Transit Analysis in Charlotte, NC

Data and code files can be found at:

https://github.com/JBrighttt/plan372_hmks/tree/main/Final%20Project

Introduction

Public transit in Charlotte has long been criticized for the lackluster service and availability offered throughout the metropolitan area. Much of the residential area within the city's borders is suburban, and the prospect of walking or biking from home to work or any other place is exceedingly difficult anywhere outside of a few small neighborhoods close to the urban core of the city. Recently dubbed as the "sprawl capital of planet Earth" by urbanist Youtuber Ray Delahanty (Newsom), Charlotte will need much more than a few bike lanes painted on asphalt to alleviate the issues of car dependency and pedestrian connectivity. Providing public transit service that is realistic and feasible for citizens to use on a daily basis will be a necessary part of solving these issues. For walking and biking to be a realistic means of travel in a city, interconnected transit networks with service intervals frequent enough to be convenient are invaluable as infrastructure and as a public service. Public transit is additionally important to mitigate environmental impacts as the increasingly car-dependent development patterns in American cities exacerbates greenhouse gas emissions, of which transportation accounts for 29% in the United States (Hosseini). For these reasons, it is important to assess the current accessibility of public transit in the city, and identify areas well-suited for expanded transit service.

In this study, I aim to explore potential disparities in Charlotte related to transit service frequency and various demographic characteristics related to income, employment, and money spent on transportation. I will utilize bivariate choropleth maps to assess the availability of transit and which areas of the city might be most in need of expanded transit service. With the first map I will examine which areas are best served by public transit in relation to gross activity density, a measurement of jobs plus housing per acre. The second and third maps will be used in conjunction to attempt to identify which areas are most in need of expanded transit service. To accomplish this, one map will display a jobs per acre statistic and the number of low wage workers' workplaces, and the other will display transit frequency and the number of low wage workers' homes. The areas of interest in these two maps will be those with high jobs per acre and high numbers of low wage workers' workplaces, and those with low transit frequency and high numbers of low wage workers' homes. By identifying areas of high employment where many people work for low wages, and areas of low transit service where many low wage workers live, the City of Charlotte can make a more informed decision about where and how to expand transit service effectively. Using these maps, I will attempt to identify areas well-suited for transit expansion. In addition to these visual analyses, I will utilize linear regression to assess what relationship, if any, the average income of an area has had on transit frequency. I will also attempt to assess a relationship between the percentage of income spent on transportation and transit frequency, to see if areas with greater need of economical transportation options are better or worse served by public transit. Finally, I will offer suggestions for potential transit corridors well-suited for expanded public transit service, based on the findings of the spatial representations.

Literature Review

In my process of literature review I found scholarly journal entries closely related to my subject matter, several authored by PhD students at UNC Charlotte. The first from Hosseini, et al. explores the possibility of Demand Responsive Transit, through the broad acceptance and use of transit mobile apps. Adoption of an app, which would need to be well-designed and responsive to user needs, could provide real-time data and aid the Charlotte Area Transit System (CATS) in optimizing bus routes and schedules. She explains how bus riders from low-income neighborhoods are particularly affected by the inefficient bus routes: “These residents often must travel first to the Center City, where job density is high but concentrated around financial and banking institutions, before transferring to other lines that serve low-skill, low-wage areas. This system disproportionately affects low- income communities, who rely heavily on public transportation for access to employment and essential services” (Hosseini, et al. 2023). The more recent publication looks more directly at demographic disparities in the bus system. Her study finds that, “East Charlotte residents and females face limited routes and longer wait times; safety concerns are more prevalent among Black participants and those in East Charlotte” (Hosseini, et al. 2025). Many areas of East Charlotte are low-income, which certainly plays a role in exacerbating these disparities. And finally, Li, et al. look at the effect of new light rail transit (LRT) development on commuter behavior in Charlotte, Dallas, and Los Angeles. The study found that Charlotte’s LRT created the largest increase in ridership out of the three, and the authors conclude that this is due to a dynamic of diminishing returns on public transit investment. Charlotte’s LRT was the first of its kind in the area, while Los Angeles already had several networks like it (Li, et al. 2024). This signifies, I believe, that Charlotte has much room to benefit from expanded transit service.

Data Collection

Data was collected from two online sources, the SimplyAnalytics web tool and the EPA’s Smart Location Database (SLD). All data collected were distinguished at the census block group level. SimplyAnalytics allowed me to find curated data for two specific variables I hoped to use in the analysis: median household income, and household average transportation costs, both from 2022. Median household income was sourced from the American Community Survey (ACS) Five-Year Estimate. Transportation costs were compiled by SimplyAnalytics with small area estimation techniques. By using Public Use Microdata from the Consumer Expenditure Survey, and ACS Five-Year Estimates to estimate how consumer units are distributed over small areas, they can derive expenditure estimates at the block group level for all Consumer Expenditure spending categories.

From the EPA's Smart Location Database, I used several variables which all source data from one or more of the following: 2017 Census Longitudinal Employer-Household Dynamics (LEHD) Residence Area Characteristics (RAC) and Work Area Characteristics (WAC), ACS Five-Year Estimates from 2014-2018, and Urban Design 4 Health Transit Data. The 'TransitFrequency' variable gathered from the SLD was difficult to obtain metadata on. It turns out that the transit data is sourced either from 2012 or 2014, but when I began this study I thought it was more recent. This may be a key limitation in the study, since construction of the Lynx Blue Line, a light rail with mostly dedicated right-of-way, began in 2013. I had hoped to gain access to more recent transit measurements, but I was unable to source transit data for the city anywhere else. Another key point on transit frequency is the way it was measured: the value is the combined hourly frequency of service for all transit routes that stop within 0.4 km (0.25 miles) of the block group, and reflects service during weekday afternoon peak commute periods of 4pm-7pm. The following variables and their descriptions were also sourced from the SLD: 'CountLowWage_home', the number of low wage workers' home locations. 'CountLowWage_work', the number of low wage workers' work locations. Low wage workers are classified as earning $1250 per month or less. ‘JobsPerAcre’, gross employment density on unprotected land, ‘GrossActivityDensity’, gross employment density plus housing unit density on unprotected land. The low wage worker statistics come from the 2017 Census LEHD RAC and WAC, while the employment and activity density statistics come from the ACS Five-Year Estimates from 2014-2018 and are refined with other SLD variables.

Methods

My approach in this study was primarily to make use of bivariate choropleth maps to create informative spatial representations of the City of Charlotte as it pertains to transit service frequency, and various demographic characteristics which I believe to be relevant to policy discussions of how and where public transit could or should be improved. Additionally, to investigate relationships between transit frequency and income spent on transportation, I decided that linear regression would be informative to the research and provide insights into the general patterns of transit service in the city. The first variable I decided to look at was average household transportation costs, coded as ‘transport_costs’ in my final dataset, ‘data’. I theorized that by analyzing the relationship between transit frequency and the percentage of income spent on transportation, it could help inform whether the city has disparities in providing public transit to lower income individuals. Therefore, I performed linear regression on these two variables, with transit frequency being the dependent variable. I created a simple percentage value of income spent on transportation by dividing ‘transport_costs’ by ‘income’, then multiplying by 100, to have a possible range of numbers from 0 to 100. I also regressed transit frequency on income to analyze what relationship there may be. This would, in theory, be informative as to whether public transit serves high-income communities better than low-income ones. I included this regression in my initial plan based on visuals I saw during the data gathering process, in which some of the areas with high transit frequency seemed to overlap with areas of high median income. To perform this regression, I created a variable, ‘income_thousands’, to make the relationship more easily interpretable. The variable was created by dividing ‘income’ by 1000.

‘GrossActivityDensity’ was the first variable I decided on for a relationship map with transit frequency. My thought process was that density and public transit go hand-in-hand, and the SLD’s Activity Density measurement seemed to provide an interesting metric for density, including not just housing but also jobs. I then identified the low wage worker variables as a metric I wanted to focus on, since one could reasonably argue that public transit is a more essential service for low income individuals in their commute to work. I decided to make two relationship maps, and compare the two. The first was between ‘JobsPerAcre’ and ‘CountLowWage_work’. This was intended to identify areas of high employment and high concentrations of low wage jobs, which might be good prospects for expanded transit service due to their role in the economic activity of the city and the volume of low wage commuters. The second relationship map was between ‘TransitFrequency’ and ‘CountLowWage_home’. This was intended to highlight areas of *low* transit frequency, but *high* concentrations of low wage households. This would help identify areas that might also be good prospects for expanded transit service, due to the theoretical demand for public transit and the ridership it would attract if the service was frequent enough to be convenient and connected to areas with high job concentrations. By visually identifying areas meeting these criteria across the two relationship maps, suggestions can be made about where and how public transit should be expanded in effective and efficient ways throughout the city. All bivariate choropleth maps will use standard quantiles to calculate breaks.

Results

Figure 1 is a relationship map of 'GrossActivityDensity' and 'TransitFrequency'. Areas of hot pink and light pink are of note because they have high or medium level of activity density, but a low level of transit service frequency. The southwest area of the county stands out; the hot pink block group at the southern border is where Carowinds is located. Surrounding the amusement park are block groups with high and medium density, and some have low transit frequency. Much of this area is rural or suburban, and car dependent. Another area of note is the large hot pink block group in the northeast. This is the University City Research Park, and is home to corporate offices and research facilities. Much of this area is lacking in transit frequency, even with such density of employment and the areas surrounding the research park being primarily residential. The large light pink block to the west is Charlotte Douglas International Airport; there is a consistent shuttle running from the Transportation Center in downtown to the airport. Lastly, the area along I-77 north of the city has an area with several light pink and a few dark pink areas. This is Northlake Mall and the surrounding businesses, and it is bordered by several interstate interchanges.

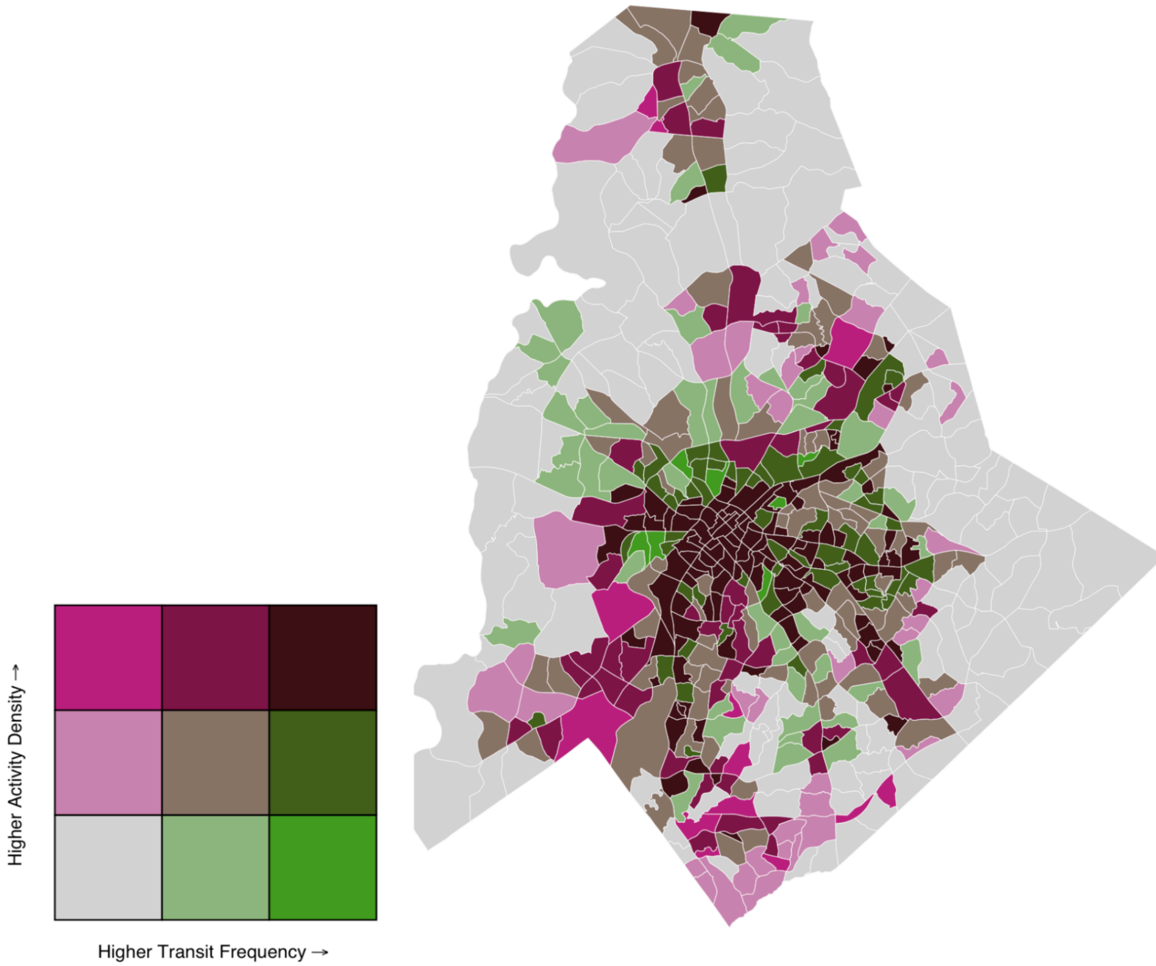

Figure 1. Activity Density and Transit Frequency

Figure 2 is the result of fit1, ‘TransitFrequency’ being regressed on ‘PctTransportCosts’. Interpreting the coefficients, the intercept is the estimated transit frequency when the percentage of income spent on transit is 0. The slope is 0.98, meaning that for every 1% increase in percentage of income spent on transit, there is a nearly 1 point increase in transit frequency, measured in combined hourly frequency of service. The results may barely pass the benchmark of .05 for statistical significance, but given the relatively weak p-values and the miniscule R-squared, a significant relationship cannot be concluded. Figure 3 is a plot of this linear model over a scatterplot of the data used. I believe that the single point outlier with over 800 combined hourly transit frequency is the block group where the Charlotte Transportation Center is located. While the trend line is positive, there appears in the scatterplot to be a curved relationship rather than a linear one.

```
##
## Call:
## lm(formula = TransitFrequency ~ PctTransportCosts, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -52.18 -31.78 -19.96  10.52 829.71
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        21.9796    11.2612   1.952   0.0519 .
## PctTransportCosts   0.9838     0.4944   1.990   0.0475 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67.51 on 301 degrees of freedom
##   (252 observations deleted due to missingness)
## Multiple R-squared:  0.01298,    Adjusted R-squared:  0.009706
## F-statistic:  3.96 on 1 and 301 DF,  p-value: 0.0475
```

Figure 2. Regression results, fit1

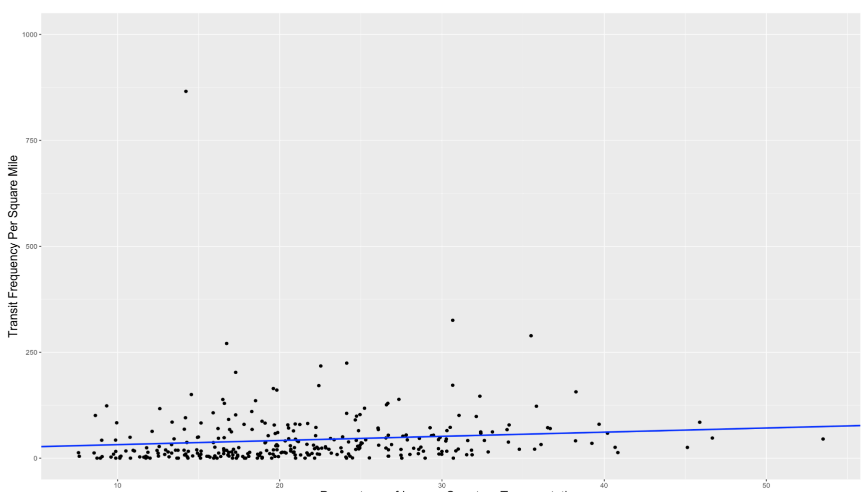

Figure 3. Plot and trend line of fit1

Figure 4 shows the results of fit2, 'TransitFrequency' regressed on 'income_thousands'. The intercept is the estimated value of transit frequency when income is 0. The slope is -0.17, meaning that for every thousand dollar increase in yearly income on average in a block group, transit frequency decreases by .17, measured in combined hourly frequency of service. This result perplexed me, and after checking the variables I discovered that the data for income is capped at 250,001. There are 555 entries with this value, and I'm certain that many, if not most, of these are much higher in reality. This does make the results of the regression less certain, but those points are still visible on the right end of the X-axis in figure 5. Many of them are quite low, so it's possible that the results would be similar, if not an even stronger negative relationship. The relationship does appear statistically significant based on the p-values, but the R-squared is once again very small.

```
##
## Call:
## lm(formula = TransitFrequency ~ income_thousands, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -49.20 -32.85 -17.81   8.83 823.71
##
## Coefficients:
##                  Estimate Std. Error t value          Pr(>|t|)
## (Intercept)      56.99550    7.32609   7.780 0.000000000000116 ***
## income_thousands -0.16863    0.07478  -2.255            0.0248 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67.29 on 302 degrees of freedom
##   (251 observations deleted due to missingness)
## Multiple R-squared:  0.01656,    Adjusted R-squared:  0.0133
## F-statistic: 5.085 on 1 and 302 DF,  p-value: 0.02484
```

Figure 4. Regression results, fit2

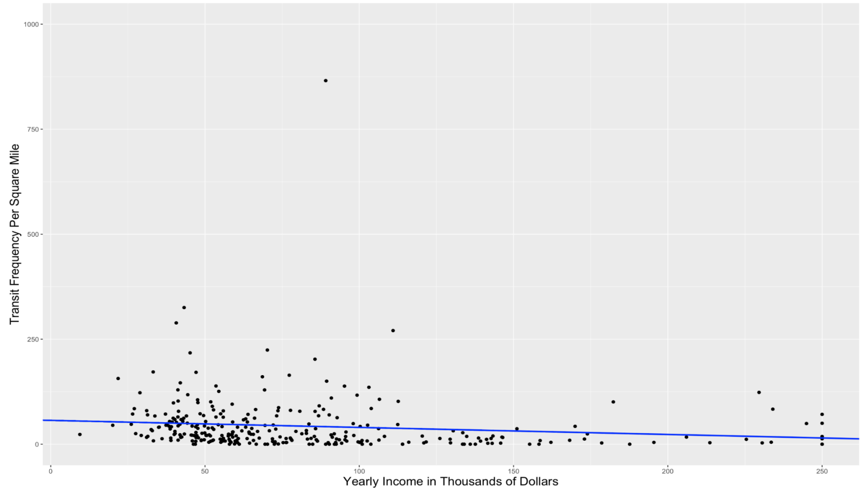

Figure 5. Plot and trend line of fit2

Now for figure 6, which represents the ‘JobsPerAcre’ variable and the ‘CountLowWage_work’ variable. In this map we are especially looking for areas with high job density and high low wage job density. These areas are commuter destinations, prime targets for transit connectivity. The University City Research Park is once again an area of note, as is Northlake Mall and much of the area south of it along I-77 heading into the city. Just southeast of that corridor are the neighborhoods of Eastland and Four Seasons, situated along Highway 27. These are areas with high numbers of low wage jobs and job density, and as we will see in figure 7, this area also has high transit frequency and high numbers of low wage homes. In fact, it is one of the few areas in the city to have a concentration of low wage workers’ homes, jobs, and a high transit frequency. Joining that club as well are the areas of Lockwood and Tryon Hills, just northeast of uptown Charlotte outside of the I-277 loop.

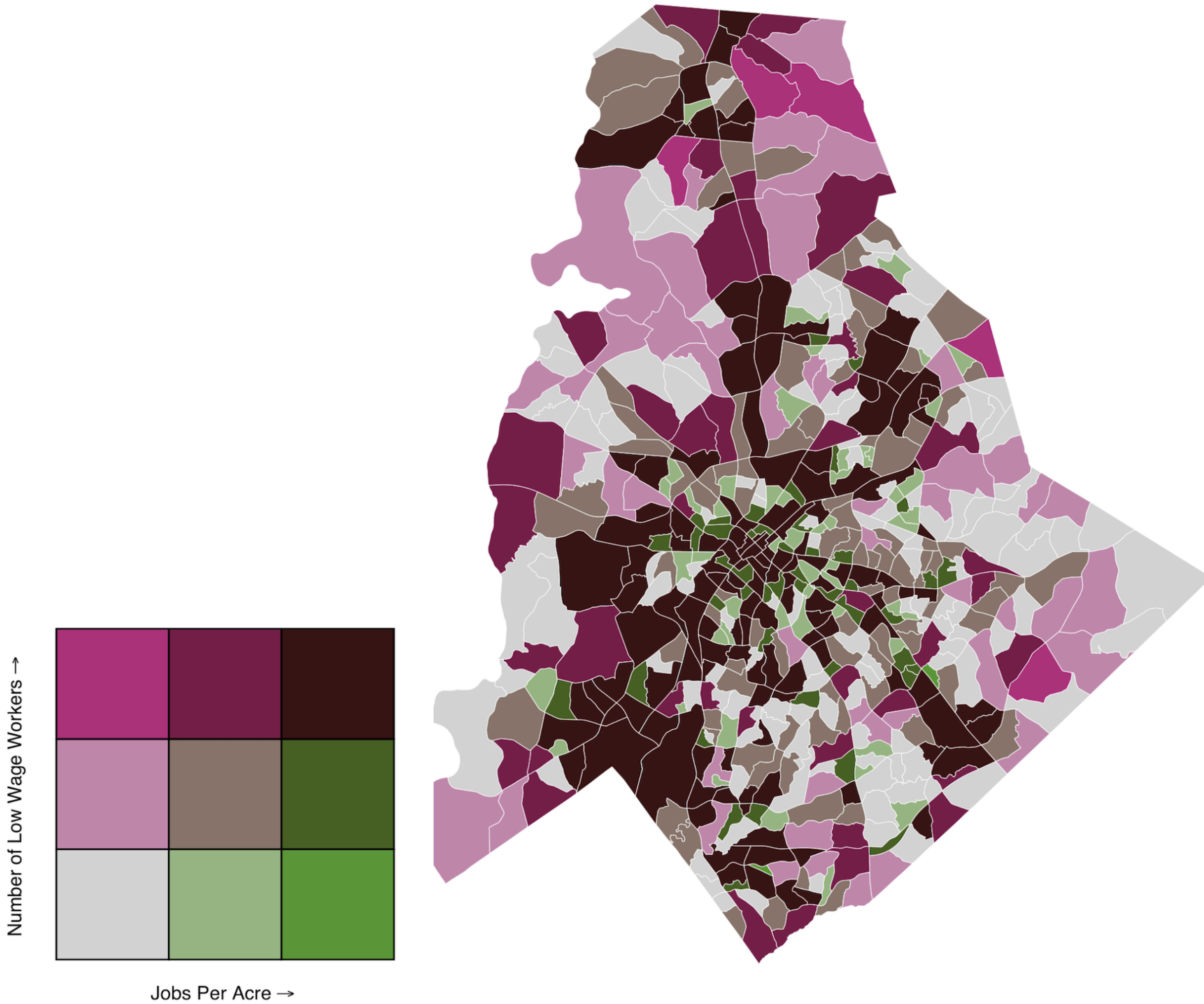

Figure 6. Low Wage Workers (Workplace) and Jobs Per Acre

Figure 7 shows the number of low wage workers' homes, and transit frequency. Rather than looking at the high end of both categories like the previous figure, here we are looking primarily for hot pink and light pink areas, showing block groups with high and medium levels of low wage worker homes, and low levels of transit service. Unlike the previous two relationship maps, much of the outskirts of this map of Mecklenburg county is filled in, and with the categories we're looking for. However, much of these areas are rural or disconnected suburban developments, so I will be focusing on the areas identified in the prior two maps and areas closer and more connected to those workplace concentrations. The University City Research Park is low on both, but some of the residential areas around it fall into the hot pink category. The same is true for Northlake Mall, which has large swaths of low wage worker homes surrounding it, but very low transit frequency in the area.

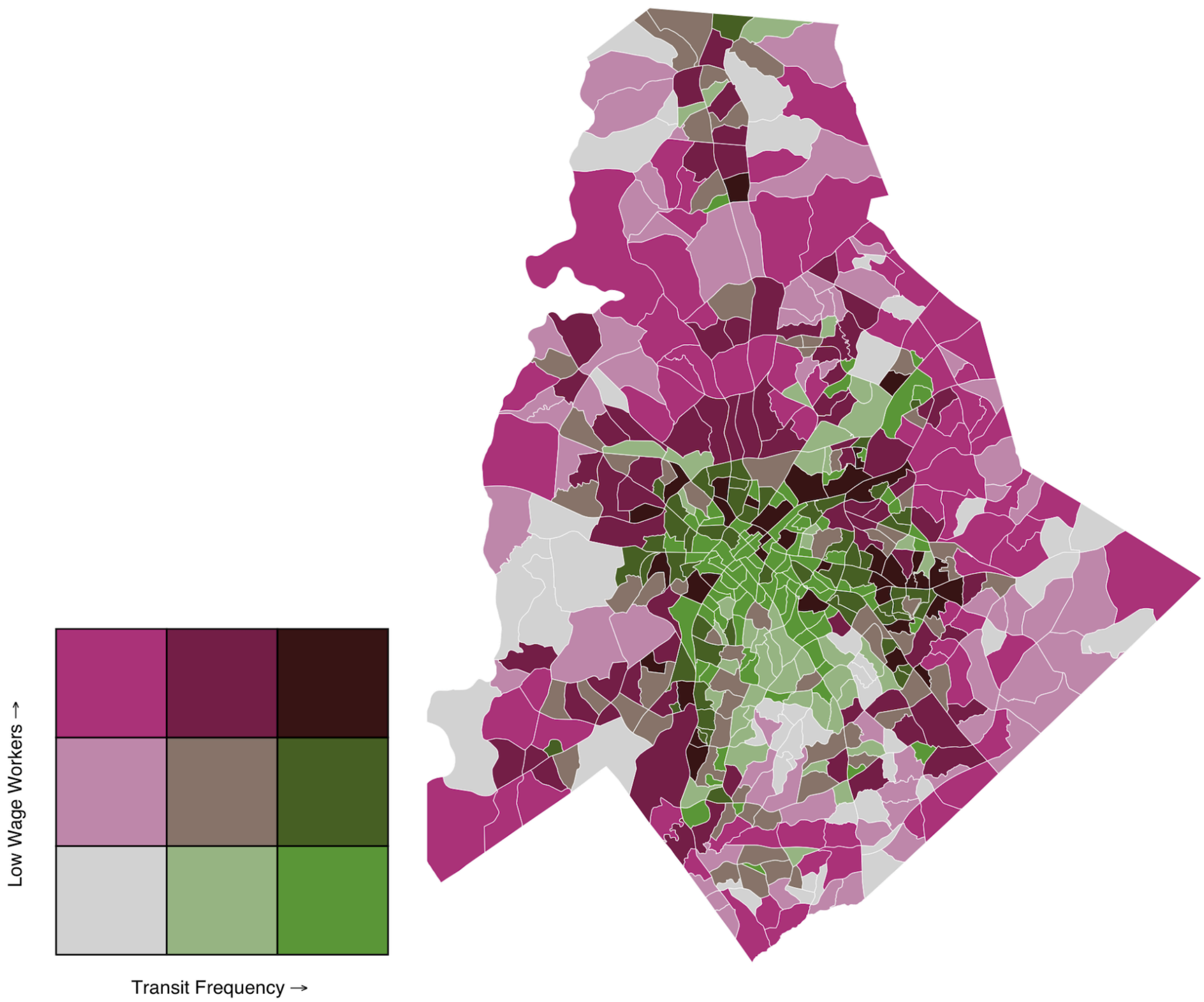

Figure 7. Low Wage Workers (Homes) and Transit Frequency

Conclusions

I believe this analysis is informative enough to draw some conclusions, while there are also limitations which make other conclusions less feasible. I will begin by discussing limitations of the study. First, the data contains some quirks which led to problems. The income variable had a max value of 250,001, which represented any income value equal to or greater than $250,001. When the regression in figure 4 reported a negative relationship, I was immediately skeptical, based on other choropleth maps I have seen of income in the center and south of the city being much higher than other areas, and seeing that transit frequency in those areas is also high in my own choropleth maps. The issue with income data may invalidate the results of that regression, but a majority of them appear to also be low transit frequency in the scatterplot, so I cannot be sure. Second, the lack of analysis of actual transportation networks is a major limitation in making this analysis informative for policy decisions by Charlotte-Mecklenburg transportation planners. There are a lot of bases which aren't covered in this study which should be covered in a robust evaluation intended to culminate in a new transit plan. This analysis is essentially a preliminary look at communities that might benefit the most from expanded transit service, while a follow-up study should focus on the logistics of the transportation networks interconnecting those areas to move towards a proposal for specific transit plans.

The results of the regression in figure 2, if taken as a statistically significant relationship, suggest that transit in the City of Charlotte does in fact improve in areas where people spend a greater percentage of their income on transportation. This is a conclusion I hoped to see at the outset of this study. If the result of the relationship was negative, I would consider that to be an issue in need of addressing in order to make public transit more equitable and accessible.

However, with the weak result of the uncertainty statistics in the observed relationship, it's likely that the city could be doing more to address these potential disparities.

To summarize the relevant findings of the three relationship maps, the University City Research Park and Northlake Mall are both areas with high activity density and low transit frequency. Those two areas are also high in job density and low wage job density, making them prime transit destinations for commuters. Several other areas in the city scoring high in these categories of job density also have high transit frequency, perhaps indicating that the city is providing effective public transit options for commuters to those areas. And finally, the areas around University City Research Park and Northlake Mall have numerous block groups where many low wage workers live, and transit frequency is low. Putting all of this information together, I would suggest to transit planners to consider expanded transit service along these areas with job concentrations and low wage homes in close proximity. Specifically, bus routes could be developed between Northlake Mall and University City Research Park, and rather than just traversing W.T. Harris Blvd/Hwy 24, create new bus stops in residential areas between these two destinations. This would provide the businesses at Northlake Mall increased activity, and allow citizens to travel to and from the mall without having to make a car trip, each individual increasing traffic and contributing to air pollution. It would also allow commuters around University City, the Research Park, and the residential areas between the Research Park and Northlake Mall who work in these areas to utilize public transit for their daily commute to work. This could provide great benefits to citizens and businesses in the area, and the economic success of the region as a whole.

Bibliography

Bivarite Mapping with Ggplot2. https://chris-prener.github.io/biscale/articles/biscale.html#bivariate-mapping-with-biscale. Accessed 24 Apr. 2025.

Hosseini, Sanaz Sadat, Babak Rahimi Ardabili, et al. "Demographic Disparities, Service Efficiency, Safety, and User Satisfaction in Public Bus Transit System: A Survey-Based Case Study in the City of Charlotte, NC." *Transportation Research Interdisciplinary Perspectives*, vol. 29, Jan. 2025, p. 101296. *DOI.org (Crossref)*, https://doi.org/10.1016/j.trip.2024.101296.

Hosseini, Sanaz Sadat, Mona Azarbayjani, et al. *Towards Understanding the Benefits and Challenges of Demand Responsive Public Transit- A Case Study in the City of Charlotte, NC*. arXiv, 2023. *DOI.org (Datacite)*, https://doi.org/10.48550/ARXIV.2304.06467.

Li, Wei, et al. "Effects of New Transit Lines on Commuting: Evidence from Restricted-Use Census Bureau Microdata." *Applied Geography*, vol. 164, Mar. 2024, p. 103202. *DOI.org (Crossref)*, https://doi.org/10.1016/j.apgeog.2024.103202.

Maps – Mecklenburg County. https://ncgenweb.us/mecklenburg/resources/maps/. Accessed 24 Apr. 2025.

Newsom, Mary. "An Urbanist Prescription for Sprawling Charlotte." *WFAE 90.7 - Charlotte's NPR News Source*, 16 June 2023, https://www.wfae.org/charlotte-area/2023-06-16/an-urbanist-prescription-for-sprawling-charlotte.

SimplyAnalytics. https://app.simplyanalytics.com/index.html. Accessed 24 Apr. 2025.

US EPA, OP. *Smart Location Mapping*. 27 Feb. 2014, https://www.epa.gov/smartgrowth/smart-location-mapping.