Ethan Arsht

8 February 2023

Research Question and Data

I will aim to answer the following research question: to what extent do CTA "ghost-buses" and/or "ghost-trains" follow a geographically clustered pattern?

At the onset of the COVID-19 pandemic, CTA users observed a dramatic divergence between the Chicago Transit Authority's scheduled service and actual service. Specifically, buses and trains had increased delays, or simply failed to arrive at any point. A CTA vehicle failing to conduct its scheduled route is known as a "ghost-bus" or "ghost-train". Service degradation has continued despite the conclusion of the early phase of the COVID-19 pandemic.

I hypothesize that the distribution of ghost-buses and/or ghost-trains is not equally distributed throughout CTA infrastructure, nor is it distributed in a way proportional to CTA's scheduled service. To prove this hypothesis, cluster analysis will be necessary to see if CTA ghosting is clustered in ways distinct from the clustering of CTA services toward population centers. Contingent on the result of that hypothesis, I hypothesize that the greatest relative intensity of CTA ghosting corresponds with disadvantaged areas.

The data for this project will likely come from the CTA API, as well as the CTA and Metra Reliability dashboard. If reliable historical CTA actuals and schedules are not attainable from those sources, I will look into a modified research question. A CTA ghost-bus/train is represented by the point of the stop or station where the bus or train failed to arrive. The first hypothesis can therefore be tested using point pattern analysis. The second hypothesis will involve merging CTA data to neighborhood-level data, so both point data and areal data will be used in the project.