The Relationship of Transit Deserts and Inequality: A Case Study of Travis County Using Simple Linear Regression Analysis

Brian Folkers, Lucas Herrera
SDS 322E - Elements of Data Science - Fall 2021

Abstract

In recent years, the subject of Transit Deserts has gathered much attention from researchers in Urban Informatics. While the link between a lack of adequate public transportation and economic inequality is often reported by researchers, it is rarely studied in-depth. This paper seeks to expand upon the collected transit desert data, by examining the relationship between transit score and variables meant to represent economic inequality, including poverty, unemployment, and minority status. Using the Austin/Travis County area as a case study, this paper focuses on three simple linear regression models to compare census block's transit score with inequality. The results suggest a weak but positive correlation between the availability of public transportation and economic inequality within the Travis County area; however, additional guided research is required to properly evaluate this relationship on a regional or national scale.

Introduction

The United States, like much of the world, is a rapidly urbanizing country. With each passing year, more Americans rely on, or could rely on, public transportation to move from one place to another within cities. However, US cities still remain heavily reliant upon single-user vehicles, while public transportation is continuously neglected. According to Sanchez (2008), extensive research has been conducted throughout the last 60 years in various attempts to describe public transit's localized socioeconomic impacts, with low rates of car ownership and inadequate public transit options being identified common characteristics of lower-income communities. This general lack of access to cars in impoverished areas, coupled with the fact that many destinations within the United States are primarily or exclusively accessible by car, indicates that increased public transportation access may have a significant impact on local populations and economic mobility. The situation whereby public transit is increasingly needed but hardly supplied has often been referred to as a 'Transit Desert.' Researchers who study these Transit Deserts note both their prevalence across many American cities, and their tendency to exacerbate economic inequality.

The goal of this study is to more closely examine the relationship between a region's transit desert score (the degree to which public transportation supply meets demand) and a number of demographic variables meant to represent economic inequality including poverty, unemployment, and the number of ethnic minorities living in that area. The general hypothesis is that areas that are transit deserts (higher demand than supply, with a negative transit score) will have higher rates of poverty and unemployment than areas that are transit oases (higher supply than demand, with a positive transit score). Similarly, it is expected that there will be a positive correlation between an area's number of ethnic minorities and a lack of adequate public transportation. This study will act as a case study focusing on the City of Austin by using the available data within Travis County (note: the terms Travis County and Austin will be used interchangeably in this paper). Given this, the population of interest is the different regions of Travis County and the people that reside in them. The stakeholders are those interested in the demographic and economic effects of improved access to adequate public transportation. This

study will analyze the relationship between transit and the three aforementioned variables using simple linear regression analysis. This paper hopes to provide a more thorough analysis of the transit desert data to examine the claim that a lack of transportation is closely linked to poverty and inequality.

Research Questions:

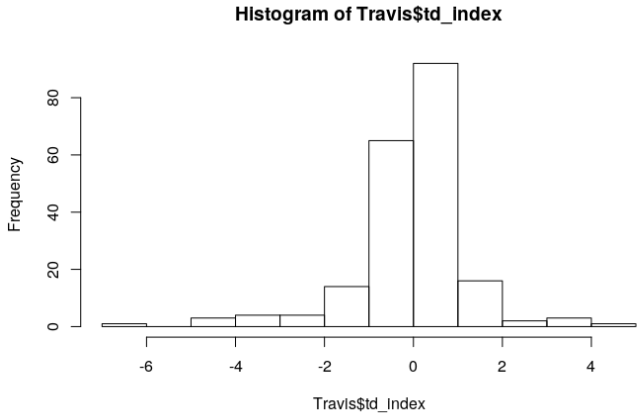
1. Is there a linear relationship between the poverty rate and the transit score in Austin?
2. Is there a linear relationship between the unemployment rate and the transit score in Austin?
3. Is there a linear relationship between the number of minorities and the transit score in Austin?

Data

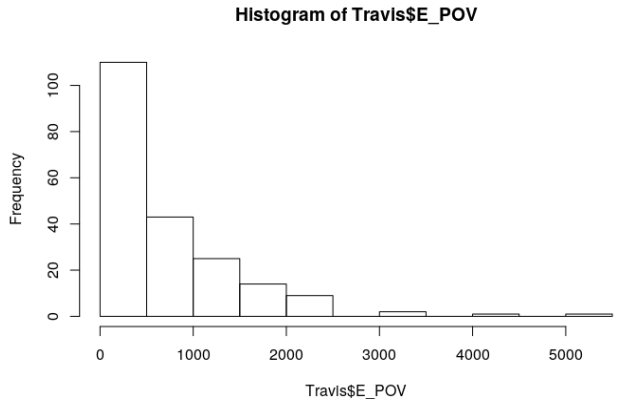
The dataset is borrowed from a series of studies on transit deserts conducted at the UT Austin Urban Informatics Lab. No further steps were taken to prepare data asides from removing NA values and limiting the sample size to areas in Travis county, the number of which is 206. The data is divided into two main sections: data related to transit score calculation and data about several demographic variables of each examined census block. This study will only focus on the transit desert score and the three previously mentioned explanatory variables.

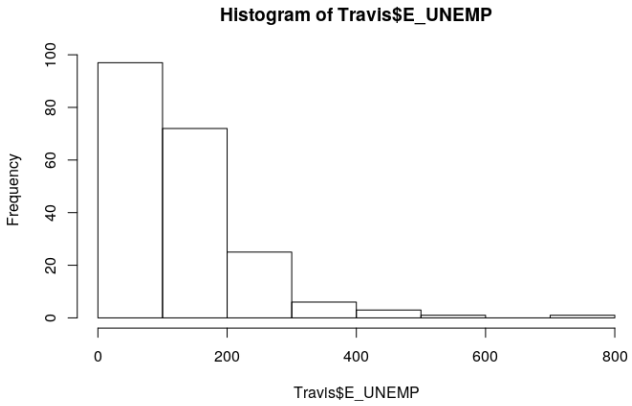
Exploratory Analysis

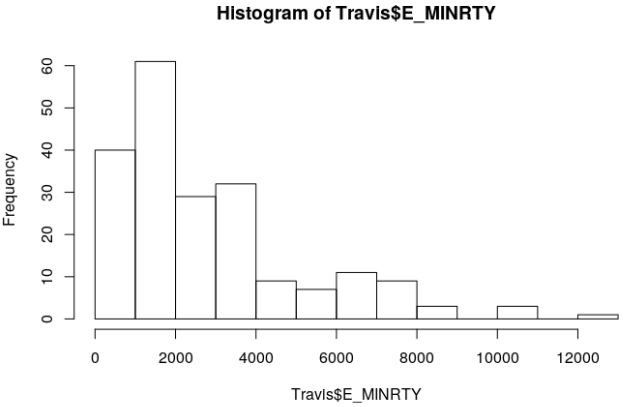
The initial distribution of transit scores is relatively normally distributed and did not require a transformation (mean=-0.05, standard deviation=1.26). One interesting attribute to note is how the transit score data does not vary considerably. Most values are just above or below 0 (adequately served).



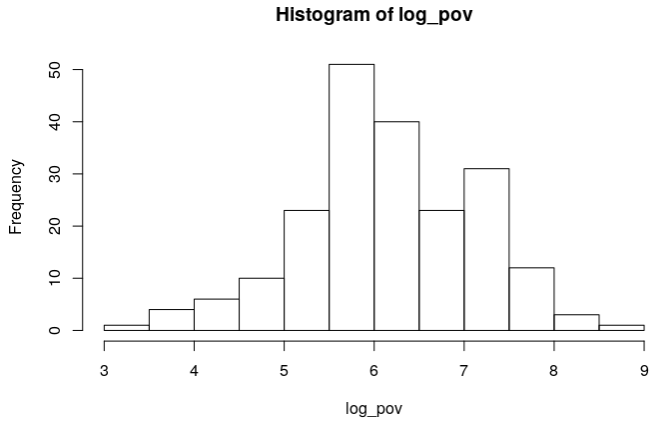
The initial distributions of response variables poverty (mean= 731.0, standard deviation= 746.84), unemployment (mean=129.4, standard deviation= 98.3), and minority (mean= 2870.8, standard deviation= 2324.1) were all heavily right skewed.

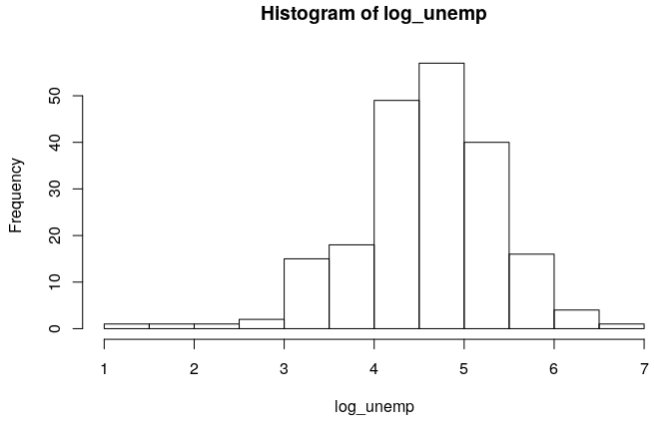


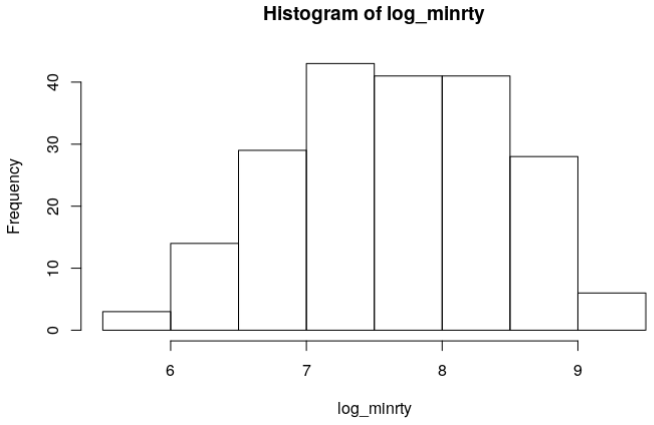




After performing a log transformation, unemployment (mean = 6.2, standard deviation = 1.0), poverty (mean = 4.6, standard deviation = 0.8), and minority (mean = 7.7, standard deviation = 0.8) were more normally distributed and better suited for regression analysis. The initial correlation tests were fairly low (poverty=0.25, unemployment=0.26, minority=0.44).







Given the previously mentioned suggestion that transit deserts (negative transit score) are closely related to inequality, the hypotheses are as follows:

1. There is a negative linear relationship between poverty and the transit score in Austin.
2. There is a negative linear relationship between unemployment and the transit score in Austin.

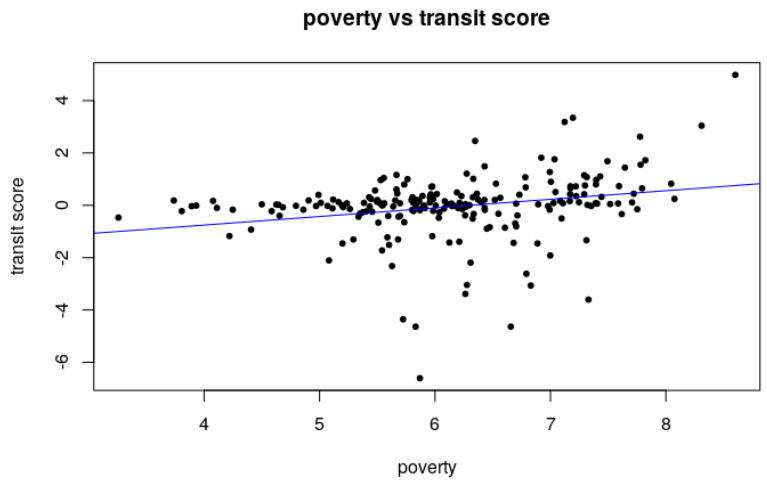
3. There is a negative linear relationship between the number of minorities and the transit score in Austin.

Modeling

To test each of the three aforementioned hypotheses, three separate linear regression analyses were performed comparing transit score to poverty, unemployment, and minority.

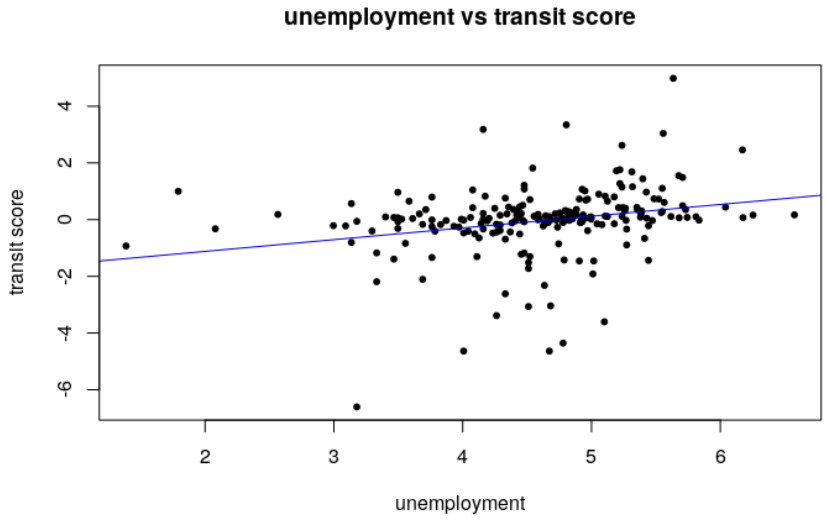
Poverty:

The data provides evidence that poverty is significantly linearly related to the transit desert score (t = 3.74, df =203, p =.000243). For every additional person in poverty, the transit desert score increases by 0.32 points, on average. For every additional increase in poverty, the transit desert score increases by between 0.156 and 0.501 points, on average, at the 95% confidence level. We can see that 6.4% of the variation in transit score can be explained by the variation in poverty. The QQ-plot for poverty is not normally distributed, so the assumption of normality is not met. The residual plot does not have an approximately symmetric cloud of points above and below the line, so the assumption of equal variance is not met. (see Appendix for assumption plots)



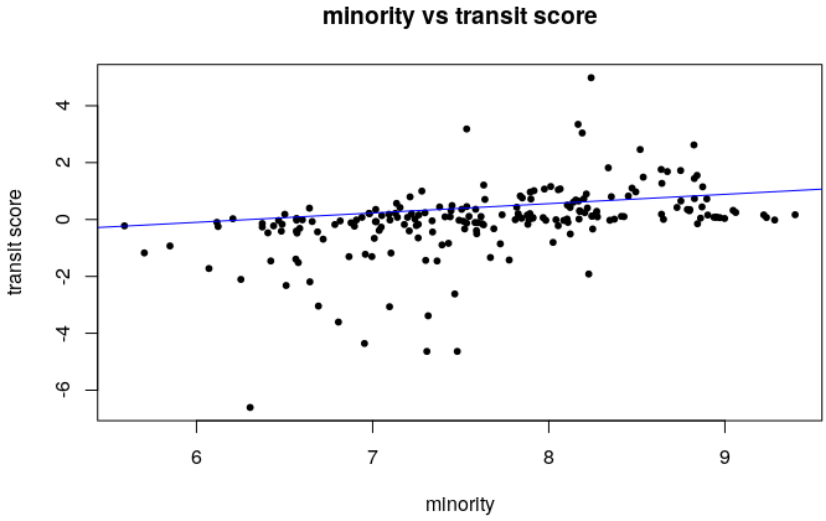
Unemployment:

The data provides evidence that unemployment is significantly linearly related to the transit desert score (t = 3.85, df =203, p =.000158). For every additional unemployed person, the transit desert score increases by 0.41 points, on average. For every additional increase in unemployment, the transit desert score increases by between 0.201 and 0.626 points, on average, at the 95% confidence level. We can see that 6.8% of the variation in transit score can be explained by the variation in unemployment. The QQ-plot for poverty is not normally distributed, so the assumption of normality is not met. The residual plot does have an approximately symmetric cloud of points above and below the line, so the assumption of equal variance is met. (see Appendix for assumption plots)



Minority:

The data provides evidence that minority is significantly linearly related to the transit desert score (t = 6.91, df =203, p =6.07e-11). For every additional minority person, the transit desert score increases by 0.67 points, on average. For every additional increase in minority, the transit desert score increases by between 0.480 and 0.862 points, on average, at the 95% confidence level. We can see that 19.1% of the variation in transit score can be explained by the variation in poverty. The QQ-plot for minority is not normally distributed, so the assumption of normality is not met. The residual plot does not have an approximately symmetric cloud of points above and below the line, so the assumption of equal variance is not met. (see Appendix for assumption plots)



Discussion

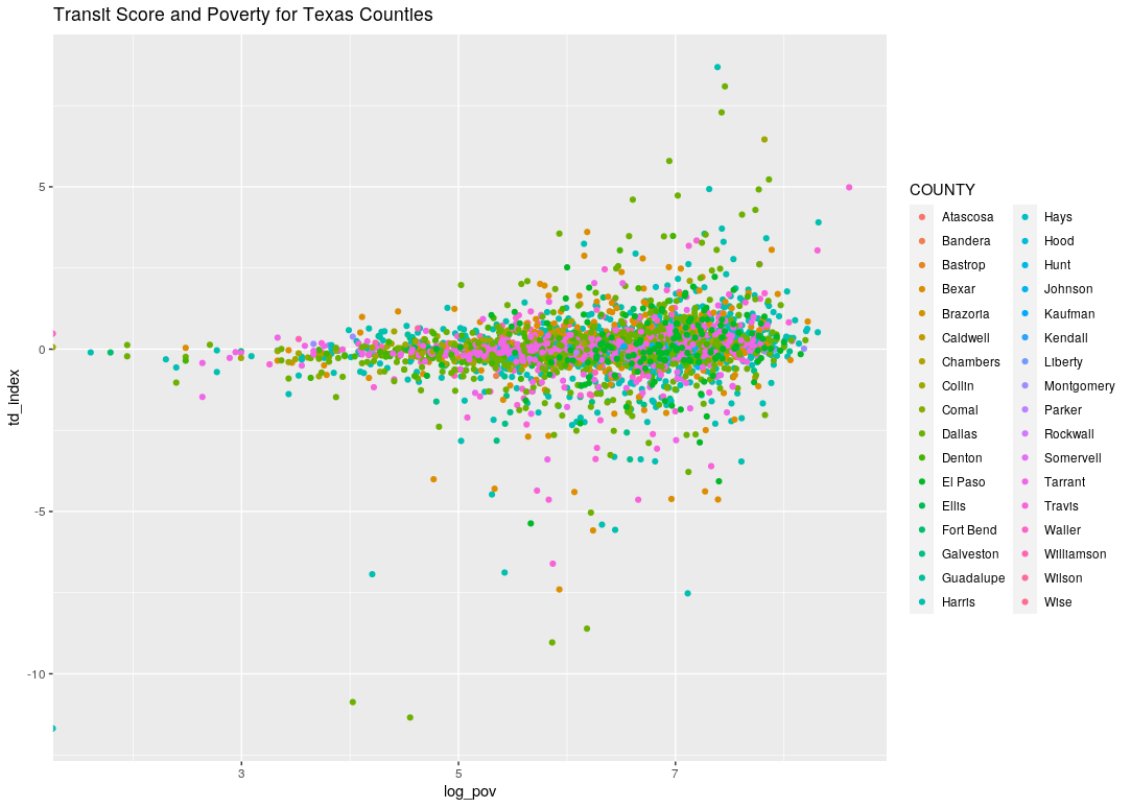
The results of this study do indicate that there is a significant linear relationship between the poverty rate, unemployment rate, and the number of minorities, and the transit score. However, contrary to the hypothesized outcomes, they all had positive correlations. Most research on transit deserts indicates a strong correlation between poverty and a lack of adequate public transportation (which is to say, a negative transit score).[1] This might indicate that Austin is an outlier in how poverty relates to public transportation. There are two main limitations for this study. Firstly, despite the significant results, most of the assumptions for the regression models were not met (see Appendix). Second, in an article comparing the public transit of five major cities in Texas, Austin was found to be the city with the least transit-dependent population (14.70%).[2] This might suggest that demographic factors would not correlate with a service that is not essential for most of the city and that Austin is an outlier.

Conclusion

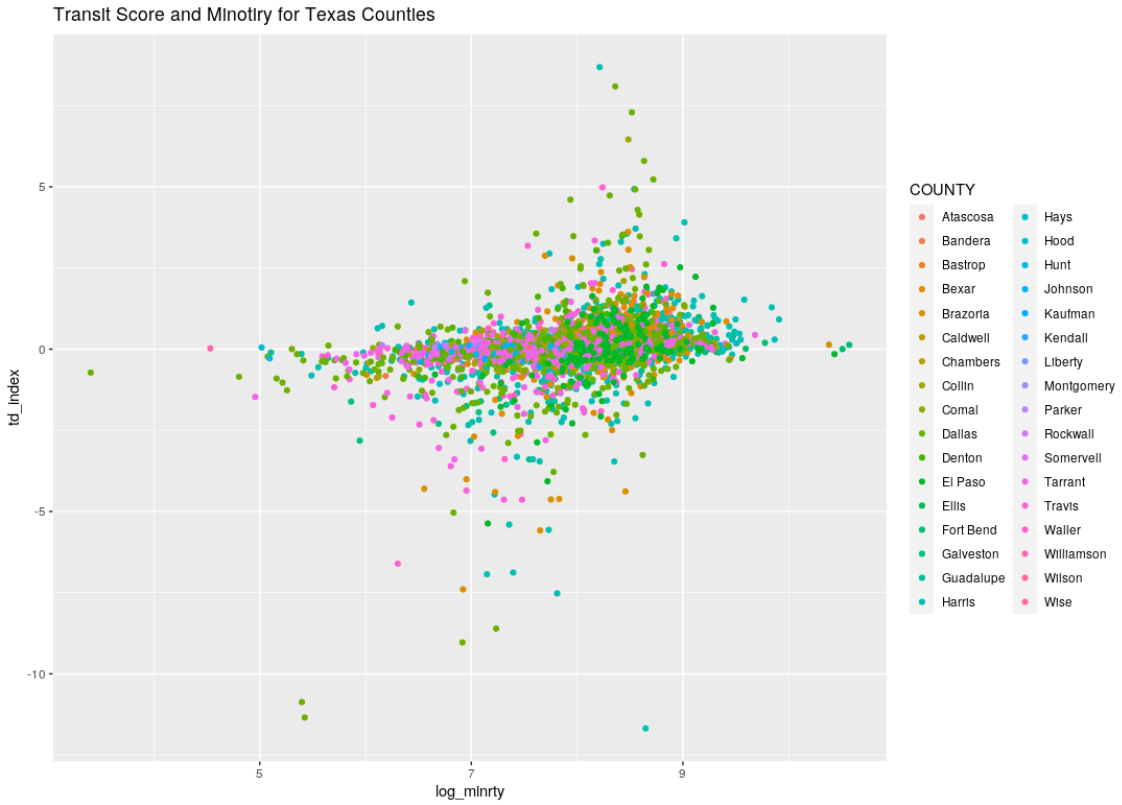
This paper acted as a case study to more closely examine the relationship between economic inequality and a lack of adequate public transportation. While a rather weak relationship was discovered, it indicated that areas with higher poverty, unemployment, and ethnic minorities had better access to public transportation than other areas. Hopefully, the findings of this study will be understood in context and act as motivation to learn more about the demographic factors that are most related to public transportation access in the city of Austin and other areas of the US.

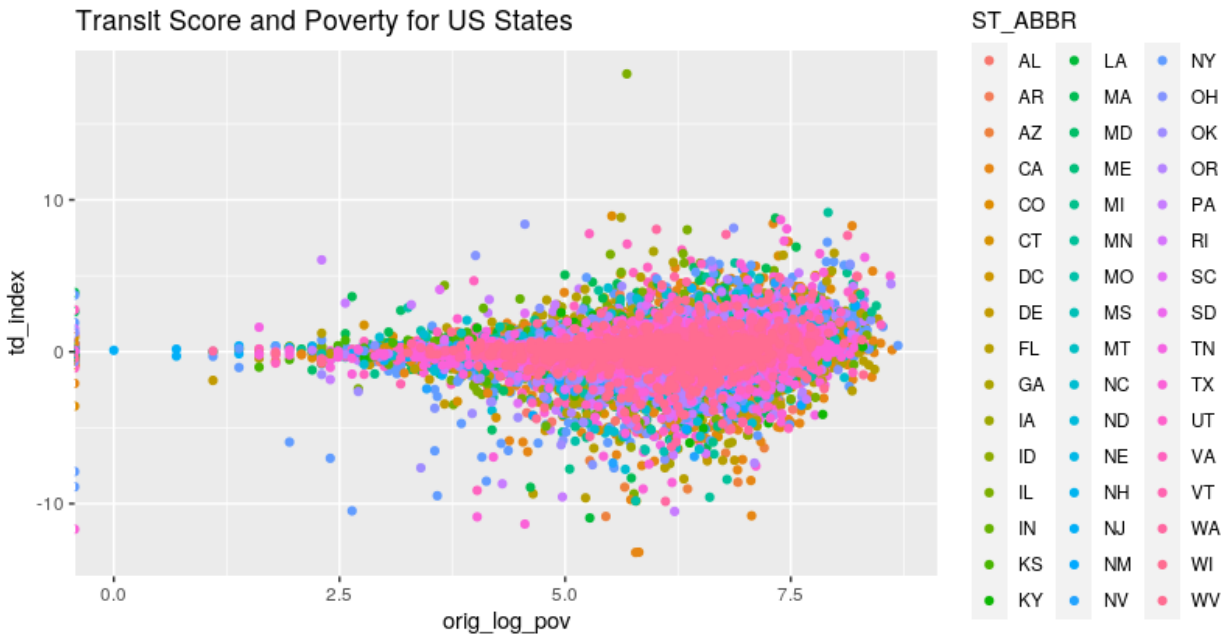
There are many avenues for future research on this subject. One might focus on more complex regression models like multiple regression to examine if there is an interaction effect between any of the variables. There are also plenty more demographic variables to work within the dataset (including age, disability, and single-parent status), so using best a correlation matrix or subsets regression to determine which variables to focus analysis on would also be a good next step. Similarly, studies that examine and compare other cities, states, or even all available data would also provide better context to the results of this project.

While the analysis of this study focused on a regression of the Austin data, this project did briefly explore the same data for Texas and the US. The results of which are worth mentioning given their potential relevance to future research. Using a subset of data from other non-Austin-area census tracts containing populations within one standard deviation of the mean Travis County tract population, linear regression analysis was performed for four different population metrics (estimated minority, unemployment, poverty, and total populations, respectively) in an attempt to build a demographic-based predictive model for a given tract's Transit Desert index score. However, the linear model proved a rather unsuccessful fit for all four variables on this national scale (all four regression coefficients obtained were smaller than 0.001), which corresponded to large mean-squared-error rates when the regression models were employed to predict new TD Index values. This may indicate that Austin's relationship between inequality and public transportation may be an outlier as similarly positive correlations between minority and socioeconomic categories were difficult to distinguish when examining other U.S. areas of similar population scale.

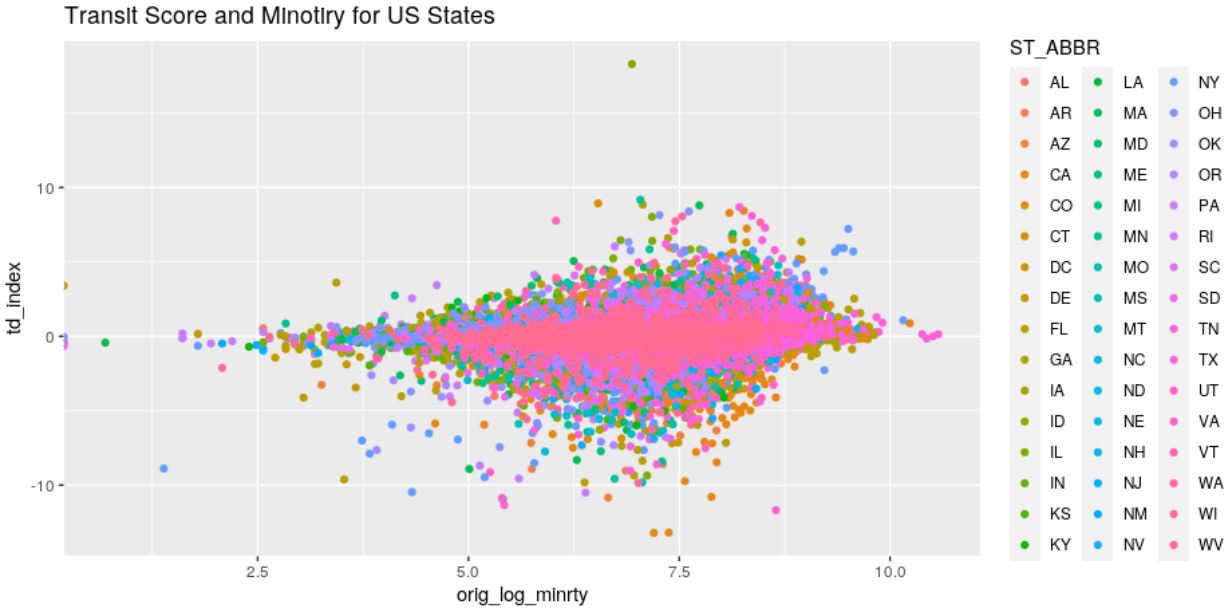

Transit Score and Poverty for Texas Counties
td_index
log_pov
5
0
-5
-10
3
5
7
COUNTY
Atascosa
Bandera
Bastrop
Bexar
Brazoria
Caldwell
Chambers
Collin
Comal
Dallas
Denton
El Paso
Ellis
Fort Bend
Galveston
Guadalupe
Harris
Hays
Hood
Hunt
Johnson
Kaufman
Kendall
Liberty
Montgomery
Parker
Rockwall
Somervell
Tarrant
Travis
Waller
Williamson
Wilson
Wise

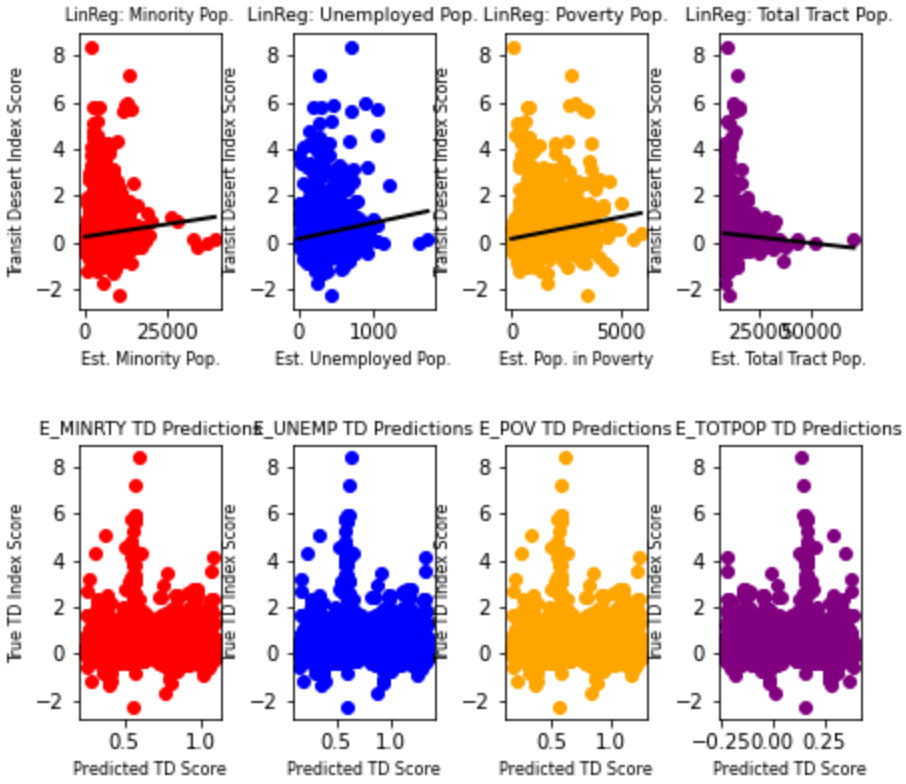

Transit Score and Unemployment for Texas Counties
td_index
log_unemp
5
0
-5
-10
2
4
6
COUNTY
Atascosa
Bandera
Bastrop
Bexar
Brazoria
Caldwell
Chambers
Collin
Comal
Dallas
Denton
El Paso
Ellis
Fort Bend
Galveston
Guadalupe
Harris
Hays
Hood
Hunt
Johnson
Kaufman
Kendall
Liberty
Montgomery
Parker
Rockwall
Somervell
Tarrant
Travis
Waller
Williamson
Wilson
Wise


Transit Score and Minotiry for Texas Countles
td_index
log_minrty
5
0
-5
-10
5
7
9
COUNTY
Atascosa
Bandera
Bastrop
Bexar
Brazoria
Caldwell
Chambers
Collin
Comal
Dallas
Denton
El Paso
Ellis
Fort Bend
Galveston
Guadalupe
Harris
Hays
Hood
Hunt
Johnson
Kaufman
Kendall
Liberty
Montgomery
Parker
Rockwall
Somervell
Tarrant
Travis
Waller
Williamson
Wilson
Wise


Transit Score and Poverty for US States
td_index
orig_log_pov
10
0
-10
0.0
2.5
5.0
7.5
ST_ABBR
AL
AR
AZ
CA
CO
CT
DC
DE
FL
GA
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
RI
SC
SD
TN
TX
UT
VA
VT
WA
WI
WV


Transit Score and Unemployment for US States
td_index
orig_log_unemp
10
0
-10
0
2
4
6
ST_ABBR
AL AR AZ CA CO CT DC DE FL GA IA ID IL IN KS KY
LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV
NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV


Transit Score and MInotIry for US States
td_index
orig_log_mInrty
10
0
-10
2.5
5.0
7.5
10.0
ST_ABBR
AL AR AZ CA CO CT DC DE FL GA IA ID IL IN KS KY
LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV
NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV



Acknowledgments

Brian Folkers - GitHub Repository, R Coding, Introduction, Data Collection/Cleaning, Exploratory Analysis, Regression Modeling, Discussion, Conclusion, Bibliography, Presentation
Lucas Herrera - Background Research, Abstract, Introduction, Bibliography, Conclusion, Discussion, Python Coding, Exploratory Analysis, Regression Modeling

Bibliography

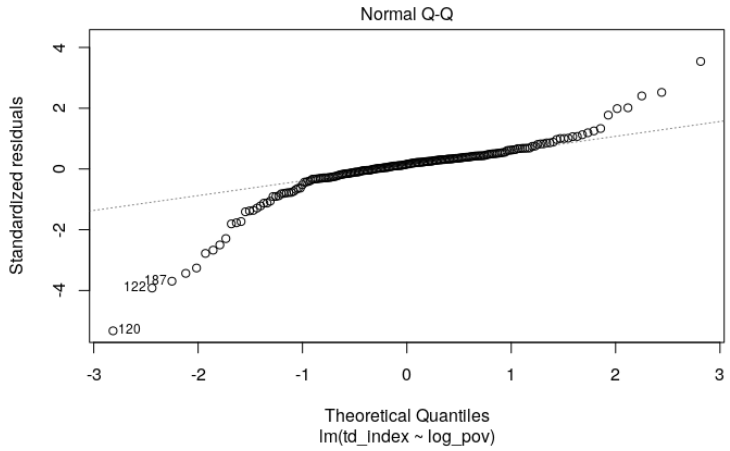
1. Junfeng Jiao Assistant Professor of Community and Regional Planning and Director, and Chris Bischak Masters of Community and Regional Planning Candidate. “People Are Stranded in 'Transit Deserts' in Dozens of US Cities.” *The Conversation*, 24 Mar. 2021, https://theconversation.com/people-are-stranded-in-transit-deserts-in-dozens-of-us-cities-92722.
2. Jiao, Junfeng. “Identifying Transit Deserts in Major Texas Cities Where the Supplies Missed the Demands.” *Journal of Transport and Land Use*, vol. 10, no. 1, 2017, https://doi.org/10.5198/jtlu.2017.899.

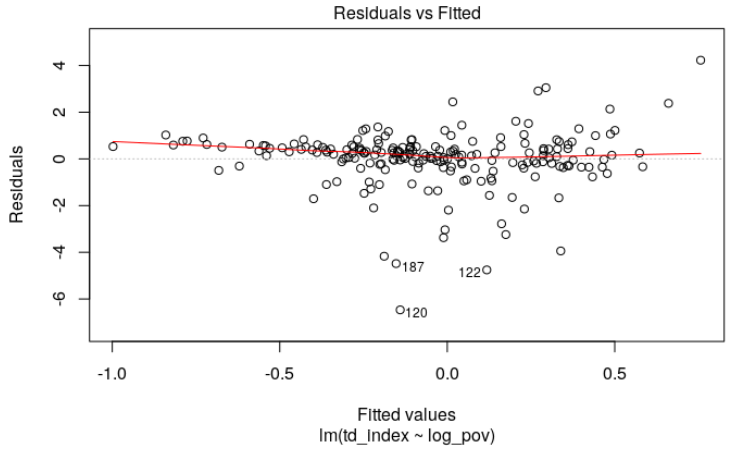
3. Sanchez, Thomas W. “Poverty, policy, and public transportation.” Transportation Research Part A: Policy and Practice, Volume 42, Issue 5, 2008, Pages 833-841. https://doi.org/10.1016/j.tra.2008.01.011.

Data Source: https://utexas.app.box.com/s/thxupji8napgmebucghkxvl0m083q3z6

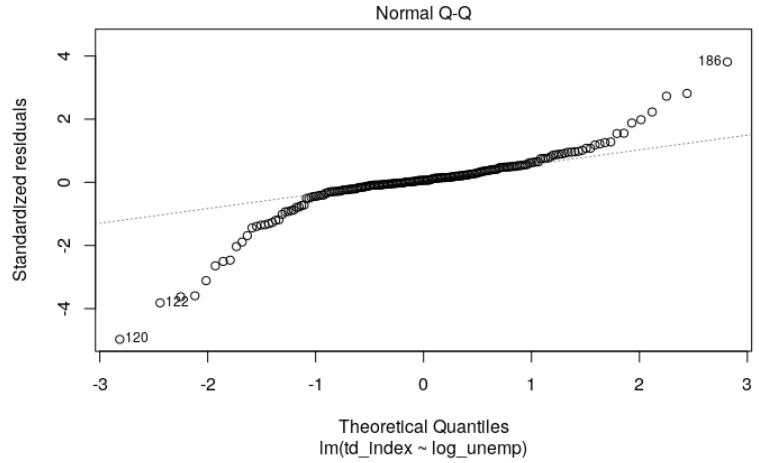
Appendix

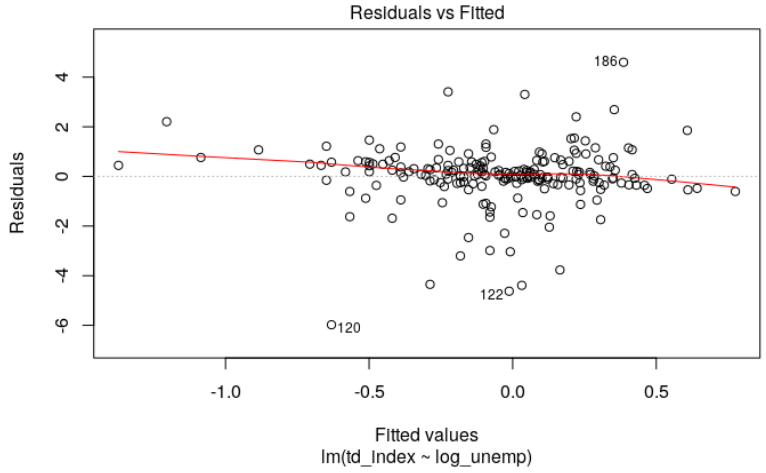
Assumptions plot for poverty simple linear regression:



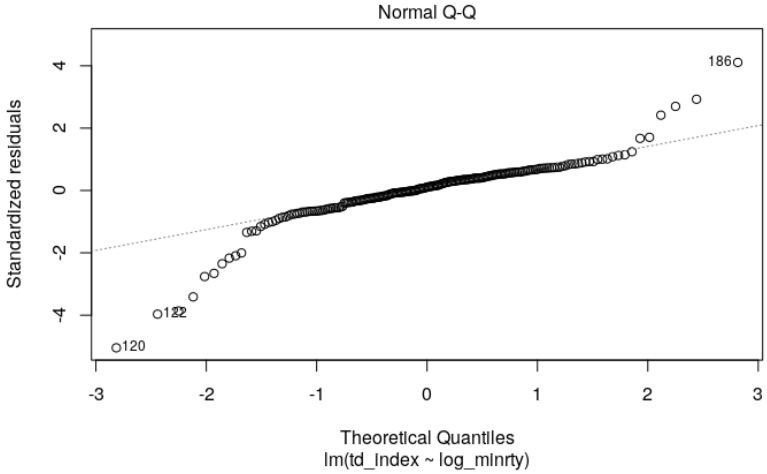


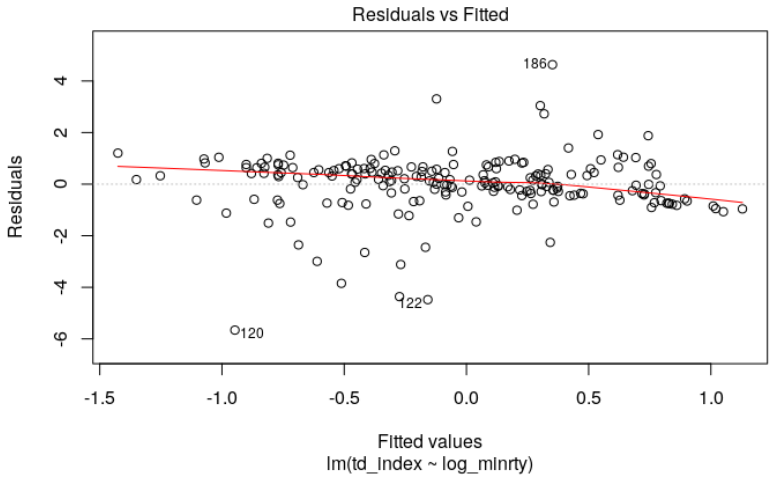
Assumptions plot for unemployment simple linear regression:





Assumptions plot for minority simple linear regression:




Residuals vs Fitted
Residuals
Fitted values
lm(td_index ~ log_minrty)
186
122
120
4
2
0
-2
-4
-6
-1.5
-1.0
-0.5
0.0
0.5
1.0