Group 1.

You did a great job on your presentation. There were some potential concerns in your results, so I’d like to comment on them here.

1. Duplicated columns when merging the two data sets

See your column names. Many column names end with “.x” because of the duplicated column entries. It often happens when merging files. You need to select the necessary columns first to avoid this issue. It may look OK to you, but it confuses others.

2. Chi-square test

It is clever to apply the chi-square test, but you need to consider what these two categories represent. In class, we discussed two examples: male/female counts and race counts. Each category sums up the number of respondents, which may represent the total population. Now, when you sum up the number of gas stations and public transportation observed, what would this number imply? You may need to consider “others” if you mean the mode of transportation.

3. Normalization

When comparing the two statistics, it would be wise to normalize them. Otherwise, we could not capture the variations of the smaller statistics. You need to address more clearly what you normalize and why, especially in the time series. For the graphic results in “Public versus private transportation”, the pattern for “private” is well captured, but “public” looks quite mundane. It is because you normalize based on the max pattern of private.

Agenda setting: Good

Presentation Quality: Good

Code Quality: Fair (-1)

Code efficiency: Fair (-1)

Further consideration: Good for future work (+0.5)