

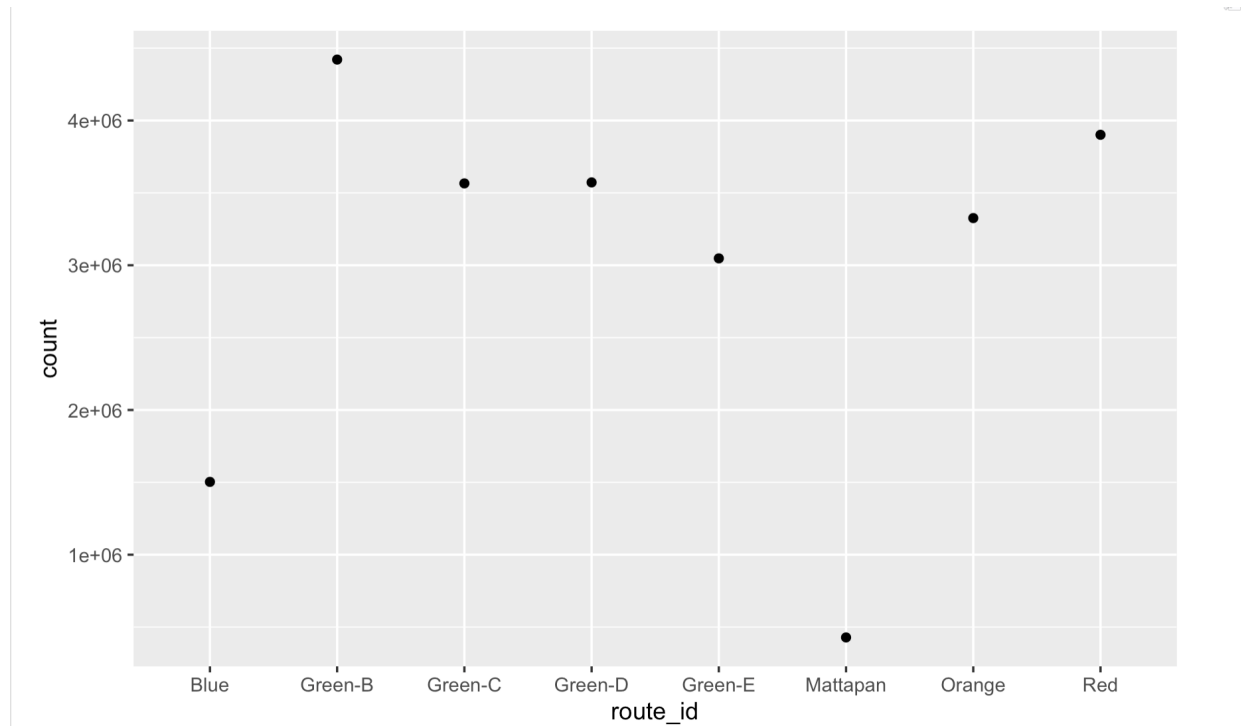
EDA Report

Hui Xiong
12/16/2022

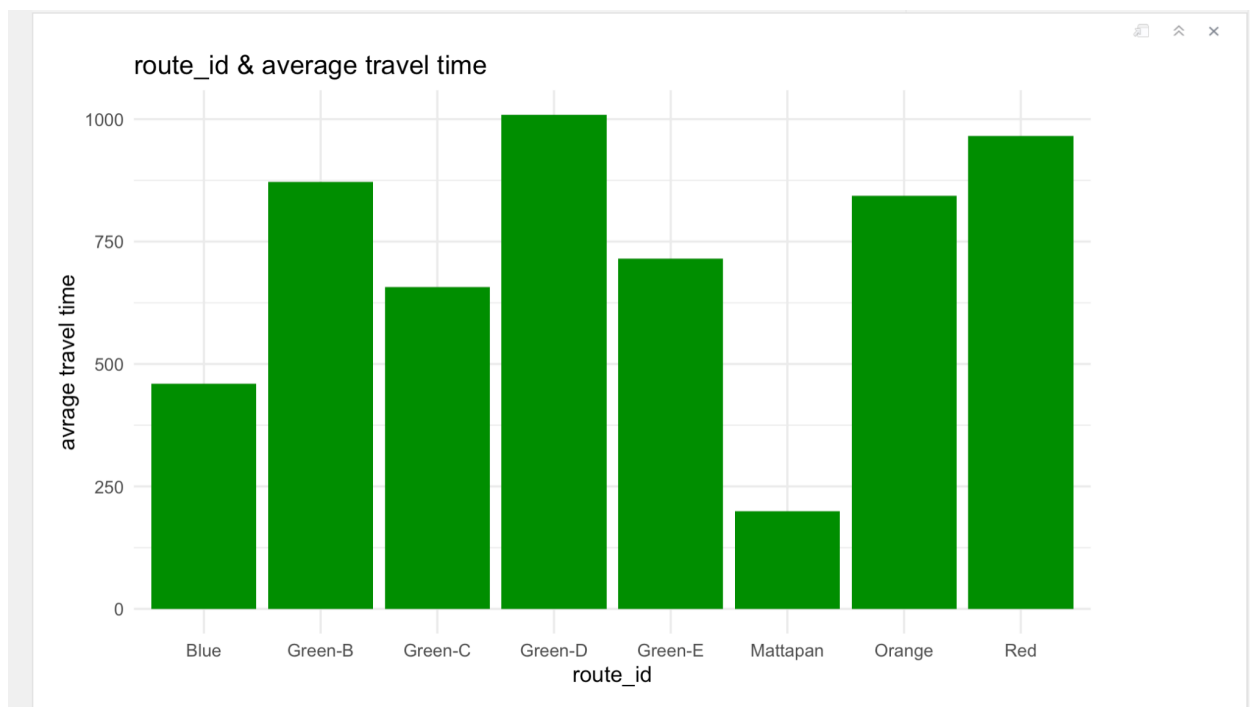
Introduction

MBTA is a public transportation agency. The MBTA operates four subway lines (the Red, Blue, Orange, and Green lines) and several rail lines that serve remote suburban areas. It also works as a bus.

About Subway

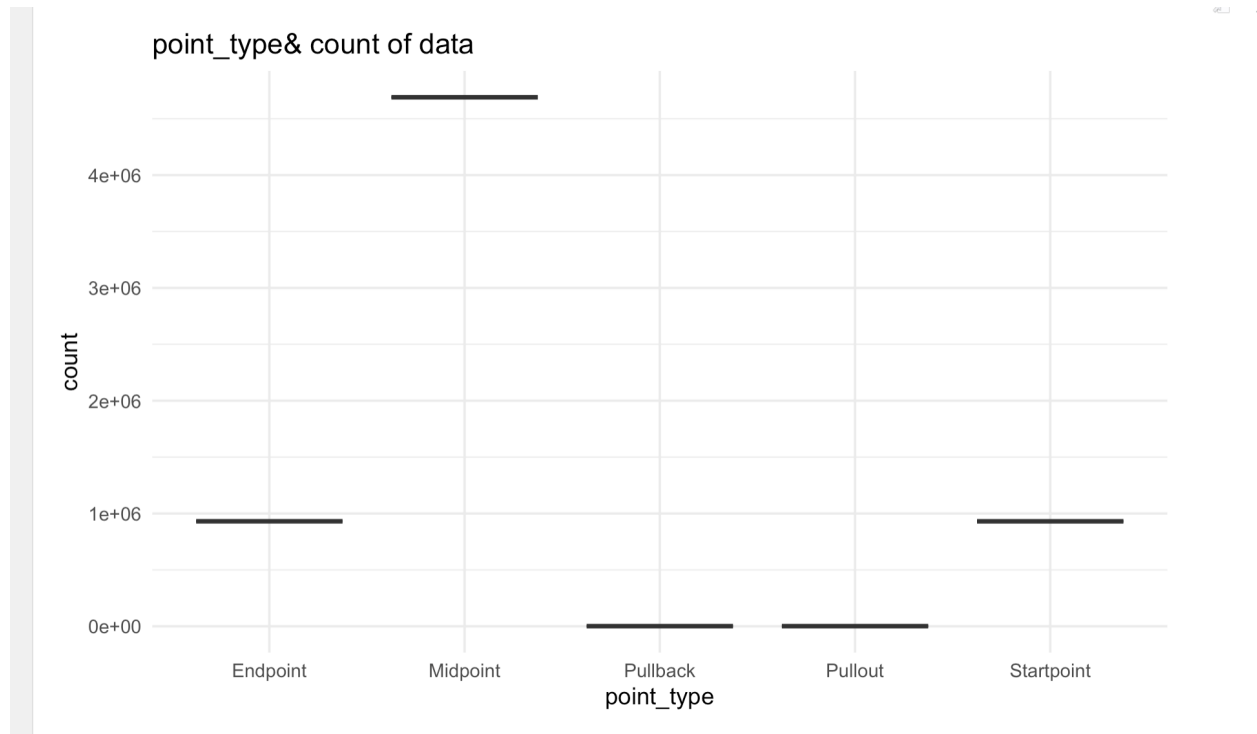


This graph mainly shows the frequency of departures. As shown in the figure, Green-B has the highest, followed by red, and Green-C is similar to Green-D. The lowest is Mattapan.



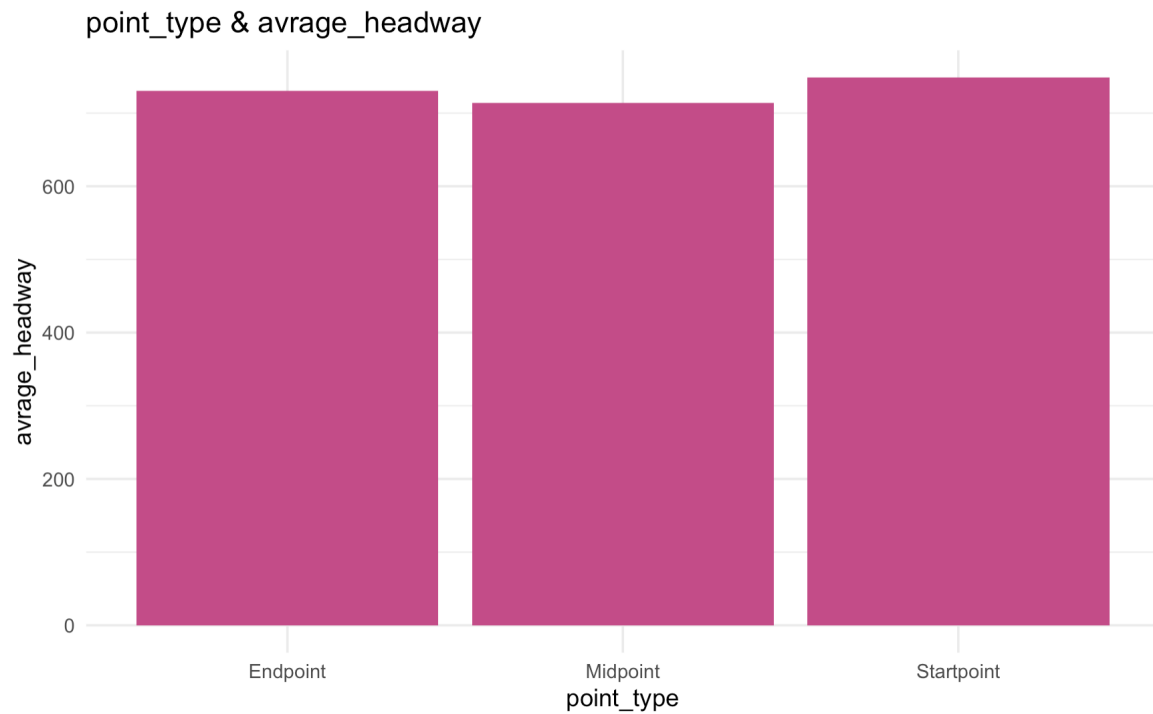
The figure shows that Green-D has the highest average travel time, followed by red, and Green-B is about the same as orange. The lowest is Mattapan.

About Bus

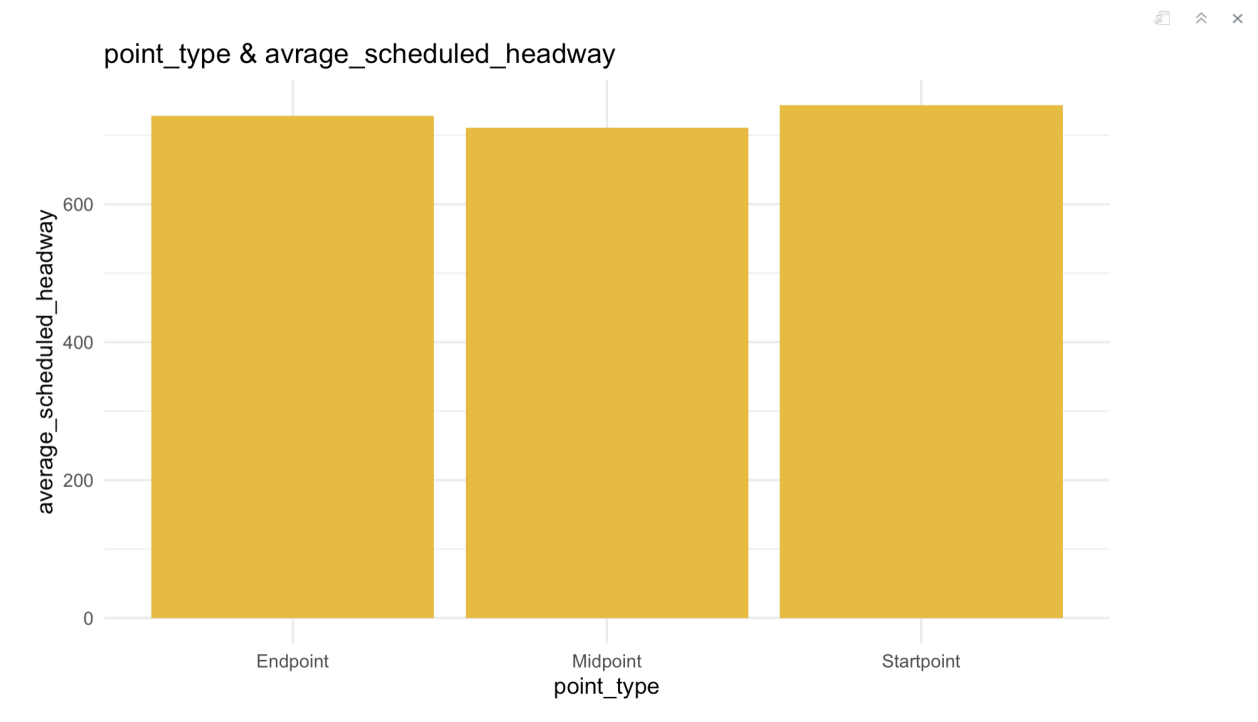


The argument I've chosen is point_type. The most point_type chosen is Midpoint, followed by similar Endpoint and Startpoint, and pullback and Pullout are the least.

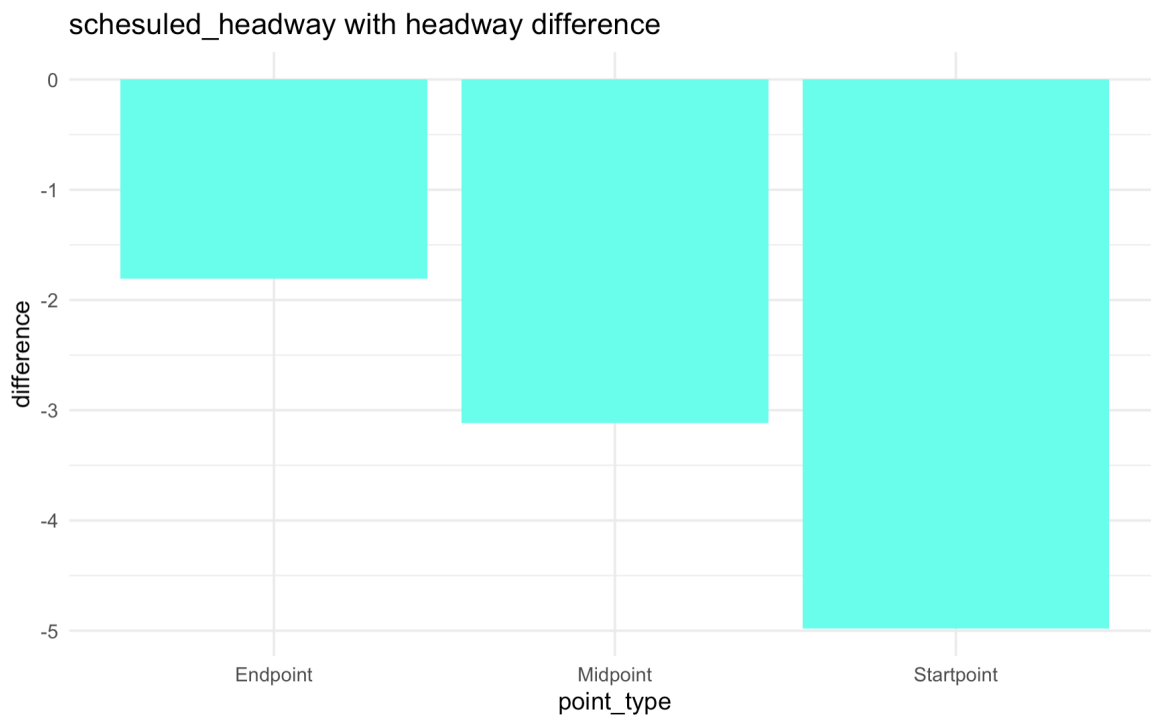
First, I cleaned up the data and removed the NA lines.



This is the graph of the point_type and the average headway. The highest in the diagram is Starpoint, the second highest is Endpoint, and the lowest is Midpoint.



This is the graph of the point_type and the average scheduled_headway.
The highest in the diagram is Starpoint, the second highest is Endpoint, and the lowest is Midpoint.



This is the graph of the point_type and the difference (average scheduled_headway & headway)

The average difference between scheduled_headway and headway is the least in Endpoint, and the largest in Startpoint.

##

1. I originally wanted to try using a radar chart, but I failed to draw the desired result after many attempts.
2. There are a lot of ideas about cleaning up data that I have yet to be able to implement in code, and I've tried a lot of different ways to clean up data, which is the only one that makes sense.
3. I don't know how to draw a bar chart with more than three variables. I tried to do it but failed.