



Bus pending

Final project for CAPP 30122

Michael Rosenbaum Keling Yue Daniel Muñoz Regina Isabel Medina Rosales

Winter 2024

For code revisions, please visit the github repo.

1. Abstract

We aim to visualize inequality in transit delays in Chicago based on CTA's data on bus positions. The Chicago Transit Authority (CTA) regularly reports on bus positions and bus schedules in the city of Chicago. We will scrape these reports on a regular interval of one minute (below the daily API call limit of 10,000) to gather bus and train locations and then calculate which stops are regularly delayed. For the analysis, we will merge CTA's datasets with publicly available data on community areas to calculate which areas regularly experience delays. We will then visualize those delay locations. If we have the capacity to do so, we will estimate if delays are correlated with demographic characteristics drawn from the American Community Survey.

2. Data Sources

2.1. Bus data

2.2. Train data

2.3. Geo data

- Neighborhood shape files
- Zip code shape files
- Census units shape files



3. Pending work

2.4. Sociodemographic data

3. Pending work

1. Scraping

- Get the schedule data downloaded
- Scrape the stops
- Link the schedule data to the stop data and to the scraper data
- Set-up the cron job on the server that James spins up
- Documentation

2. Cleaning

3. Analyzing

4. Visualizing