THE UNIVERSITY OF CHICAGO

Chicago Public Transportation Accessibility Analysis

By

Guangyu Cai

April 2024

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree

in the Master of Arts Program in the Social Sciences

Preceptor: Marshall Ryan Jean

Introduction:

As Chicago strives to become one the busiest cities in the world, the Chicago Transit Authority (CTA) has also become the second largest transportation system in the United States. The CTA has 1,868 buses that operate 127 routes and 1,514 route miles. Buses make about 15,943 trips a day and serve 10,633 bus stops. On the rapid transit system, CTA's 1,480 rail cars operate eight routes and 224.1 miles of track. CTA trains make about 1,888 trips each day and serve 145 stations (CTA Facts at a Glance, 2023). Public transportation like buses and subways connect people to their jobs, schools, healthcare, recreation, and other essential services. They are affordable and convenient choices for people that cannot easily utilize private vehicles and subways. They also provide a more sustainable mode of transportation, reducing traffic congestion and carbon emissions. The accessibility of public transportation gives us feedback on urban planning and public need for transportation. It can identify transportation deserts, by analyzing the coverage and distribution of public transportation services. Public transportation accessibility metrics, such as frequency of service, travel times, and connectivity, can also indicate the quality of transit service in different parts of the city. Urban planners and policy makers require such specialized feedback to make informed decisions and create more inclusive and sustainable cities. Residents depend on reliable and convenient public transport to access public services, job opportunities, educational institutions, healthcare facilities, and other important destinations (Ermagun & Tilahun, 2020). Answers to public transportation accessibility distribution will directly impact these parties by informing them how well public transports serve Chicago's diverse population and where improvement are needed. 'Transportation agencies typically have stated goals around public participation and depending on the entity and funding source, usually federal requirements as well. In fact, various federal mandates require participation that includes of people of color, but equitable participation may have limited impact in practice’ (Lowe & Jones, 2023). Therefore, this study will identify potential disparities in public transportation accessibility in Chicago and explore the relationships between these disparities and demographic distributions such as race and poverty. I will conduct a comprehensive analysis of the public transportation accessibility in Chicago to uncover the underlying issues that may be contributing to disparities in access. The transportation accessibility will be quantified by calculating the public transportation surplus index in each census tract area on the base of GTFS (general transit feed specification) data. I will then find correlations between the surplus index and demographic variables using a regression model. All the visualization outputs will be organized in an ArcGIS story map. This research can address the very essence of social justice within Chicago's public transportation infrastructure.

Literature Review:

There are numerous ways to interpret accessibility. For instance, one meaning of it is how easy it is for people with disabilities to access public infrastructures. Geographic researchers often define accessibility as the relative ease of reaching a location or facility. This study focuses on accessibility in terms of distribution and frequency rather than just the convenience of access. The aspect of accessibility in terms of distribution and frequency considers how well public transportation services are spread out across a geographic area and how frequently they operate. On the other hand, convenience of access focuses more on the ease with which individuals can reach the public transit stops. It considers factors such as walking distance and presence of barriers. Different measures are used to calculate the accessibility index. Jay and his research team used a methodology adapted from the London Borough of Hammersmith and Fulham (Shah, J S., & Adhvaryu, 2016). The method involves with following steps: defining POI (points of interest) and SAP (service access points), identifying valid routes to reach SAP, and calculating minimum time to walk to SAP, adding average wait time to the walking time to obtain minimum access time, and obtaining the accessibility index by incorporating routes and frequency data of transportation. While the methodology provides a systematic approach to measure public transportation accessibility level, it has some deficiencies. It relies on several assumptions and parameters such as walk speed, peak-hour frequencies, and reliability factors, which may not accurately reflect the diverse conditions and dynamics of public transport in a busy city like Chicago. Other accessibility studies using network analysis also have the same problem of having presumptions of people's behavior and data. Network analysis is not simply calculating the walking buffer surrounding each point of interest. It includes various spatial analysis techniques that utilize the connectivity and topology of linear networks, such as roads, rivers, or public transportation routes. A study team from University of Leicester conducted network analysis on the accessibility of urban green spaces in Leicester, particularly focusing on how access varies across different ethnic and religious groups (Comber et al., 2008). They used a Geographic Information Science (GIS) tool called SANET to calculate distances between green space access points and output area centroids along the road network. The centroids refer to the center points of the census data area. By calculating the distances between these centroids and nearest green spaces, researchers can find how accessible each census area are to the green spaces. The analysis considers actual access routes and barriers like rivers or railways, offering a more realistic view of accessibility than simple point-to-point distance calculations. However, a network analysis model such as this one also makes assumptions about the studied population and data. It assumes that individuals from different demographic backgrounds have similar travel behaviors when they visit green spaces. However, culture and social factors can influence travel patterns, potentially leading to errors in the demographic aspect of the analysis. Additionally, the analysis assumes that the provided road networks and location data are correct. In reality, road network data is frequently updated, and any inaccuracies could influence the results and conclusions of the analysis significantly.

An alternative approach to service access analysis and network analysis is to calculate the surplus or gap of the public transportation. Areas that are lacking public transportation services are named transit deserts. A research team from Canada calculated the gap of public transportation. A public transportation gap is calculated by the difference between supply and demand index (Kaeoruean et al., 2020). The demand is calculated by comparing the number of workers who use public transportation to the size of the service area. The supply is determined by evaluating the transit service capacity per hour. After normalizing both the demand and supply indices, they were divided to identify areas in the city that lack access to public transportation. This public transportation gap study method has advantage of avoiding assumption of the patterns of human behaviors. It provides a more holistic understanding of the dynamics between transit demand and supply, offering actionable insights for improving transit service quality and accessibility. However, there are still assumptions about public infrastructure in this analysis. The study assumes that all public transportation operates at its maximum capacity during each trip. In reality, buses and subways are often not fully occupied, especially during non-peak hours.

Another team of researchers from the University of Texas at Austin came up with a different approach of calculating the supply index to avoid assumption. Junfeng took in a list of public transportation features into consideration when calculating the supply index in his transit desert study (Jiao, 2017). He combined the number of transit stops, frequency of transit service (weekday service), number of transit routes, total length of sidewalks (in miles), total length of bike routes (in miles), total length of low-speed limit roads (in miles), and intersection density to calculate the supply index of public transportation in each census block. His study considered multiple aspects of public transportation supply, providing a more comprehensive and accurate transit gap output. Integrating a methodology similar to this public transport gap study would be beneficial to my research. My study only considers public transportation for commuting methods. Therefore, I will modify the supply index calculation to focus solely on bus and subway routes. Another study done by Junfeng used another calculation formula for the public transportation supply index. In his study of measuring social vulnerability in transit deserts of United States metro areas, he and his team calculated the transit supply index by adding up the normalized value of the number of transit route stops per area of census tract, number of intersections per area, and number of points of interest per area (Jiao et al., 2021). However, due to the limitation of data sources, they did not take transit frequency into consideration. The frequency of transit service plays a crucial role in defining the accessibility and usability of public transportation. Areas with high demand for transit but low frequency may be mistakenly classified as adequately served, resulting in inaccuracies in the assessment of transit gaps. Therefore, my study will add frequency of transit into consideration.

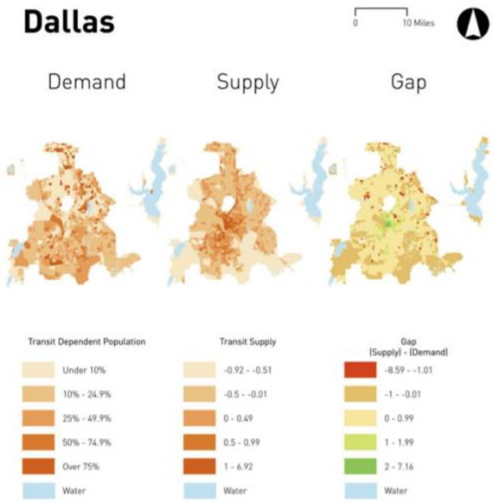


Figure 1. Transit desert analysis in Fort Worth, Texas (Jiao, 2017).

Other spatial calculation can also visualize the spatial pattern of public transportation accessibility. For example, kernel density calculation can estimate the density of point features across a continuous surface. Lukar and his research team conducted research on analyzing the association between destinations distribution and physical activity outcomes, specifically walking frequency and physical activity sufficiency (King et al., 2015). The output raster image indicates areas with greater destination intensity through dark shading, while lower destination intensity is represented by a grey color. Bus and subway stops can also be represented on the map as points and hence are suitable for kernel density analysis.

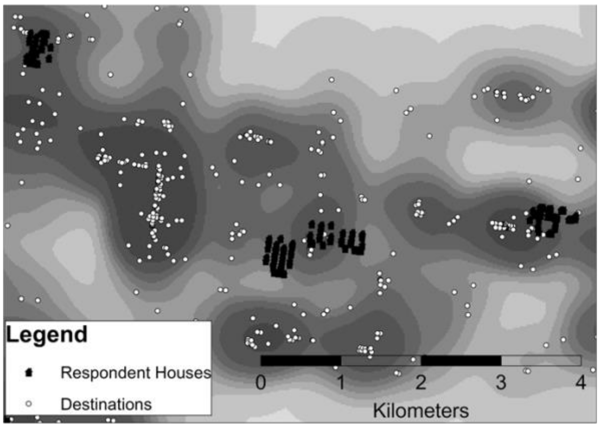


Fig 2. Kernel density map representation (King et al., 2015).

The accessibility of public transportation and the desert index can reveal spatial patterns of underserved and well-served areas in the city. My study also aims to analyze how these patterns relate to demographic distribution, including race and poverty features. The research team from University of Leicester who conducted the research on green space accessibility, incorporated race and ethnicity data in their study (Comber et al., 2008). By incorporating census tract data into their study, they produced reliable and highly detailed output at the census tract level. Another research team on green land accessibility used American Community Survey (ACS) data that contains racial information. According to their research team, in comparison with the decennial census, the ACS offers updated and high-quality data covering major socioeconomic aspects of the small areas for the United States, the demographic information of all census tracts in the study area can be extracted from the ACS (Liu et al., 2021). The study used statistical methods, such as linear regression analysis, to measure the connection between access to urban green spaces and racial categories. It calculated regression coefficients to assess how much accessibility to urban green spaces varies between census tracts with a White majority and those with a minority majority. The same calculation can be implemented into public transportation accessibility. Ermagun and his research team utilized demographic data separated by block group and compared that to the transit accessibility values. Their findings revealed that African Americans with low incomes predominantly reside in the southern part of Chicago, where access to transportation, public infrastructure, and grocery stores is relatively limited (Ermagun & Tilahun, 2020). Dajun analyzed demographic data to assess racial/ethnic and socioeconomic disparities in green space access using bivariate correlation and multivariate regression analyses (Dai, 2011). His study found significant negative bivariate correlation between green space access and multiple socioeconomic variables such as black population, households with more than one occupant per room, female-headed households, populations below the poverty line, and households without cars. This suggests that neighborhoods with higher percentages of these demographics tend to have less access to green spaces.

Visualization plays a crucial role in GIS research, offering insights into spatial patterns and relationships. ArcGIS Story Maps emerge as a powerful tool for visualizing map outputs, providing an intuitive platform to present geographic data in a narrative format. With its interactive maps, multimedia content, and storytelling capabilities, ArcGIS Story Maps enable researchers to effectively communicate their findings, and convey complex geo-spatial concepts in clarity. Story maps have greater potential for educational purposes compared to traditional image map outputs and graphs. Cope and his team discovered that students respond more positively when introduced to geographical content through story maps rather than traditional textbooks (Cope et al., 2018). Edwards and his research team also states that online, interactive GIS Story Maps are the best tool for encouraging citizen engagement, providing meaningful context to geo-spatial topics and concepts, and empowering informed decision making (Nguyen et al., 2016). The objective of this study is to inform policy makers and residents about the public transportation accessibility distribution and its relation to demographic features. The target audiences might not have a professional background in GIS. An online interactive web map application is suitable for visualizing the results, making it easier to convey the information hidden behind the data and graphs.

Data acquisition and processing:

The transportation data utilized in this study is derived from General Transit Feed Specification (GTFS) files provided by Chicago Transit Authority. GTFS data is the common data source for representing public transportation schedules and stops including CTA buses, subways, and trains. The GTFS data consists of five major text documents. The stops.txt file contains information about transit stop locations, including stop IDs, names, and geographic coordinates. The routes.txt file provides detail information about the ID and service type of the transit routes. The trips.txt file contains specific trips made by each transit units. The stop_times.txt file lists the arrival and departure time of each vehicle. The frequency.txt file contains details about the frequency of each trip. Before analyzing the GTFS data, it needs to be processed using GIS toolsets. ArcGIS Pro's GTFS transit feed tool can convert features from text files into stop and route geometry. The tool can also incorporate frequency and schedule information into the layer based on the GTFS data.

This study retrieves demographic data from the American Community Survey (ACS) from the year 2018 to 2022. ACS is a survey conducted by the United States Census Bureau that collects and provides detailed demographic, social, economic, and housing data. The ACS employs a rolling sample design, with approximately 3.5 million addresses surveyed each year, ensuring that data are continuously updated and reflective of changing population trends. This study examines three features of the ACS dataset: means of transportation to work, population below poverty rate, and race composition. The transportation table lists the different methods used by workers aged 16 and over to travel to work, including driving, walking, taking a taxi, working from home, and using public transportation. The poverty rate table lists the total population in each census tract and the number of people who are below poverty rate. The race and ethnicity table contains the number of white, black, Asian, and other population in each location point. I downloaded the TIGER shapefile in census tract level to turn these demographic attributes into geospatial layer. TIGER shapefile (Topologically Integrated Geographic Encoding and Referencing) is a digital vector format developed by the United States Census Bureau to represent geographic features and boundaries within the United States. I then joined the demographic data into the TIGER shapefile and clipped the spatial extent to Chicago boundary.

Method:

The initial step in the analysis involved conducting a kernel density analysis for transit stops. Kernel density calculation uses a kernel function to calculate the magnitude-per-unit area from point or polyline features. A univariate dataset at a specific point x is calculated using the formula:

$$KDE(x,y) = n \cdot hx \cdot hy1 \cdot \sum K(hxx - xi) \cdot K(hyy - yi)$$

Where

KDE(x) = Kernel density estimates at a given location (x) on the spatial surface

N = Total number of public transportation points

h_x and h_y = Bandwidth parameters that control the smoothness of the KDE

(x_i, y_i) = Coordinates of each public transportation point in the dataset

The visualization tools in ArcGIS pro can plot the kernel density values in a raster map where deeper color represents higher density of stop points.

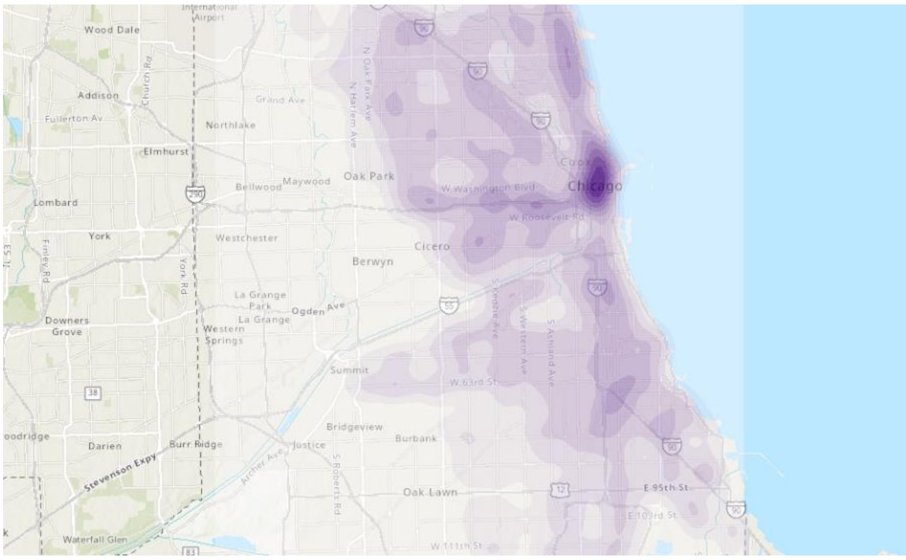


Fig 3. Kernel density map

The transit surplus index is measured by the difference between public transportation demand index and supply index in each census tract using the following formula:

Surplus Index = Normalized Supply Index – Normalized Demand Index

When the surplus index is negative, it indicates a lack of public transportation access in the area. Conversely, a positive surplus index value for a census tract area indicates adequate public transportation access. The demand index in each census tract is calculated based on the ways of commute to work data layer. The equation can be written as:

Demand Index = number of workers taking public transportation to work / Total workers

Public transportation supply index is calculated based on the combination of normalized transit values including transit stops, transit route counts, transit route length, and transit frequency. Transit stop counts are obtained using the count points in polygon tool in QGIS. The number of transit routes is calculated by summing the intersections between census tract boundaries and transit routes. The total length of transit routes is determined using the sum line lengths tool in QGIS. The frequency data is derived from GTFS text file. To ensure a balanced representation of the components contribution to the supply index, weights (w) are assigned to each variable in the supply index calculation. These weights are determined based on the relative significant of each factor in influencing transit accessibility. The transit supply formula can be written as below:

$$\text{Supply Index} = (w_{stops} * S_{stops} + w_{routes} * S_{routes} + w_{length} * S_{length}) * x_{frequency\_normalized}$$

$$S_{stops} = (X_{stops} - X_{min\_stops}) / (X_{max\_stops} - X_{min\_stops})$$

$$S_{routes} = (X_{routes} - X_{min\_routes}) / (X_{max\_routes} - X_{min\_routes})$$

$$S_{length} = (X_{length} - X_{min\_length}) / (X_{max\_length} - X_{min\_length})$$

where

x_{stops} = Number of stops in each census tract

x_{routes} = Number of routes in each census tract

x_{length} = Sum of the length of routes in each census tract

$x_{frequency}$ = The frequency of buses and trains

w = weight assigned to each variable

The demographic layer includes the poverty rate (poverty population divided by total population) and race population rate (race population divided by total population). In this study, regression coefficients will be estimated using Ordinary Least Squares.

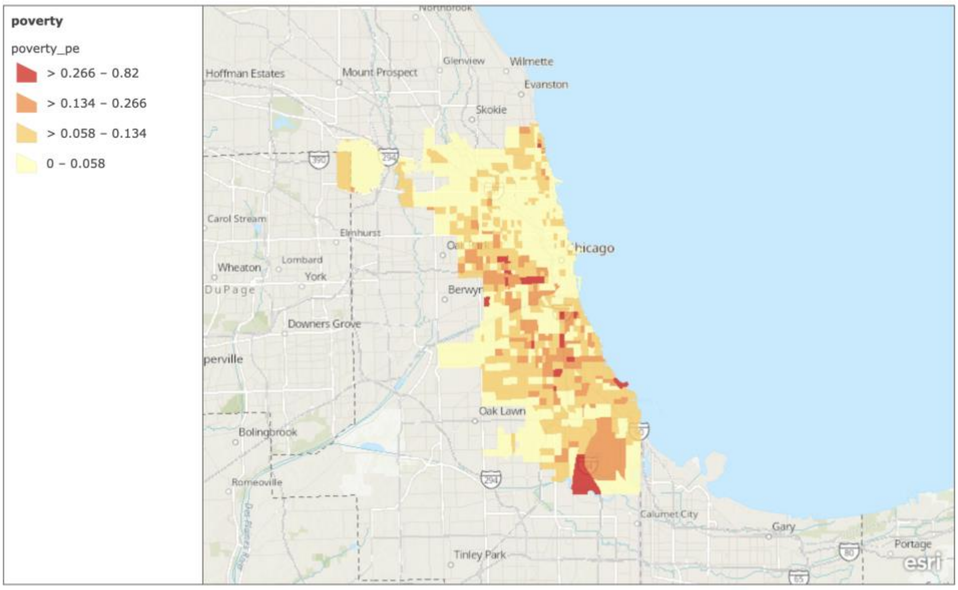
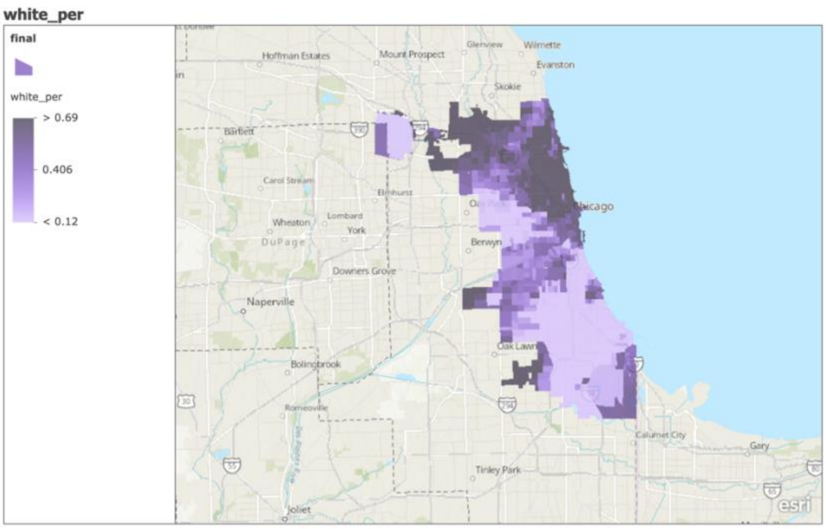
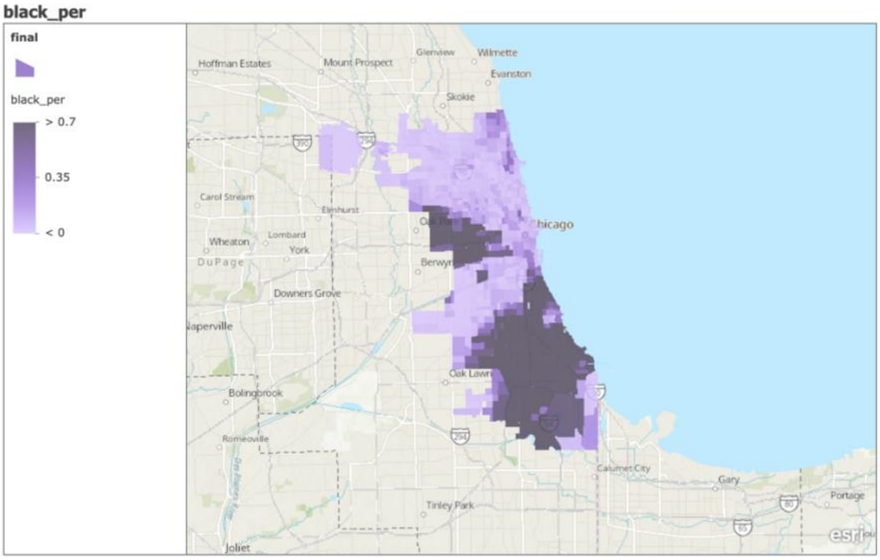


Fig 4. Poverty rate map





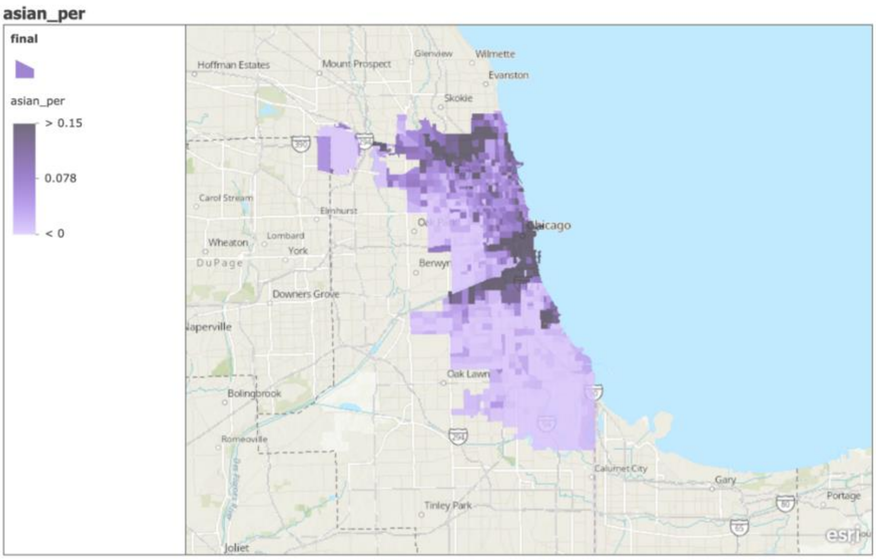


Fig 5, Race population rate map

Result:

The kernel density map shows the distribution of public transit stops in Chicago where the downtown Chicago has a significantly higher number of bus and train stops. However, when analyzing accessibility, transit stop numbers are not the only attribute to consider. The public transit surplus index map indicates that there is a lack of adequate public transportation access in most areas of northeastern Chicago. The high demand for public transportation is coupled with a low supply of available public transport options in these areas. The demographic correlation assumption is that the surplus of public transportation has an influence on the poverty rate. In this case, I treated the poverty rate as the dependent variable and the surplus as the independent variable in the regression model. resulted in an R^2(formula) value of 0.0036625, a standard error of 0.007, and a p-value < 0.0001. The R^2 value suggests that only 0.4% of the variation in poverty rate can be explained by variations in public transportation surplus. This indicates a very weak relationship between the two variables. The low standard error of 0.003 suggests that the coefficient estimate for surplus is relatively precise. A p-value = 0.00745715 indicates that more evidence would be needed to conclude that there is a relationship between the two variables. For race distribution data, I set the public transit surplus value as the dependent variable and the race population rate data as the independent variable. I wanted to evaluate how the racial composition influenced the urban planning process. The resulted R^2 value = 0.0029174 for Asian population rate, 0.0105682 for White population rate, and 0.0026302 for black population rate. The results show that the racial composition has little correlation with the distribution of public transit surplus. All the maps and graphs in this study are organized using ArcGIS story map, which offers clear and interactive visualizations related to the topic. Story map link: https://storymaps.arcgis.com/stories/64f01227558741d28680489cd6e228f5.

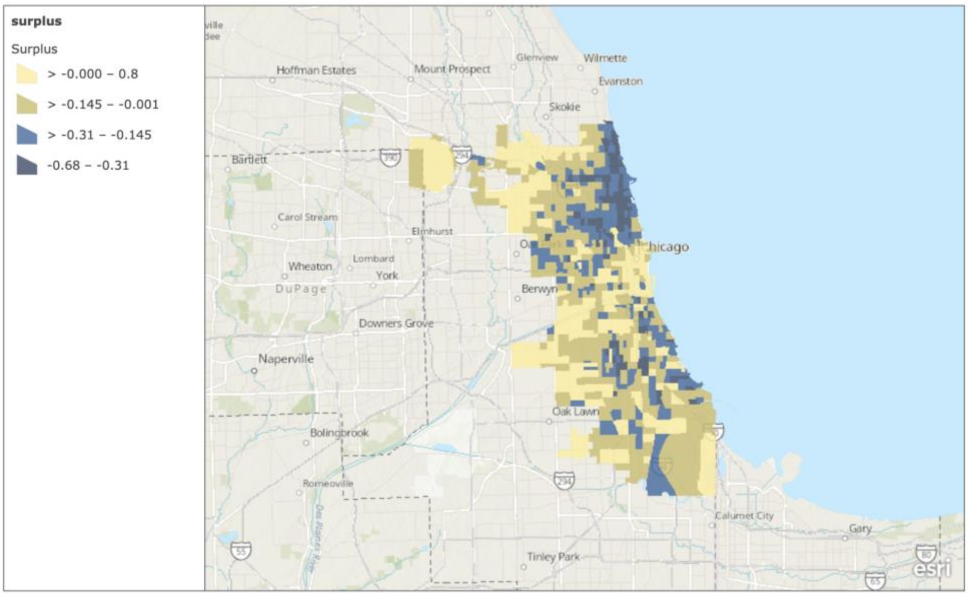


Fig 6, Transit surplus map

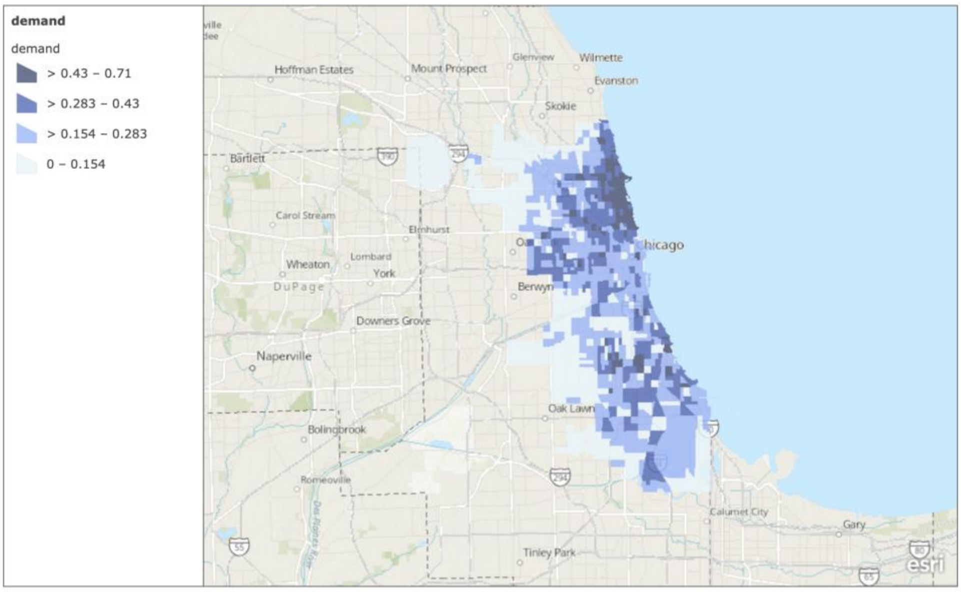


Fig 7, Transit demand map

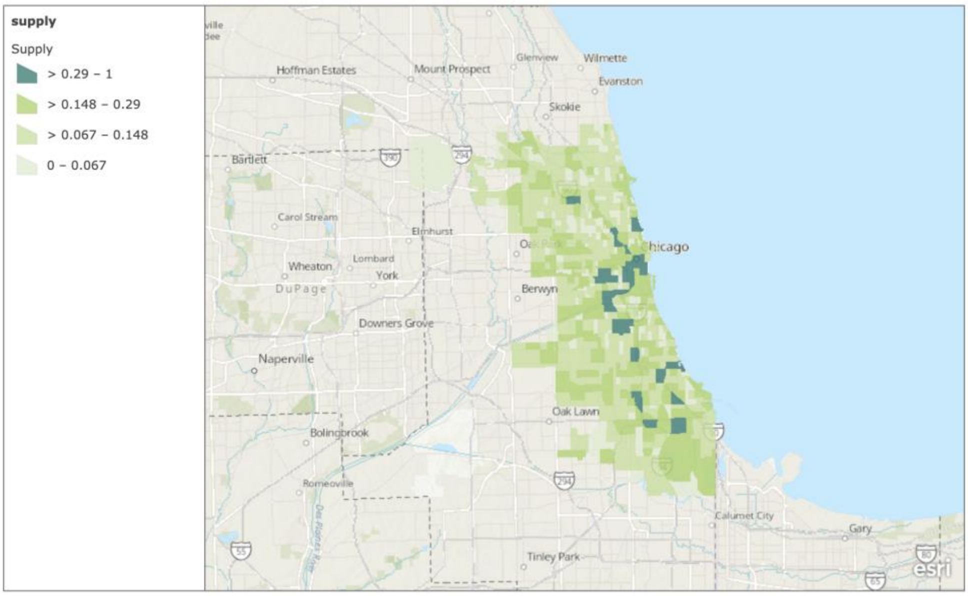


Fig 8, Transit supply map

Discussion:

In this study I utilized GTFS data and demographic data to calculate the public transportation surplus in Chicago area. This surplus calculation method provides a comprehensive assessment of the balance between the accessibility of public transportation services and the level of demand within a specific area. This allows for a more nuanced understanding of transportation system performance. It also allows for customization by adjusting the weights assigned to different supply factors (such as stops, routes, and route length) based on their relative importance. Compared to network analysis approach, the surplus index can be easily interpreted where positive surplus values indicate areas where public transportation services exceed demand and negative surplus values signal areas of unmet demand. However, the surplus formula neglects some influencing factors such as transit reliability and access to the transit station. The demographic analysis revealed minimal correlation between the transit surplus and the two demographic variables. More demographic variables such as home ownership and age can be added into the formula to look for correlations. The web map application that combines all the data output can help urban planners identify areas with transportation surplus or deficit, guiding decisions on infrastructure investment, route planning, and service optimization to enhance efficiency of the public transportation. Investigating the public transportation gap in the Northern part of Chicago would be a good starting point.

References:

CTA Facts at a Glance. (2023, January 1). https://www.transitchicago.com/facts/

Ermagun, A., & Tilahun, N. (2020, September 1). Equity of transit accessibility across Chicago. https://doi.org/10.1016/j.trd.2020.102461

Lowe, K W., & Jones, P C. (2023, August 4). Participation and Equity in Municipal Transportation Decisions: A Chicago Case Study and a Politicized Capacity-Building Alternative. SAGE Publishing. https://doi.org/https://doi.org/10.1177/1087724x231193721

Shah, J S., & Adhvaryu, B. (2016, September 1). Public Transport Accessibility Levels for Ahmedabad, India. , 19(3), 19-35. https://doi.org/https://doi.org/10.5038/2375-0901.19.3.2

Comber, A., Brunsdon, C., & Green, E. (2008, May 1). Using a GIS-based network analysis to determine urban greenspace accessibility for different ethnic and religious groups. Elsevier BV, 86(1), 103-114. https://doi.org/https://doi.org/10.1016/j.landurbplan.2008.01.002

Kaeoruean, K., Phithakkitnukoon, S., Demissie, M G., Kattan, L., & Ratti, C. (2020, September 3). Analysis of demand–supply gaps in public transit systems based on census and

GTFS data: a case study of Calgary, Canada. Springer Science+Business Media, 12(3), 483-516. https://doi.org/https://doi.org/10.1007/s12469-020-00252-y

Jiao, J. (2017, January 12). Identifying transit deserts in major Texas cities where the supplies missed the demands. University of Minnesota, 10(1). https://doi.org/https://doi.org/10.5198/jtlu.2017.899

Jiao, J., Conrad, J., & Azimian, A. (2021, April 15). Measuring Social Vulnerability in Transit Deserts of United States Metro Areas. , 8(1). https://doi.org/https://doi.org/10.26153/tsw/12527

King, T., Thornton, L., Bentley, R., & Kavanagh, A. (2015, September 10). The Use of Kernel Density Estimation to Examine Associations between Neighborhood Destination Intensity and Walking and Physical Activity. Public Library of Science, 10(9), e0137402-e0137402. https://doi.org/https://doi.org/10.1371/journal.pone.0137402

Liu, D., Kwan, M., & Kan, Z. (2021, April 1). Analysis of urban green space accessibility and distribution inequity in the City of Chicago. Elsevier BV, 59, 127029-127029. https://doi.org/https://doi.org/10.1016/j.ufug.2021.127029

Dai, D. (2011, September 1). Racial/ethnic and socioeconomic disparities in urban green space accessibility: Where to intervene?. Elsevier BV, 102(4), 234-244. https://doi.org/https://doi.org/10.1016/j.landurbplan.2011.05.002

Cope, M., Mikhailova, E A., Post, C J., Schlautman, M A., & Carbajales-Dale, P. (2018, July 19). Developing and Evaluating an ESRI Story Map as an Educational Tool. Wiley, 47(1), 1-9. https://doi.org/https://doi.org/10.4195/nse2018.04.0008

Nguyen, T A., Edwards, S., Rahall, N J., Scott, M., & Cragle, J. (2016, October 1). GIS

Story Maps: A Tool to Empower and Engage Stakeholders in Planning Sustainable Places.

https://rosap.ntl.bts.gov/view/dot/34788/dot_34788_DS1.pdf