Changes to Course Project

Joshua Soriano

2023-05-16

1 Alteration to y-scale

The code for the y-axis was altered to display the correct scale for the y-axis. A "fill" aesthetic was also added to differentiate between the MBTA Lines.

2 Proportionality of track length and restriction distance

This comment was removed, since the data set was changed to reflect a more date-accurate recollection of the speed restriction data.

3 Alteration to side-by-side plots

Instead of displaying four plots side-by-side, the code was altered to visualize data from a combined data set. This was feasible since the data was more date-accurate this time around. The use of a single plot also made it more readable, compared to the previous iteration which had somewhat small text. Since reporting the total track section length was not that useful, it was omitted in favor of the more visually-inclined fill aesthetic which detailed each MBTA Line by their respective color.

4 Comment on normality

The comment regarding normality was applicable to the new plot since the x-axis represented dates instead of ID values. A comment was also added regarding the shape of the plot, and how the sudden incline made the possibility of a normal distribution rather debatable.

5 Restriction reason per branch

The scatter plot was replaced in favor of a stacked bar plot which visualized the combined data. This made the manual legend unnecessary, and also made reading the plot easier, as the previous one had small text.

6 Z-test

Key statistics were added to the tables regarding the z-tests, and the code was altered to reflect a z-test instead of a t-test.

7 Alteration to Figure 2.1 and Figure 2.2

Figure 2.1 and Figure 2.2 were altered since they were too bulky, and did not make much sense due to the x-axis representing an arbitrary value. A box plot was used instead, and the x-axis represents a date instead of an ID value.

8 Figure 2.2 caption error

There was an error on the plot caption for Figure 2.2. Restriction speed was changed to restriction distance.

9 Changes to linear regression model

Track ID was once again changed to a date value to ensure accuracy for the linear regression model as well as the comments regarding the regression analysis.

10 Changes to chi-square test

The chi-square test was altered to test the correlation between restriction speed and location description. Since each location is tied to a specific branch, we were able to revisit Figure 1.2 and make a comment regarding this correlation. The density plot was also changed to a balloon plot to visualize this analysis.

11 Further Analyses

While further analyses were indeed feasible, unfortunately I could not apply them due to time restrictions.