Gretchen Klock

Professor Koylu

GEOG: 4580

14 December 2023

Chicago Mass Transit Database Project

Introduction

Chicago, Illinois is the third largest city in the US and has the second largest heavy rail transit system and the third largest bus system by ridership in the country. Chicago transit is managed by the Chicago Transit Authority (CTA) which operates the cities ‘L’ system (elevated rail system) and bus system. Despite the large size of the city’s transit system, studies have indicated that access to transit is unevenly distributed throughout Chicago. Specifically, areas with a higher percentage of African-American, Hispanic, Asian, low-income, low-educated and elderly residents have disproportionately lower levels of accessibility to these mass transit systems in Chicago. These cohorts have lower access to important public spaces including jobs, parks, groceries, hospitals, and libraries.[[1]](#footnote-1)

This project is exploratory. It could be used to fuel other research questions. Mainly it serves as an exploration of the data available from the city of Chicago via the online Chicago Data Portal in the context of the city having unequal access to transportation. Potentially, a project like this could be used by traffic planners to assess which areas of the city are underserved, but that purpose would require more data and more polishing. Expanding the project to be used by transportation planners would help to satisfy Ermagun and Tilahun’s call for decision makers and planners to prioritize equity and accessibility in future planning endeavors.

While this project is exploratory, the goals of this project are to explore relationships between the distribution of bus and rail lines/stops throughout Chicago in relation to the distribution of difference demographic groups and to explore the density of transportation options by neighborhood. Specifically, this includes assessing areas well served by CTA bus lines and CTA rail lines within the city boundaries of Chicago. Ultimately these goals and objectives are outside of the scope of the data available, however, these questions guide and motivate the project.

Database Design and Implementation

All data is sourced from the Chicago Data Portal and the Chicago Health Atlas which are both owned and managed by the city of Chicago. Datasets from the Chicago Data Portal include bus stops, bus routes, rail stops, rail stops, and hospital locations. Datasets sourced from the Chicago Health Atlas include demographics, community area (neighborhood) boundaries. Data concerning ridership was only available as a measure of daily total boarding across both the bus and rail systems and thus was not useful for this project.

All spatial data was transformed into the coordinate system EPSG: 3435— NAD83 Illinois East (ftUS). The datum is NAD83. All measurements for queries are in feet as this is the default unit of the coordinate system.

Bus Routes: Shapefile, 128 Rows, Columns: id, geom, route, name. https://data.cityofchicago.org/Transportation/CTA-Bus-Routes-Shapefile/d5bx-dr8z/about\_data.

Bus Stops: Shapefile, 10760 Rows, Columns: id, geom, street, cross\_street, systemstop, routesstpg. <https://data.cityofchicago.org/Transportation/CTA-Bus-Stops-Shapefile/pxug-u72f/about_data>.

Rail Lines: Shapefile, 153 Rows, Columns: id, geom, name. <https://data.cityofchicago.org/Transportation/CTA-L-Rail-Lines-Shapefile/53r7-y88m/about_data>.

Rail Stations: Shapefile, 144 Rows, Columns: id, geom, name. <https://data.cityofchicago.org/dataset/CTA-L-Rail-Stations-Shapefile/vmyy-m9qj/about_data>.

Demographics: csv, 77 Rows, Columns: id, name, whitepct, blackpct, asianpct, hispanicpct, native\_americanpct, pacificpct, 2ormore races pct, otherpct. <https://chicagohealthatlas.org/indicators/PCT?topic=demographics>.

Hospitals: shapefile, 44 Rows, Columns: id, facility, address, commonname. <https://data.cityofchicago.org/Health-Human-Services/Hospitals-Chicago/ucpz-2r55>.

Neighborhoods: shapefile, 98 Rows, Columns: id, geom. <https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Neighborhoods/bbvz-uum9>.

Public Schools: shapefile, Rows: 661, Columns: id, geom, address, short\_name. https://data.cityofchicago.org/Education/Chicago-Public-Schools-School-Locations-SY1718/4g38-vs8v

A diagram on a piece of paper

Description automatically generated

Analysis

Data was sourced from the Chicago Data Portal and Chicago Health Atlas, imported into qgis, projected, and imported the database with pgadmin. The column routesstpg in the Bus Route table violates first normal form as it lists the routes that use each stop as a list separated by a comma within the tuple for each individual stop. I looked into using python to separate these values but was unsuccessful unfortunately. I was able to run the code for the first row of the csv I was looking to separate the values for, however I could not seem to create a script that repeated with the help of ChatGPT.

Results

This project is exploratory so the queries are example queries that demonstrate the relationships within the datasets.

* Query 1: How many bus stops are in each of Chicago’s neighborhoods?

A close-up of a white background

Description automatically generated

A map of a city with red squares

Description automatically generated

There is a large range of values for the total number of bus stops in each neighborhood ranging from 4 to 365. This indicates a possible uneven distribution of mass transit access. However, these are raw counts that are not normalized. It is improper practice to create a choropleth map of raw counts. In the future this data could be normalized by population or a similar measure.

* Query 2: How many rail stations are in each of Chicago’s neighborhoods?

A close up of a text

Description automatically generated

A map of a city

Description automatically generated

The values of the count of rail stations per neighborhood ranges from 0 in many outer (western/southern) neighborhoods to 16 in the Near South Side neighborhood that is just south of the Loop (the downtown central business district). This indicates a possible uneven distribution of mass transit access, especially as many neighborhoods on the South Side (which tend to have higher proportions of Black residents) lack rail stations. Again however, these are raw counts that are not normalized. It is improper practice to create a choropleth map of raw counts. In the future this data could be normalized by population or a similar measure.

* Query 3: What percent of residents are White in the 4 neighborhoods with the most rail stops?

A computer screen shot of text

Description automatically generated A screenshot of a computer

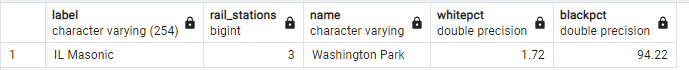
Description automatically generated

The percentage of White residents in the 4 neighborhoods with the most rail stops ranges from 1.72-52.38 percent. This is a wide range. This does not indicate a disproportionate amount of rail stops for White populations.

* Query 4: Which hospital is most accessible by the ‘L’?
* Which neighborhood is it located in and what is that neighborhood's demographics?

A close-up of a computer code

Description automatically generated



This query counted the number of rail stations within half a mile of a hospital. The IL Masonic Hospital is the only hospital with 3 stations. However, each of these stations service the Brown Line. Therefore, only 1 station is actually practical for reaching the hospital. Therefore, this query is not comprehensive enough to answer the question of hospital access.

Conclusion

Overall, this project needs more refining. Additional data on population density and CTA ridership would better equip this database for actually answering questions on the accessibility of the CTA. As the project stands now, it is clear that there is a need for more research as there is the possibility that mass transit access is uneven in the city. I believe the querying went well and some of the queries are as complex as ones would be on assignments. However, each has its limitations for actually answering questions. These queries do work for data exploration purposes. If I had more time, I would keep trying to normalize the choropleth maps and to clean the bus stops table with python. Both tasks I repeatedly attempted but with no results. Overall, this is a topic I am interested in and a project like this could have practical use for policy makers if the limitations and data shortcomings were addressed.

1. Ermagun, Alireza, and Nebiyou Tilahun. “Equity of Transit Accessibility across Chicago.” *Transportation Research Part D: Transport and Environment* 86 (September 1, 2020): 102461. <https://doi.org/10.1016/j.trd.2020.102461>. [↑](#footnote-ref-1)