EECS 298: Consequences of Computing
April 20, 2025

Final Project Report
By: Varun Sadasivam, Bella Palumbi, Margaret Wozniak, and Shalini Krish

## Introduction

In New York City, over 55% of workers rely on public transit for their daily commutes [1]. This dependence highlights the critical role of an efficient and equitable transit system in meeting everyday needs, such as traveling to work, attending medical appointments, or navigating the city. However, when public transit access is inequitable, it creates barriers that disproportionately impact individuals based on their socioeconomic status. Addressing these disparities is crucial to ensuring that public transportation is accessible to all residents, regardless of their financial situation.

In response to this issue, we developed a framework to evaluate how effectively existing transit systems facilitate access to specific destinations. In this report, we utilize our tool to explore the question: Are there inequities in access to schools via public transit in Queens, New York City? Our findings are visualized through a heatmap that displays the correlation between median income and transit access across zip codes in Queens. The methodology we established is versatile and can be applied to various regions and types of services, allowing for a broader examination of transit equity beyond this initial study.

This report presents a comprehensive overview of our data preparation and heatmap generation process, alongside an analysis of our findings and a discussion of the strengths and limitations of qualitative analysis. We also propose recommendations for integrating this tool into the democratic process by publicizing its insights to facilitate public discourse and feedback. Emphasizing modularity and flexibility, our tool can accommodate various data sources, making it adaptable for use in discussions about legislative changes, proposed transit maps, or budget allocations. By offering easily interpretable outputs, this tool could empower communities to participate in shaping transit policies. It holds potential for application in analyzing transit systems nationwide, providing a valuable resource for advancing equitable transportation access.

## Background

Looking at the existing data and analyses regarding transit access in New York, it's evident that access to public transportation is heavily influenced by income level, creating significant disparities. After cross referencing the National Neighborhood Data Archive (NaNDA): Public Transit Stops by Census Tract and ZIP Code Tabulation Area data set [12] and census data from the 2020 census, we identified that in New York City, individuals earning less than $50,000 per year have, on average, only 0.25 public transit stops per capita, while those earning between $203,000 and $225,000 benefit from about 2.9 stops per capita. This inequitable distribution of transit resources results in longer commutes, economic disadvantages, and environmental burdens for lower-income individuals.

For residents in low-income neighborhoods, unreliable bus service often necessitates more expensive commuting alternatives like rideshare services or taxis, exacerbating financial strain on these households. Moreover, the environmental impacts of inadequate public transit, such as increased air pollution from traffic congestion and outdated bus fleets, disproportionately affect poorer communities, leading to adverse public health outcomes. Addressing these issues requires a multifaceted approach that combines technological innovation, policy reform, and community engagement to develop a fairer, more efficient, and universally accessible bus network for all New Yorkers.

The stakeholders in this system include residents of lower-income neighborhoods, the Metropolitan Transportation Authority (MTA), city government and policymakers, bus operators and transit workers, community organizations, and urban planners. Residents are the primary beneficiaries, as improved bus service would enhance their mobility and economic opportunities. The MTA and city government play crucial roles in implementing policies, allocating budgets, and making infrastructure investments. Bus operators and transit workers would be directly affected by changes in routes and service frequency, while community organizations serve as advocates to ensure that transit improvements align with the needs of the people.

New York City's bus system has been characterized as being in crisis, with documented inequities affecting marginalized communities. A 2017 report by NYC Comptroller Scott Stringer revealed that NYC buses are the slowest among major U.S. cities, averaging only 5.5 mph in Manhattan [2]. This significant underperformance has led to the loss of approximately 100 million individual rides over eight years, with service deterioration disproportionately impacting low-income and minority communities [2, 4].

Research consistently shows that transit inequity in NYC has deep socioeconomic dimensions. Bus commuters are predominantly people of color (75%) who tend to be the most economically vulnerable residents with lower incomes and less education than subway commuters [6]. During the COVID-19 pandemic, lower-income neighborhoods in the Bronx and Brooklyn maintained higher bus ridership levels than wealthier areas, highlighting their greater dependency on public transit [7]. Additionally, gentrification has exacerbated these problems, with bus ridership declining in Manhattan and gentrified Brooklyn neighborhoods while simultaneously displacing lower-income communities to areas with poorer transit access [3, 6].

Several initiatives have attempted to address these issues, including NYC Transit's 2018 Bus Action Plan and the 2019 Streets Master Plan law mandating 150 miles of new bus lanes [8]. The 14th Street busway project demonstrated how dedicating street space primarily to buses could significantly improve travel times and reliability [9]. However, implementation of such improvements has been inconsistent across the city, with the Adams administration adding just 9.6 miles of new bus lanes in 2022-2023, making it virtually impossible to meet mandated targets [5]. Despite technical solutions being well-documented (including transit signal priority, dedicated bus lanes, all-door boarding, and route optimization), political factors and coordination challenges between agencies have limited progress toward creating a more equitable bus transit system in New York City [2, 8].

## Methodology

The methodology of our project was made up of three phases: data preparation, algorithmic processing, and the construction of the heatmap. We needed to be able to pass in a set of zip codes and generate ‘transit density scores’ per zip code according to a specified formula, then output a heatmap to visualize the data.

Our goal with the data pre-processing was to align the school and median income datasets with the zip code data. We had two streams of incoming data: median income per zip code, and a list of all the bus stop locations in Queens. Each had to be prepared separately, ensuring that zip code was consistent across all our datasets. Zip Code Tabulation Areas (ZCTAs) are statistical geographic entities that are approximations of U.S. Postal Service (USPS) zip code service areas. Unlike traditional zip codes, which are defined for postal delivery purposes, ZCTAs provide a way to represent these areas in demographic and economic data collection. They unite census data with geographic locations, making them especially useful for analyses that combine socioeconomic information with spatial data, such as our study on transit access in New York City. ZCTAs were crucial in our analysis, as they allowed us to match median income data with bus stop locations consistently, facilitating a more precise evaluation of transit density and socioeconomic disparities.

The main external tool we used in this work was the Python library GeoPandas, an extension of the popular data analysis library Pandas designed specifically for working with geographic data. Since we already had experience with Pandas through the labs for 298, it was relatively simple to adjust to using GeoPandas. The library enables effective handling of spatial data, such as points, lines, and polygons, allowing for operations like mapping and spatial analysis. While Pandas deals primarily with tabular data, GeoPandas integrates geometrical data, making it easier to perform tasks like plotting maps and spatially joining datasets based on location. Crucially, GeoPandas can handle file types specific to geographical information, such as shapefiles, and perform spatial operations like buffering and intersection, which are not possible with standard Pandas functions. These capabilities make the library especially useful in fields like urban planning, environmental science, and any area where geographical visualization and analysis are required.

*School Data Pre-Processing*

*Script:* school_data_processing.py
*Data Sources:*

- New York City Department of Education. (November 6, 2024). *School Point Locations* [Data set]. New York Open Data. https://data.cityofnewyork.us/Education/School-Point-Locations/jfju-ynrr/about_data

- U.S. Census Bureau. (2020). TIGER/Line Shapefile, 2020. *2010 5-Digit ZIP Code Tabulation Areas (ZCTA5)* [Data set]. U.S. Department of Commerce. https://catalog.data.gov/dataset/tiger-line-shapefile-2020-nation-u-s-2010-5-digit-zip-code-tabulation-areas-zcta5

The program school_data_processing.py extracts the name, latitude, longitude, and ZCTA for all schools in New York City and stores them in csv format. The program starts by using GeoPandas to read in the the census Shapefile (tl_2020_us_zcta510.shp) and a geopandas read on the school .shp file (SchoolPoints_APS_2024_08_28.shp) in order to create two GeoDataFrame objects containing the information from both files.

The program goes on to standardize the two dataframes to be in the same coordinate reference system using the geopandas `GeoDataFrame.to_crs()` function. It then performs a spatial join with `GeoDataFrame.sjoin()`. In this way, we can map the school locations onto the zip codes present in our other data sets. A ZCTA is assigned to each school based on spatial overlap with or nearest proximity to the ZCTAs in the census data. The program then drops the columns not needed from the join, and writes the output (name, lat, long, ZCTA) to a csv.

A note on data file types: Shapefiles (.shp) are a format for geographic information system data, developed by the company Esri. A shapefile is not actually a single file but a collection of files, including .shp, .shx, and .dbf.. These files work together to store a variety of spatial data and enable complex spatial analysis and visualization. Our Github repository contains all the files except the tl_2020_us_zcta510.shp itself because the file size was too large to be uploaded.

### *Bus Stop Data Pre-Processing*

*Script:* routes_processing.py
*Data Sources:*

- U.S. Census Bureau. (2020). *TIGER/Line Shapefile, 2020, Nation, U.S., 2010 5-Digit ZIP Code Tabulation Areas (ZCTA5)* [Data set]. U.S. Department of Commerce. https://catalog.data.gov/dataset/tiger-line-shapefile-2020-nation-u-s-2010-5-digit-zip-code-tabulation-areas-zcta5
- Metropolitan Transportation Authority. (2024). *MTA Bus Timepoints* [Data set]. New York Open Data. https://www.mta.info/open-data

We performed a similar data pre-processing procedure on the bus stop dataset, mapping each location to a ZCTA present in the census data. One notable difference was that the bus stop data was originally in GeoJSON format as opposed to a .shp file. Because we were using a flexible spatial data processing library in GeoPandas, we were able to read this file format into a GeoDataFrame in almost exactly the same way we did for the .shp data. This does add an extra step where after standardizing to the same coordinate reference system, the latitude and longitude columns are swapped. GeoJSON stores the data in the form as a longitude, latitude pair, whereas the US Census uses latitude, longitude. We were able to remedy this by swapping the columns in the bus stop dataframe. Then like the previous step, we saved latitude, longitude, route_id, and ZCTA to a .csv file.

*Generating Transit Density Scores*

To determine the transit density for the density analysis, the number of unique transit routes within a specified radius was counted for each location. For a given school, we found how many transit stops fell within a 0.5 mile radius. From this count, a transit density score was calculated, reflecting the abundance of transit options

*Feature Normalization*

Feature normalization was applied to ensure comparability between median income and transit density scores. We utilize the formula to convert the range of both income and transit density scores into a standardized scale between 0 and 1.

$$X' = \frac{X - Xmin}{Xmax - Xmin}$$

This normalization allowed for an equitable assessment of both metrics, facilitating the comparison of their impacts across different regions.

*Heatmap Construction*

To generate the heatmap visualization, we used both GeoJSON files and the MatPlotLib Python library to map the borough of Queens. The normalized scores for income and transit density were key inputs, as they highlighted areas with varying levels of access and income. The aim was to identify any areas that stood out due to significant differences in income or transit density, thereby flagging potential inequities between zip codes.The heatmap illustrates these normalized scores, categorically displaying regions with high and low accessibility and income levels.

*Implications and Considerations*

It's crucial to critically assess our chosen outcome variable - the correlation between median income and bus stop density around schools, segmented by zip code - to understand its broader significance and potential limitations. We will briefly analyze our outcome variable through the lens of the framework proposed in Martin 1.6, which poses questions such as “What is the phenomenon we are trying to represent with this outcome?” [13] Our choice of variable aims to reflect public transportation accessibility and equity in educational settings. However, several assumptions underpin this decision. We presume that median income reliably indicates socioeconomic status, despite its potential to overlook individual household variances.

Additionally, while bus stop density suggests availability, it does not account for service quality or frequency, crucial factors for true accessibility. This outcome primarily serves urban planners and policymakers, though it might not fully represent families or districts with unique transportation needs. Addressing these assumptions is essential for a fair and nuanced understanding of the data.

## Results

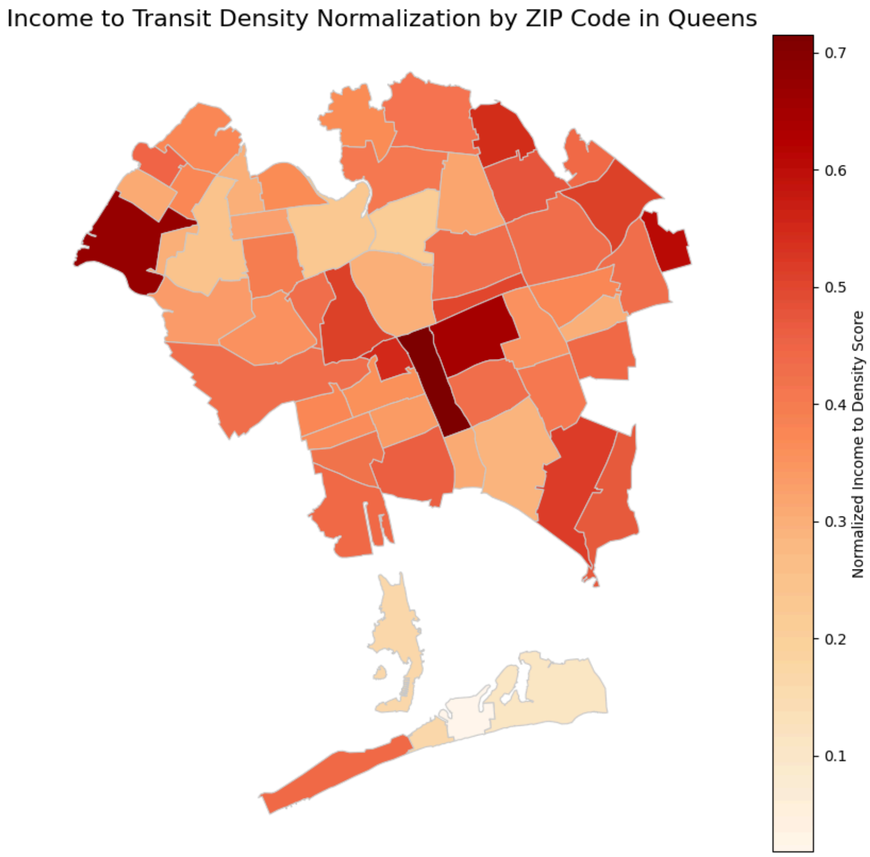
Income to Transit Density Normalization by ZIP Code in Queens
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Normalized Income to Density Score

Figure 1: Heatmap showing Income Correlation with Transit Density Score in Queens, New York

Our original example question was: Are there inequities in access to schools via public transit across zip codes in Queens? To interpret the heatmap, take the intensity of color in each zip code as a measure of how strongly the two data points are correlated. If a borough is wealthy and also has a relatively high transit density score, it will be a darker orange. So will a borough with a low median income and low transit density score. On the other hand, if a borough displays little to no correlation between its median income and transit density, it will be a lighter color.

The heatmap produced by our framework indicates that median income and transit density around schools are moderately correlated across Queens. Based on our analysis, we conclude that while Queens as a whole shows a moderate level of correlation between income and transit density, significant variations exist at the zip code level, highlighting potential inequities. High correlation areas illustrate potential systemic bias where wealth dictates access to resources, whereas inconsistencies in other areas may point to gaps in public infrastructure or allocation. The heatmap provides a qualitative visualization of these trends, and quantitatively, the moderate correlation suggests a systemic issue that merits attention from urban planners and policymakers.

## Recommendations

To address these identified inequities in public transit access for school children, strategic interventions should be considered. Firstly, urban planning should prioritize bolstering transit options in low-income areas currently lacking adequate services. Enhancing school bus routes or developing new transit corridors in these zones could mitigate accessibility issues. Additionally, policies could focus on ensuring equitable resource distribution regardless of income levels, possibly through subsidies or partnerships aimed at improving transit infrastructure.

From a conceptual standpoint, this study underscores the importance of integrating socioeconomic factors into urban planning. Future work could apply our framework more broadly, adapting the methodology to other districts or even different infrastructural elements like healthcare or community centers. For tangible outcomes, a collaborative effort involving local government, public transport authorities, and community stakeholders would be needed to implement these recommendations effectively and foster more equitable access to education across Queens

## Discussion and Future Work

If we were to expand the functionality of this tool, we would add options for drilling into individual zip codes. Instead of producing a static heatmap, the output would be an interactive map with clickable regions. Upon selection of a region, the user would be able to look at the underlying data and see the data points that led to the correlation score. Ideally, a user could select an individual school (or whatever resource datapoint you are representing) and then see a visualization of which transit stops are in its proximity radius.

Another benefit of this kind of data visualization is that it condenses a lot of information into something that is more easily digestible to a wide audience. We think the output of this sort of tool (with additional features) could be a good way to present information to the public for comment and feedback. Publicizing these kinds of reports regularly could also be a good way to evaluate the existing public transit networks. The public and the body of taxpayers make up a large proportion of the stakeholders in this sociotechnical system. Visualizations like these can provide a great opportunity for decision makers (e.g. legislators, administrators, transit system designers) to communicate with their audience and incorporate feedback.

As we demonstrated, one such map in isolation can tell you a lot about a certain question, but it would be interesting to look into ways to layer the heatmaps across different questions. For instance, could you use an extended version of this process to identify zip codes that have poor access to schools, hospitals, parks, and supermarkets all at the same time? At this point, you do start to reach the bounds of the amount of information that can be displayed at one time. It's crucial to balance legibility and information level of detail, especially when developing materials for a broad audience.

There were some limitations and roadblocks we encountered during the production of this report. We had originally intended to implement a discrete transit density score system instead of focusing just on the count of transit stops within the radius. This type of change would allow for more customizability in the scoring. For instance, a school with only 1 transit stop in its radius could receive only 25% of the number points a school with 2 transit stops might receive. This would be a scaled system, and we could experiment with either discrete demarcations or basing the score on a continuous function. In fact, we could make the measure even more granular by calculating using the exact distance measurements instead of a binary yes/no for whether the stop falls within the radius.

## Conclusion

Our investigation into transit equity in Queens highlights the correlation between socioeconomic status and public transit access, using a framework that visualizes disparities through a heatmap. We identified a  moderate correlation between median income and transit density around schools, suggesting systemic biases that call for targeted interventions. The findings underscore the necessity of including socioeconomic considerations in urban planning and improving transit services in underserved areas. Our adaptable framework holds promise for broader application in evaluating infrastructure nationwide. In the timeframe for this project, we made a prototype of such a tool, and future work would involve increasing the interactivity and level of detail. We recommend that such a data visualization project utilizing tools akin to the one presented here be included as part of dialogues on social justice and transit equity. As a final note, we further emphasize that identifying harm is merely the first step toward ameliorating the systemic biases in our society, and must be followed up with dedicated efforts to implement equitable solutions.

# References

[1] U.S. Census Bureau. (2021). *2020 American Community Survey 1-year estimates: Comparison profiles*. U.S. Department of Commerce. https://www.census.gov/content/dam/Census/library/publications/2021/acs/acs-48.pdf

[2] New York City Comptroller's Office. (2017, November 27). *The other transit crisis: How to improve the NYC bus system*. New York City Comptroller. https://comptroller.nyc.gov/reports/the-other-transit-crisis-how-to-improve-the-nyc-bus-system

[3] Sledge, M. (2021, June 8). *Racism has shaped public transit, and it's riddled with inequities*. Kinder Institute for Urban Research. https://kinder.rice.edu/urbanedge/racism-has-shaped-public-transit-and-its-riddled-inequities

[4] Crain's New York Business. (2018, October 18). *Slow city buses get failing grade*. https://www.crainsnewyork.com/transportation/slow-city-buses-get-failing-grade

[5] Meyer, David (2025, February 7). *Report: Efforts to speed up bus speeds have stalled — like bus speeds.* Streetsblog NYC. https://nyc.streetsblog.org/2025/02/07/report-efforts-to-speed-up-bus-speeds-have-stalled-like-bus-speeds

[6] New York League of Conservation Voters. (2018, July 23). *Comptroller's report highlights NYC bus problems.* New York League of Conservation Voters. https://nylcv.org/news/comptrollers-report-highlights-nyc-bus-problems/

[7] Halvorsen, A., He, Q., Ratner, K., & Li, H. (2023). Examination of New York City Transit's bus and subway ridership trends during the COVID-19 pandemic. *Transportation Research Record, 2677*(4), 51-64. https://doi.org/10.1177/03611981211028860

[8] Bus Turnaround Coalition: https://busturnaround.nyc/

[9] Cortright, J. (2020, November 8). *Car-free, bus-only street in NYC: A case study in transportation equity*. Kinder Institute for Urban Research. https://kinder.rice.edu/urbanedge/car-free-bus-only-street-nyc-case-study-transportation-equity

[10] Eizaguirre, S., Pradel, M., Terrones, A., Martinez-Celorrio, X., & García, M. (2012). Multilevel governance and social cohesion: Bringing back conflict in citizenship practices. Urban Studies, 49(9), 1999-2016.

[11] Urry, J. (2007). Mobilities. Polity Press.

[12] Pan et al.. National Neighborhood Data Archive (NaNDA): Public Transit Stops by Census Tract and ZIP Code Tabulation Area, United States, 2016-2018 and 2024. Inter-university Consortium for Political and Social Research [distributor], 2024-12-11. https://doi.org/10.3886/ICPSR38605.v2

[13] Martin, K. (2022). Algorithmic bias and corporate responsibility: How companies hide behind the false veil of the technological imperative. In K. Martin, Ethics of data and analytics: Concepts and cases (Chapter 1.6). CRC Press.