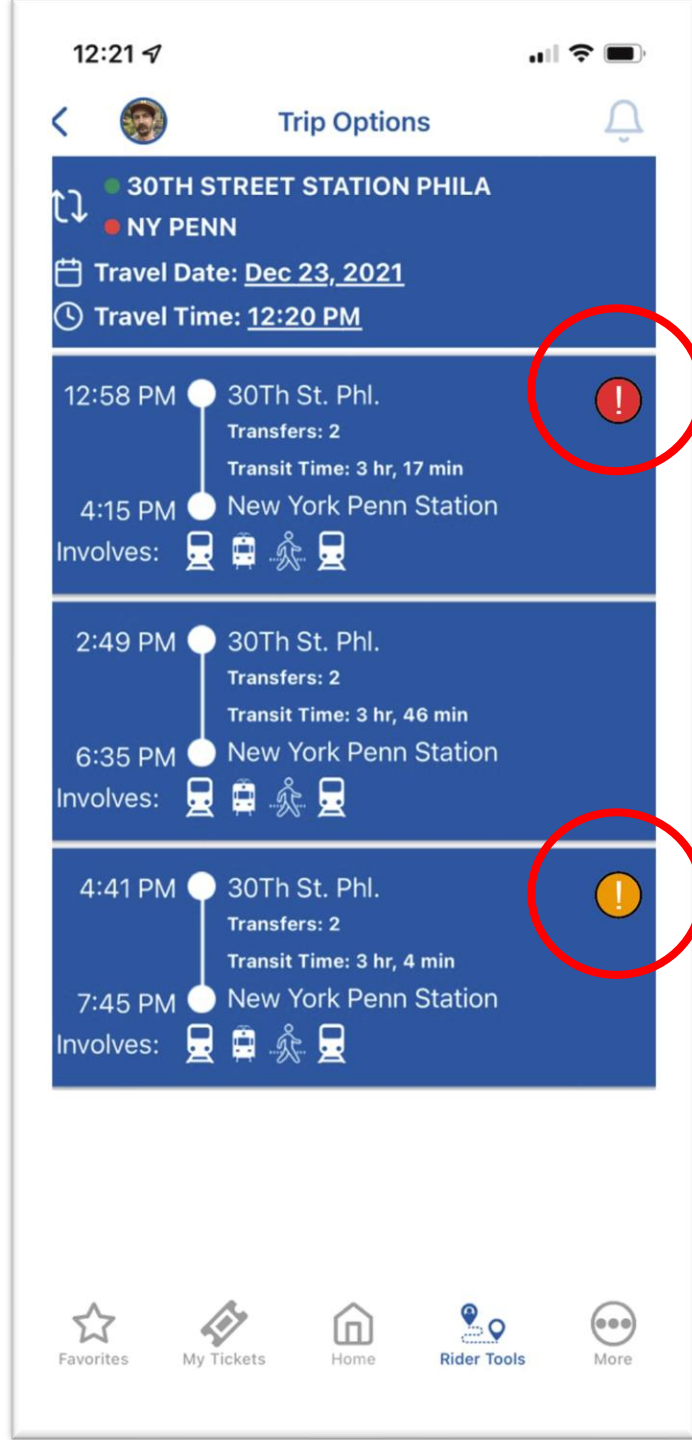


# NJ Transit

- New Jersey system differs from Amtrak
  - One-way tickets or passes, with no expiration date
  - Ticket prices vary based on distance but not time of reservation
- Tickets can only be bought on the NJ Transit Mobile App or in-person
- Trip Planner gives users scheduled trains based on date and time





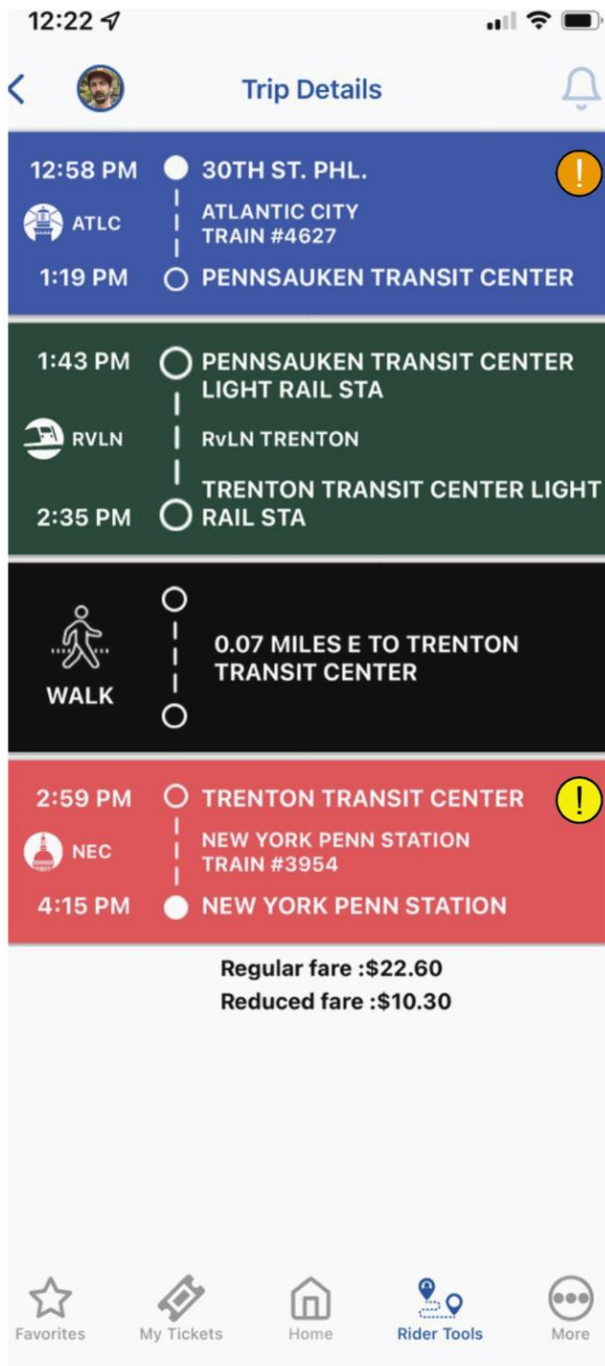


## Trip Option Delay

- Each time option shows the total predicted delay for that trip

- Red = >30 minutes combined predicted delay
- Orange = 20-30 minutes combined predicted delay
- Yellow = 10-20 minutes combined predicted delay





## Detailed Trip View

- Each leg of trip shows predicted possible delay
- Categorized below and summed for the overall trip
- Note: current data is only commuter rail, not light rail

- ! Red = >30 minutes combined predicted delay
- ! Orange = 20-30 minutes combined predicted delay
- ! Yellow = 10-20 minutes combined predicted delay



# Data Structure

---

- NJ Transit [DepartureVision](#) Real Time Train Status Service
  - Aggregated on Kaggle by Pranav Badami
- Sample months: January and February 2021

| Basic Data           | Spatial Data              | Temporal Data               |
|----------------------|---------------------------|-----------------------------|
| Train ID             | Start Station Name and ID | Scheduled Time of Departure |
| Type (Amtrak or NJT) | End Station Name and ID   | Actual Time of Departure    |
|                      | Stop Sequence Number      | Calculated Delay (minutes)  |
|                      | Line Name                 | Date                        |
|                      |                           |                             |

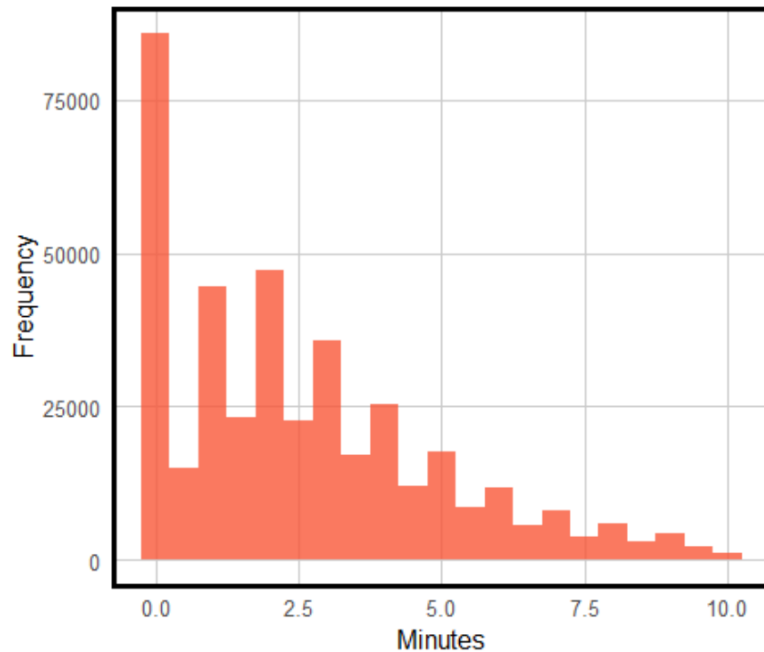


# Reviewing the Data

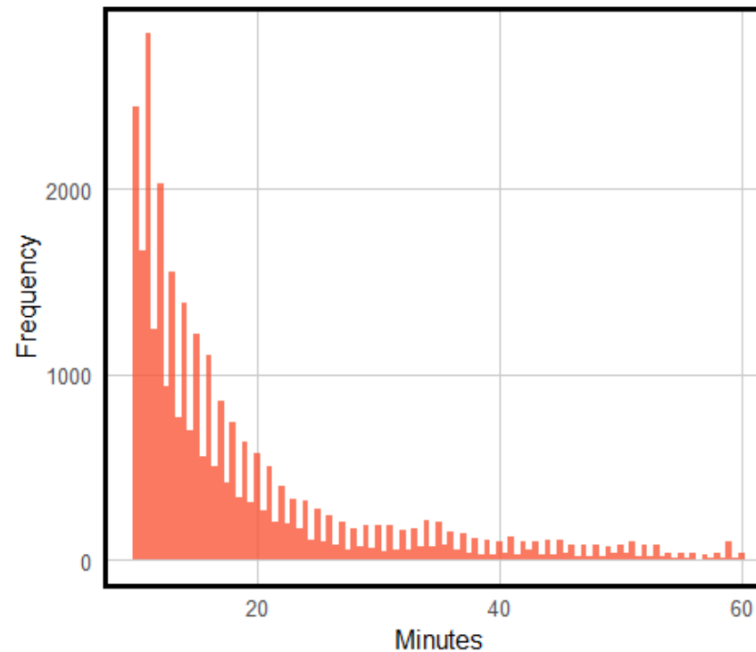
- Is delay often / severe enough to warrant a tool for it?

Delay Time of Trains at New Jersey Transit Stations, Jan-Feb 2020

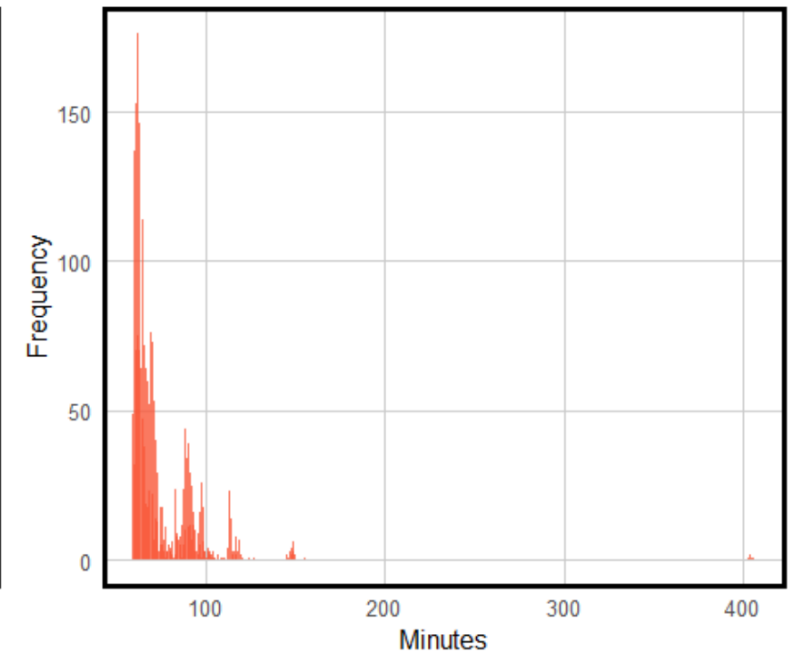
*within 10 minutes*



*10 minutes - 1 hour*



*over 1 hour*



# On Time

- **On time:** % of trains arriving at the scheduled time at each station stop on a journey

On-Time Rate of Trains at New Jersey Transit Stations

Jan-Feb 2020

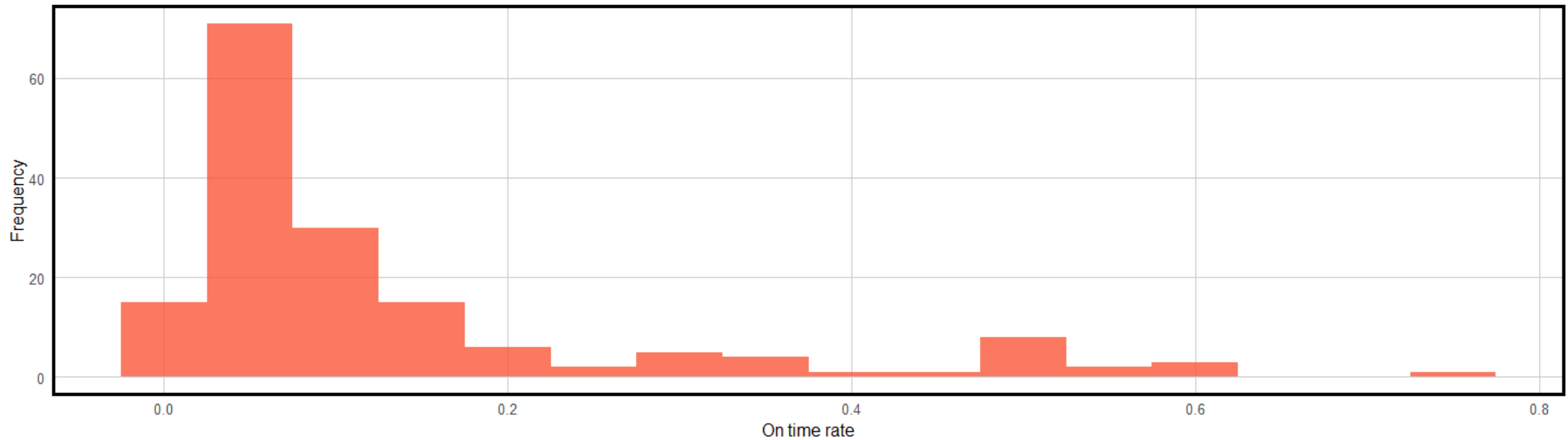


Figure 3.1



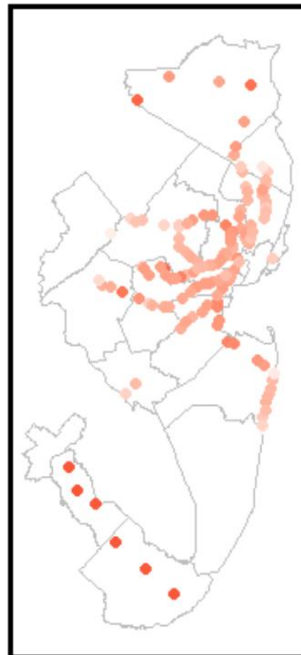


# Public Performance Measure (PPM)

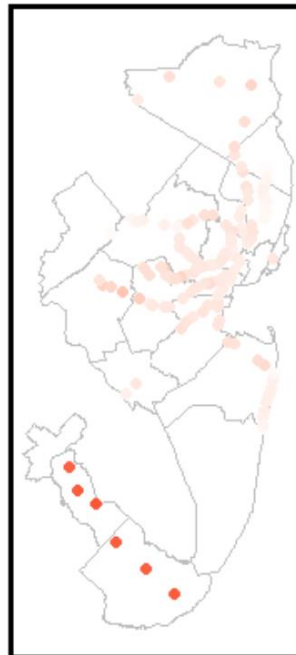
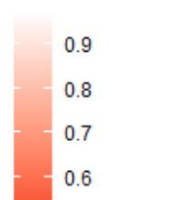
- **Public performance measure (PPM):** % of trains arriving at their destination within 5 minutes of schedule
- Philadelphia-Atlantic City Line has the lowest PPM

## Punctuality Rate of Trains by Station

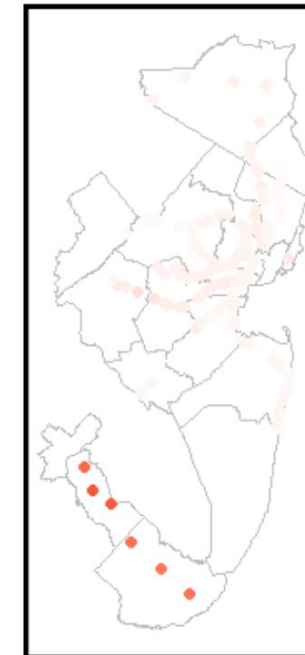
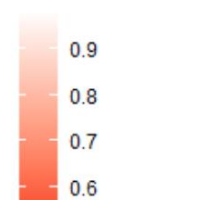
Jan-Feb 2020



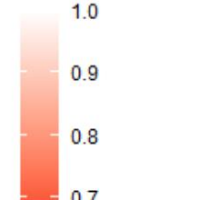
PPM 5 minutes



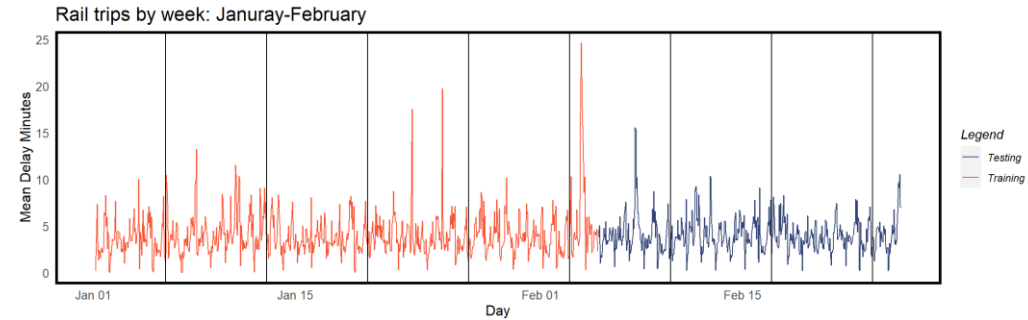
PPM 10 minutes



PPM 20 minutes



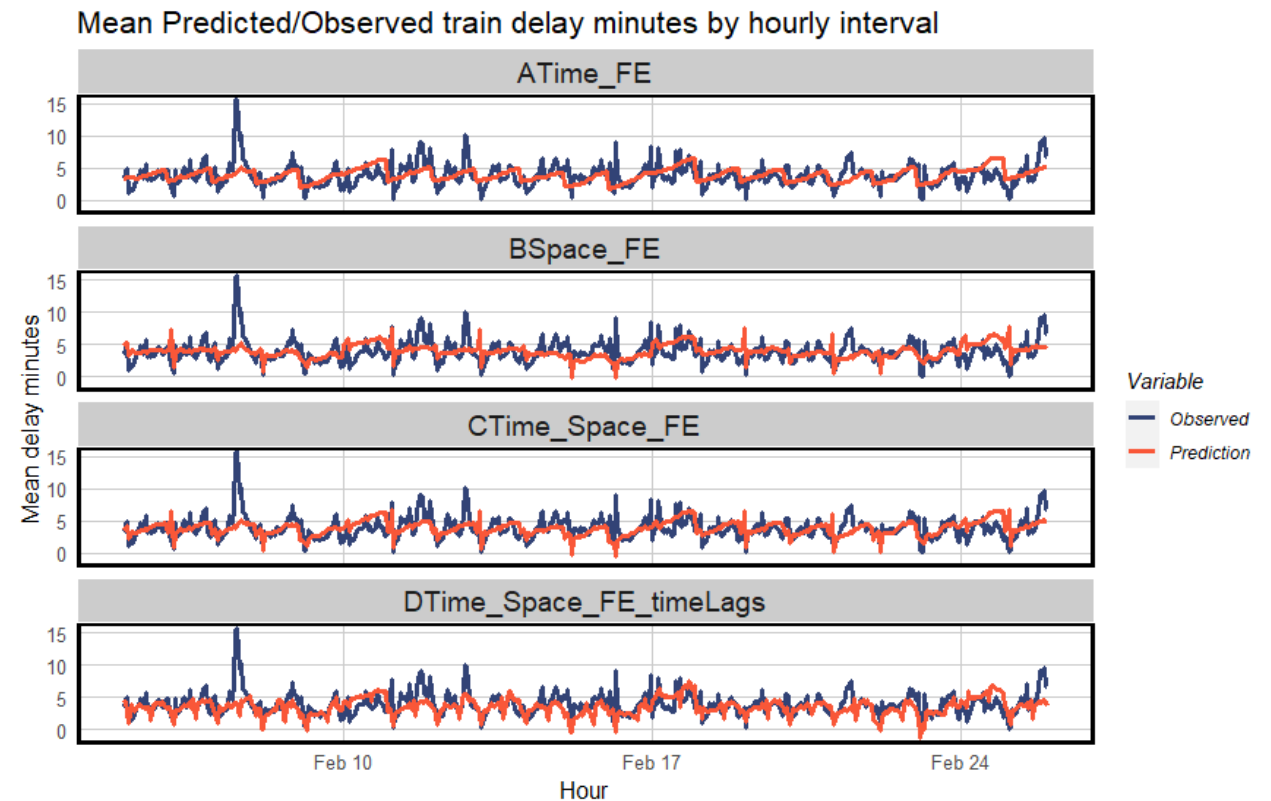
# Regression Model



**Split: Train on first five weeks, test on last three**

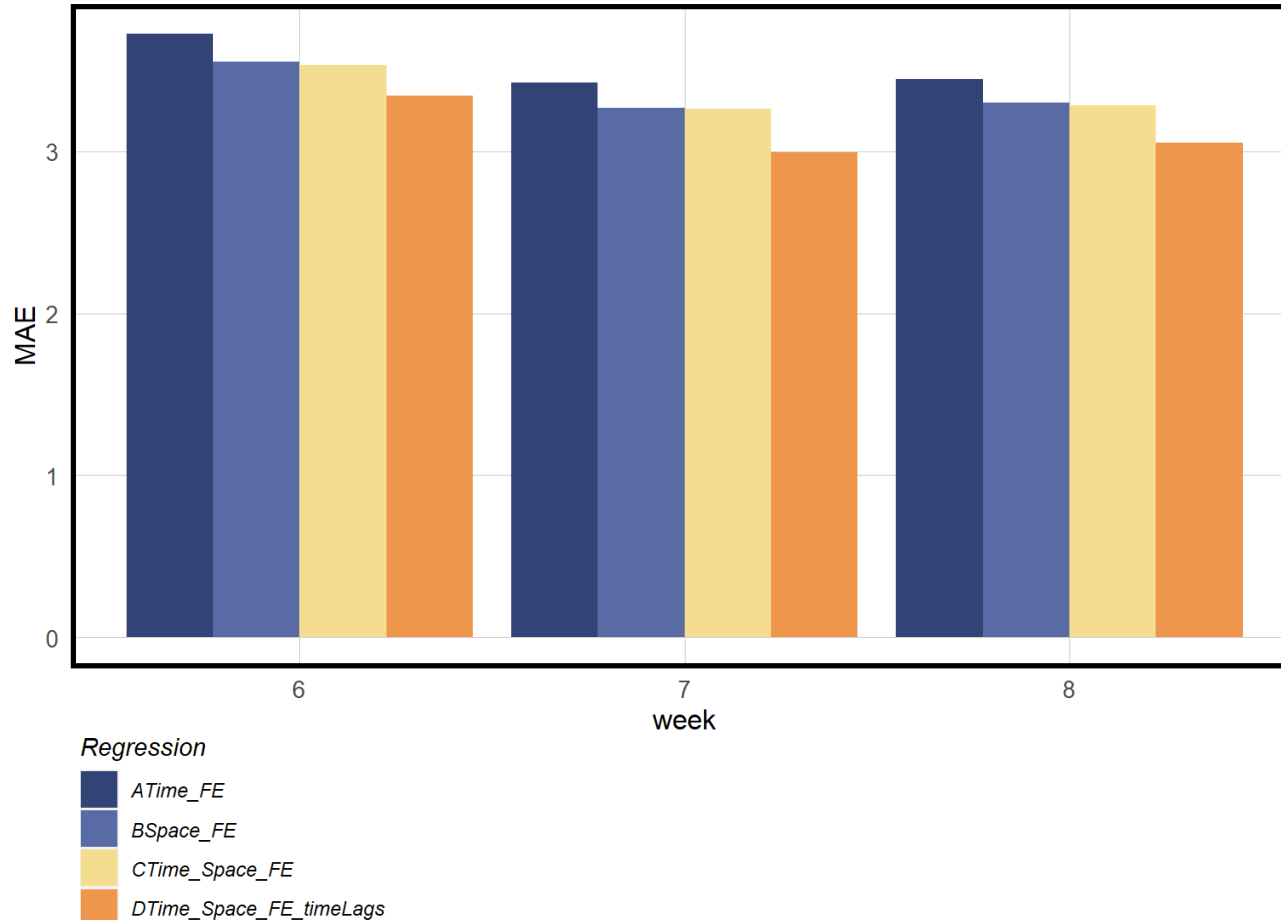
## Four different models:

- A.** Time – time of day, day of the week
- B.** Space – origin and destination stations
- C.** Time, Space, Weather (temperature/precipitation)
- D.** Time, Space, Weather, & Time Lag



# Results

Mean Absolute Errors by model specification and week



## Improving Accuracy

- Greater computing power
  - Use a larger sample size of data to ideally gain more accuracy year-round
- Improve on the current model
  - Add more variables, e.g. Number of previous stops
  - Add more independent data



# Future Possibilities

- Increase scope
  - Find scheduling data for Amtrak
  - Expand data scope to SEPTA or NJ light rail, which connect to NJ Rail Transit
- Other uses
  - Use to figure out the possible causes delay, and how to mitigate it, before it has the chance to affect customers

