Predicting Broadband Speed in Louisiana

Robert Amponsah, Sharon Gloyer, Emily Myers, Elysa Vargas

12/12/2021

Louisiana's Digital Divide

Introduction

The COVID–19 pandemic and subsequent social distancing mandates resulted in a surge of internet-enabled remote services and opportunities within healthcare, education, and workforce systems. Some sources indicate that, "Internet services have seen rises in usage from 40% to 100%, compared to pre-lockdown levels" (De et al., 2020). Out of necessity people began taking advantage of telehealth services, online schooling, and a myriad of office-based organizations pivoted to work from home models. Unfortunately, the flexibility these technological advances offered were not accessible to everyone. According to the Federal Communications Commission (FCC), even though the digital divide is narrowing, there are still an estimated 21 million people in the United States without internet access or sufficient internet speed. These non-users, as noted in Kate Farrish's (2020) article "Health Care and Education Suffer When There's No Internet Access," are more likely to be poor, living in rural areas, and are Black or Hispanic. Without reliable internet services during this time, and in the future, these marginalized groups are subject to additional inequities within the healthcare, education, and workforce sectors.

Having travelled significantly down the path of virtual operations, it is likely the US will continue to see increased telehealth services, online schooling, and work from home opportunities in the years to come. Another likelihood will be additional government incentives to improve network infrastructure as has been evident with the Connect America Fund and the Rural Digital Opportunity Fund (RDOF). Though these programs are in their infancy, the reverse bidding process associated with RDOF has already been criticized by experts for overpromising on speeds that are unrealistic (Rivkin-Fish, 2021). Thus, a critical question to consider during future expansion is where providers will choose to leverage this funding and what the effect will be on the communities that receive it.

One potentially vulnerable population in need of stronger internet infrastructure is the state of Louisiana. This year the US News and World Report ranked Louisiana 46th for healthcare utilizing access to care, quality of care and the overall health of the population as indicating metrics. Additionally, they ranked Louisiana 48th for education, factoring in performance in higher education as well as primary and secondary schooling and pre-K education. Further, the U.S. Bureau of Labor Statistics' most recent unemployment rates reveal the state is ranked 38th for unemployment. If the country continues to shift towards virtual services, Louisiana, who is ranked 40th in the country for broadband coverage, prices, and speeds (Cooper & Tanberk, 2021), needs significant updates to their network to ensure all residents have reliable service. The FCC currently identifies those with less than a 25 megabits per second (mbps) broadband download speed as unserved while those below 100 mbps are considered underserved. A current initiative, the Broadband for Everyone in Louisiana (BEL) commission, aims to ensure that by 2029 everyone in the state is equipped with a speed of at least 25 mbps, a step in the right direction. The images in Figure 1 indicate continued limited access to acceptable broadband speeds across the state including adequate coverage throughout highly-populated, urban areas (red reflects unserved and underserved areas).

As providers continue to bid for federal and state funds, a primary concern is the impact to investment return (IRR). This suggests the most important factors for growth within a region include number of customer locations, population density, and commuting averages. Without overlaying other demographic characteristics

Figure 1: Broadband Coverage - Louisiana (State view, New Orleans, Baton Rouge, Shreveport) Source: National Telecommunication and Information Administration

of a US Census tract (a geographic US Census unit), will providers inadvertently perpetuate systemic inequities?

To address the digital divide within Louisiana which may further exacerbate other systemic inequities, the central research question of this work concerns **the degree to which regional characteristics predict an area's broadband speed. Is there a relationship between broadband speed and Louisiana census tracts as classified by the rural-urban commuting codes?** Ultimately, we intend to offer recommendations to providers regarding additional indicators to consider when strategizing growth and network improvements within Louisiana.

The Data

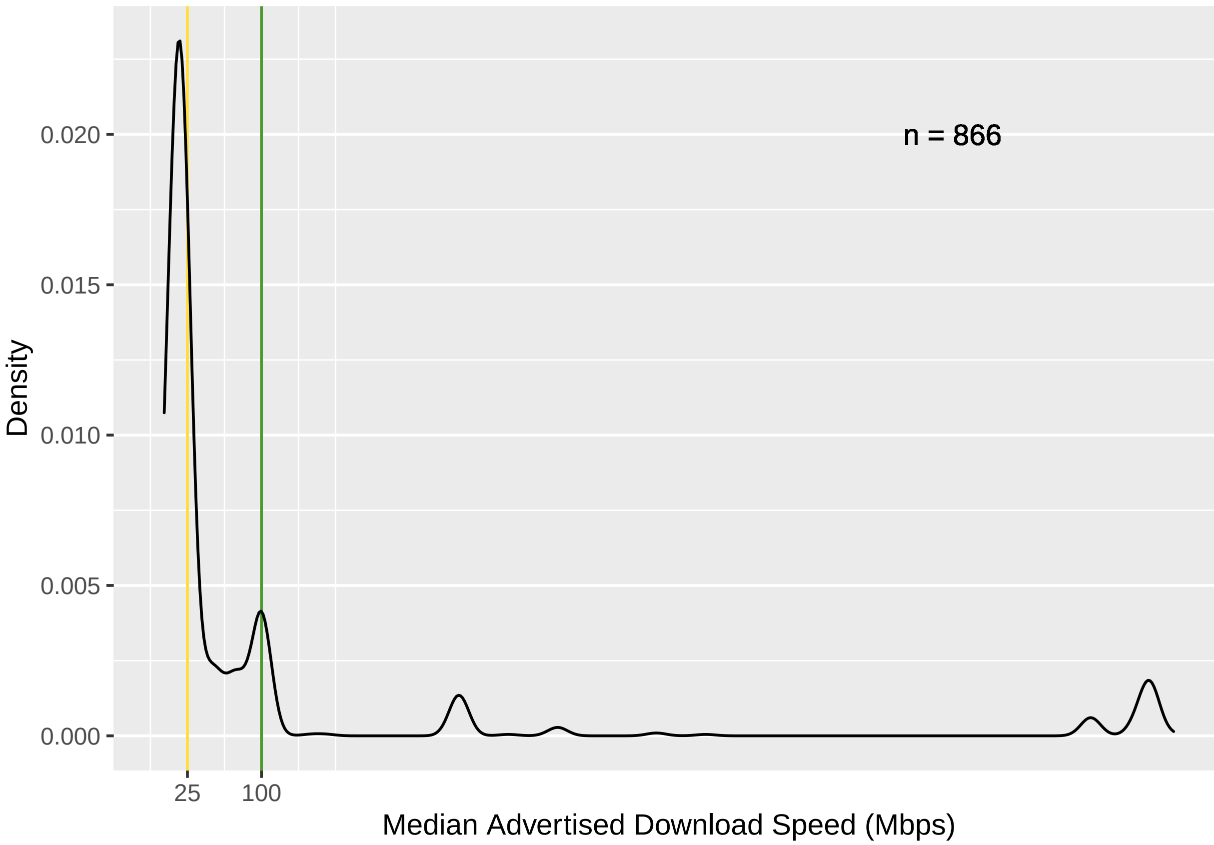
The first data set used in this report is the Fixed Broadband Deployment data from the FCC. This data set gives information on broadband maximum advertised download speeds broken down by county, zip code, and FIPS (Federal Information Processing Standard) code. From the FIPS code we could extract tract codes. Tracts are small, mostly permanent sub-county geographic areas designated by the US Census Bureau. We matched tract codes in the FCC data set with tract codes in the Rural-Urban Commuting Area (RUCA) codes dataset (the latest from 2010) to summarize geographic area types. RUCA codes are developed by the US Deparment of Agriculture (USDA) and classify tract codes into 11 different categories ranging from rural to metropolitan areas. RUCA codes also classify U.S. census tracts using measures of population density, urbanization, and daily commuting. This focus on geographic area types and their additional measures helps us determine what discrepancies exist in access to broadband at reasonable advertised download speeds in various types of geographic areas based on population density and commuting pattern. With the first step of the analysis, we attempt to predict if geographic areas of varying population and commuting sizes experience different broadband speeds and whether or not federal grant funds issued in the Obama era were appropriately allocated to address rural broadband access.

For our secondary analysis, we use US Census data to investigate how/if demographics of similarly classified RUCA code areas have an effect on the root mean squared error of our original analysis. In particular, we focus on median income, race/ethnicity (specifically Black and Hispanic populations based on the literature review), and population count as we conduct our secondary analysis. We chose these categories because we feel the information provided can be used to determine and ensure funds from new grants and legislative allocations are equitably distributed across the state of Louisiana.

In terms of data preparation, we made a key decision to summarize the primary dependent variable, broadband speed, by taking the median advertised broadband speed per Tract Code. We used median instead of mean due to account for the wide and inconsistent range of possible speeds (0 to 1000 mbps) that could create outliers which would increase the chances of a skewed mean. We used advertised speed vs. actual speed because there is wide variation in the amount of speed experienced for customers over even short periods of time. Advertised speed allowed for a consistent benchmark across providers and tracts. We also made the decision to segment out broadband types that do not reflect the typical household experience (for example, we excluded satellite broadband). Finally, we kept in the RUCA category of "No RUCA" because though the USDA definition means that there was no population in the area in 2010, there could be today, and the FCC data which is more recent may reflect broadband availability in these areas.

Descriptive Statistics

Distribution of Median Download Speed in Louisiana



The density plot does not show a normal distribution of median download speed. The state of Louisiana has a higher frequencies of speed between 0 and 100 (which the FCC qualifies as underserved internet speed) which further demonstrates the need for greater infrastructure in the state as most of the tract areas do not have sufficient download speeds, according to the reported median download speeds of the tract.

Conditional mean of broadband speed at the level of geographic area classification

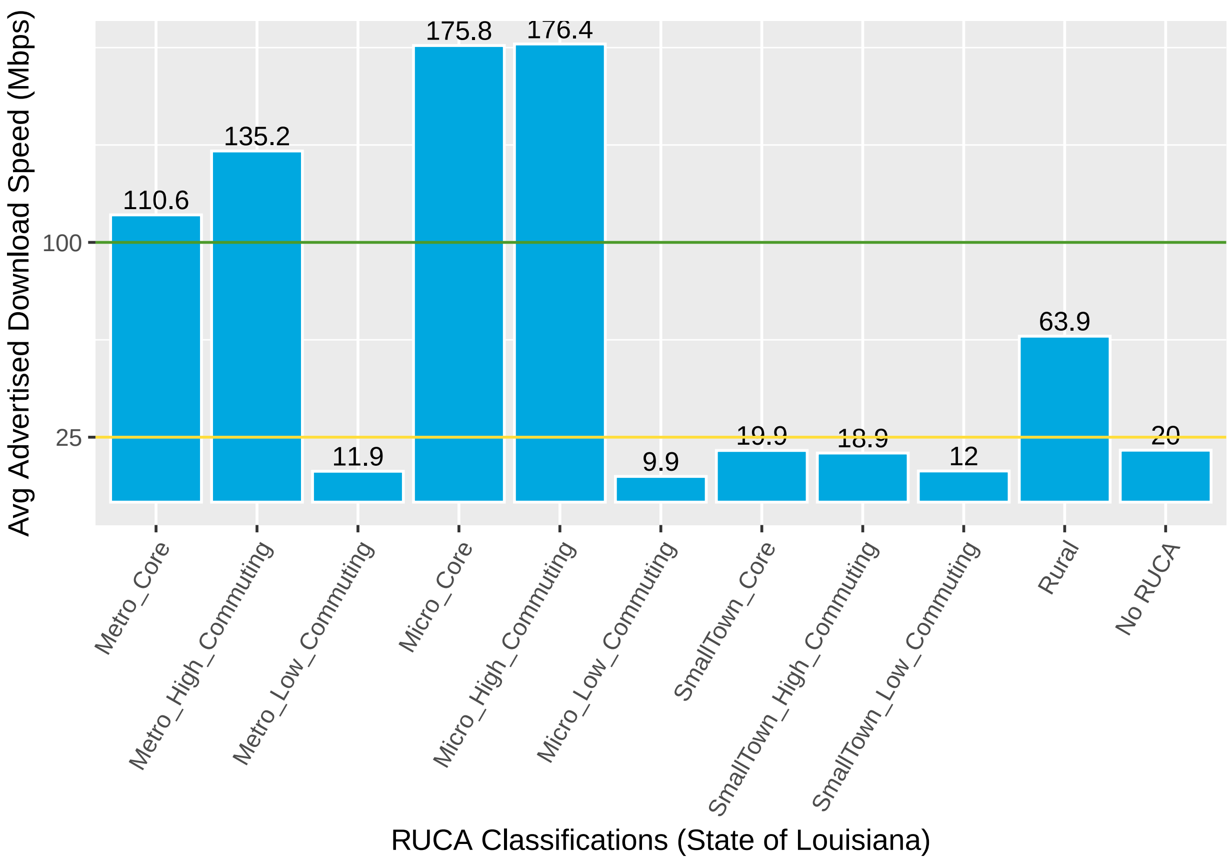
Next we look at RUCA classification as a prediction of broadband speed. We first further define the classifications:

RUCA_Classification	Area_Definition
Metro_Core	primary flow within an urbanized area (UA)
Metro_High_Commuting	primary flow 30% or more to a UA
Metro_Low_Commuting	primary flow 10% to 30% to a UA
Micro_Core	primary flow within an urban cluster of 10,000 to 49,999 (large UC)
Micro_High_Commuting	primary flow 30% or more to a large UC
Micro_Low_Commuting	primary flow 10% to 30% to a large UC
Rural	primary flow within an urban cluster of 2,500 to 9,999 (small UC)
SmallTown_Core	primary flow 30% or more to a small UC
SmallTown_High_Commuting	primary flow 10% to 30% to a small UC
SmallTown_Low_Commuting	primary flow to a tract outside a UA or UC
No RUCA	census tract has zero population and no rural-urban identifier information

This table reflects the average advertised broadband speed for each classification:

RUCADescription	avgspeed
Micro_Low_Commuting	9.857143
Metro_Low_Commuting	11.875000
SmallTown_Low_Commuting	12.000000
SmallTown_High_Commuting	18.900000
SmallTown_Core	19.850000
No RUCA	20.000000
Rural	63.857143
Metro_Core	110.633731
Metro_High_Commuting	135.205000
Micro_Core	175.846154
Micro_High_Commuting	176.400000

The graph below is a bivariate graphical representation comparing average broadband speed to RUCA classification. Vertical threshold lines represent where the FCC characterizes areas as *unserved* (<25Mbps, yellow line) and *underserved* (<100Mbps, green line). This is ordered by largest geographic area type to smallest so that we can look at insights across different rural/urban area types.



One of the more obvious findings from the bivariate analysis is that there is not a constant decrease in broadband speed from more populated to least populated areas. In fact, the Rural geographic area has a higher average broadband speed greater than any of the Small Town areas which have higher populations. The focus on rural area infrastructure during the Obama presidency may contribute to this outcome, however, small town infrastructure may have been neglected. It is also interesting to note that high commuting areas always have faster average broadband speeds than low commuting ones. We would have thought that the opposite was true and given the RUCA code data is from 2010, it would be interesting to investigate what happens in the current pandemic era when high commuting area become low as work and school from home have increased.

Conditional means of additional predictors

In this section we use the FCC's criteria for categorizing the median download speed of a tract as either underserved (<=100 megabits per second) or served (>100 megabits per second). By understanding the relationship of our independent variables to median download speed as categorized by service level, we can improve our linear regression model to ultimately how broadband speed impacts a particular area. This will also provide clarity around which characteristics may not currently be layered into a provider's IRR and thereby potentially perpetuate systemic inequities within a given tract.

Population

Pop_Quartiles	Served	Underserved
1	0.0280374	0.9719626
2	0.1028037	0.8971963
3	0.1401869	0.8598131
4	0.1915888	0.8084112

The population quartiles reflect groupings of tract population from the lowest populated tracts (1) to the highest populated tracts (4). This crosstab output indicates that nearly 97% of tracts within the lowest population quartile have a median download speed that would be considered underserved. This decreases to 81% for tracts that fall within the highest population quartile. Overall, this indicates a relationship between tract population and broadband speed with the more populous areas having faster broadband speeds. However, even in the most populous tracts, only 19% are adequately served (>100 mbps).

Income Levels

Income_Quartiles	Served	Underserved
1	0.0467290	0.9532710
2	0.1308411	0.8691589
3	0.1308411	0.8691589
4	0.1542056	0.8457944

Median household income is the independent variable associated with levels of poverty within a given tract. The income quartiles reflect groupings of tract median household income from the lowest income tracts (1) to the highest income tracts (4). Additionally, this crosstab demonstrates that these poorer areas also have a substantial proportion of lower broadband speeds. While the difference between the lowest and highest quartile is not as great as with population, we do see that in areas in the highest quartile of median income have fewer areas that are considered underserved.

Percentage of Hispanic/Latino Residents

Hispanic_Quartiles	Served	Underserved
1	0.1168224	0.8831776
2	0.1635514	0.8364486
3	0.0794393	0.9205607
4	0.1028037	0.8971963

The Hispanic/Latino quartiles reflect groupings of tract percentage of Hispanic/Latino residents from tracts with the lowest population (1) to the highest population (4). This crosstab demonstrates an inconsistent relationship between areas with lower or higher concentration of Hispanic/Latino residents and broadband speed. However, in general, areas with more Hispanic/Latino residents also have a greater proportion of download speeds that are considered to be underserved. Thus, if providers are only looking at income and population they might miss the nuance of providing reliable internet service to a significant community of Hispanic/Latino residents.

Percentage of Black Residents

Black_Quartiles	Served	Underserved
1	0.1168224	0.8831776
2	0.1401869	0.8598131
3	0.1168224	0.8831776
4	0.0887850	0.9112150

The Black quartiles reflect groupings of tract percentage of Black residents from tracts with the lowest population (1) to the highest population (4). Similarly with Black residents it is unclear if there is a negative relationship between higher percentages of Black residents and download speeds. However, the quartile with the greatest percentages of Black residents also represents areas where download speeds fall below the FCC's metric for acceptable internet capabilities.

Analysis

Based on our results from the conditional means exercise, we run a model that predicts the advertised download speed (dependent variable) as a function of the RUCA classification, population, and income (independent variables that seemed to predict a relationship). We first check to ensure that there is a significant difference between the independent variables we are taking into our model. Running a Kruskal-Wallis test reveals p-values less than the significance level of 0.05, and thus we can conclude that there are significant differences between the variables. Before running the linear regression, the No RUCA values were removed as even though they have advertised broadband speeds, their absence of median income and population cause a prediction/fit error in regression analysis. In addition, their small count (n=3) does not have a significant impact on regression results.

Regression Results

term	estimate	std.error	statistic	p.value
(Intercept)	-23.3017915	35.3192486	-0.6597477	0.5097844
Tract_Population	0.0427871	0.0078431	5.4553821	0.0000001
Household_Med_Income	-0.0000497	0.0004767	-0.1041713	0.9170841
RUCADescription_Metro_High_Commuting	-33.7946935	36.9563197	-0.9144496	0.3610155
RUCADescription_Metro_Low_Commuting	-90.3657896	117.4270254	-0.7695485	0.4420087
RUCADescription_Micro_Core	-6.1782258	48.7861318	-0.1266390	0.8992879
RUCADescription_Micro_High_Commuting	47.4498178	71.7780118	0.6610634	0.5089411
RUCADescription_Micro_Low_Commuting	-116.0678443	165.5248254	-0.7012111	0.4835671
RUCADescription_Rural	-85.9779851	83.7004752	-1.0272102	0.3049242
RUCADescription_SmallTown_Core	-91.8608050	57.3620298	-1.6014218	0.1100500
RUCADescription_SmallTown_High_Commuting	-138.3365794	96.5454382	-1.4328650	0.1526548
RUCADescription_SmallTown_Low_Commuting	-73.4703699	135.5186935	-0.5421420	0.5880138

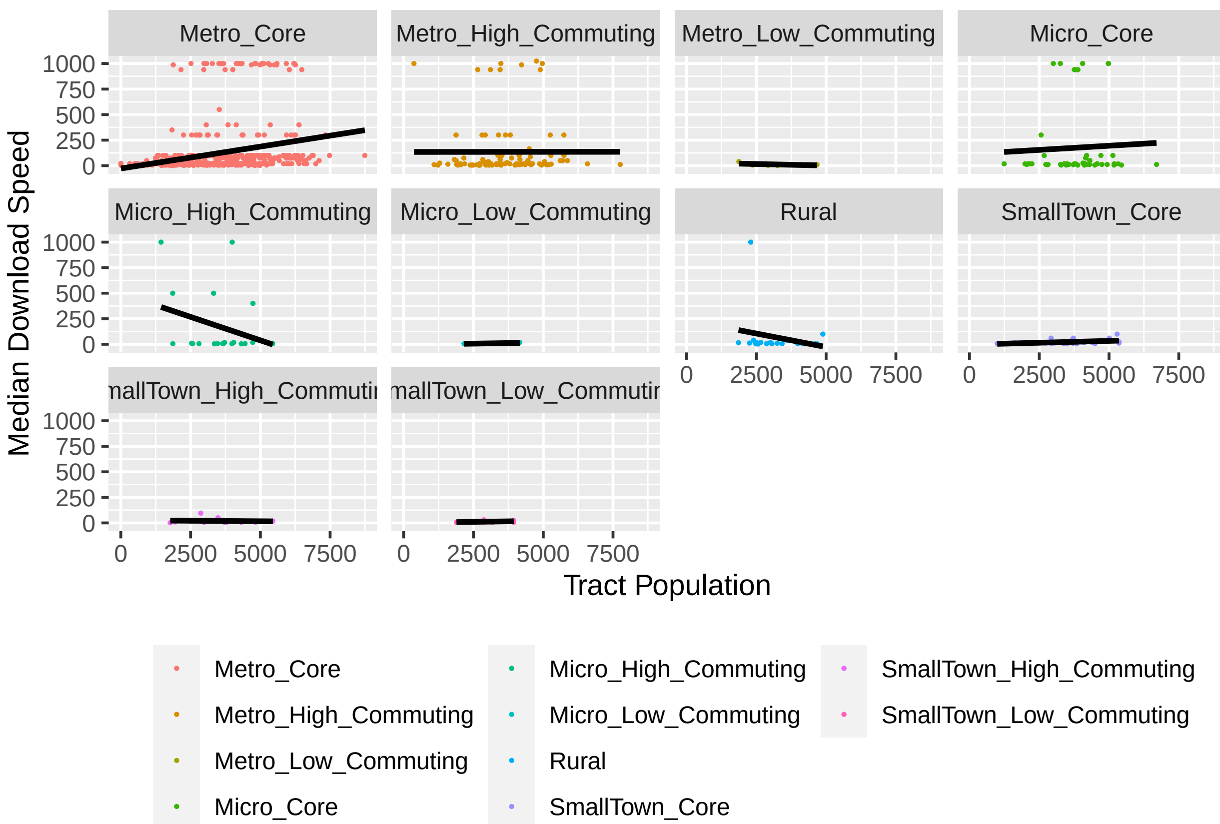
r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.residual	nobs
0.0837466	0.0592836	233.1422	3.42339	0.0001438	11	-2907.042	5840.085	5892.731	22394370	412	424

.metric	.estimator	.estimate
rmse	standard	246.9855

The predictor most related to the outcome is tract population and is the only predictor with a p-value less than .05 representing its significance. The regression model predicts that for every 100 unit increase in tract population, the median advertised download speed increases by 4 mbps, controlling for all other independent variables. The R squared of the overall multiple regression model accounted for 8% of the model variance.

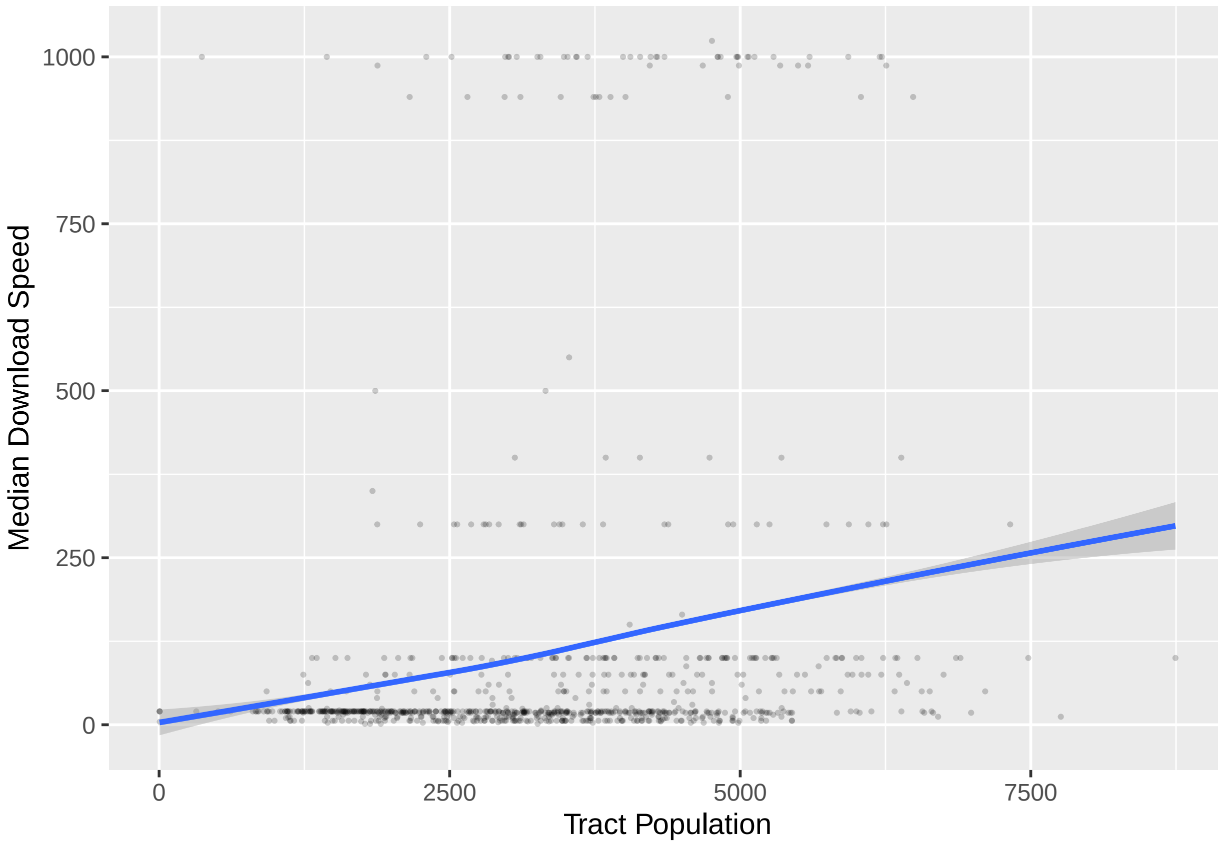
The root mean squared error (RMSE) represents the square root of the variance of the residuals, which provides insights to the average distance between observed values and predicted values for the dependent variable: median advertised download speed. The actual value of 247 is subjective but more useful when comparing other models to determine the better predictor model. However, this is a large RMSE which could be due to the large spectrum of advertised download speeds that ranges from 1 mbps to over 1000 mbps.

Graphic Visualization of Tract Population in each RUCA



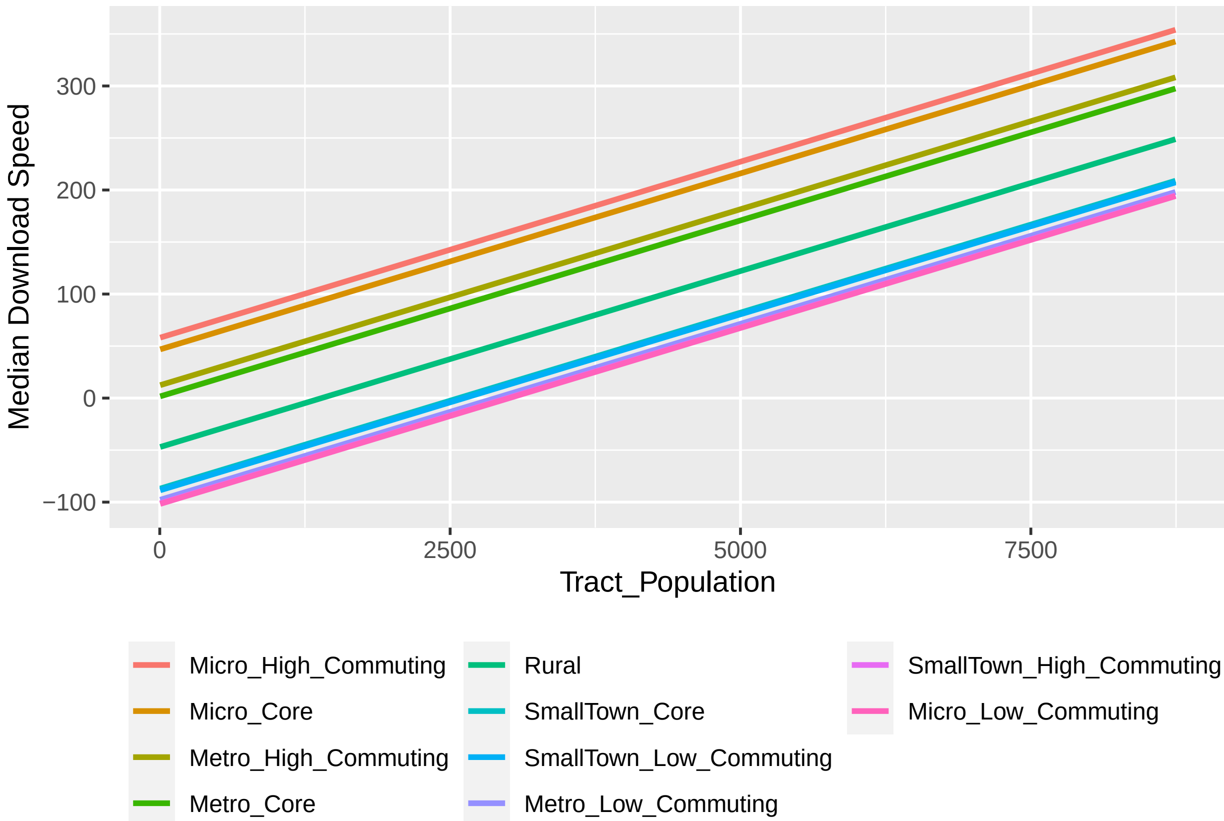
The majority of RUCAs have a flat relationship between median download speed and the population as categorized by RUCA. The metro core tracts show a positive relationship wherein population increases, so does median download speed. The micro high commuting tracts show a negative relationship wherein as population increases, median download speeds of the areas decrease. One explanation for this may be that providers have not invested in infrastructure because most people are working outside of the area and therefore might not need strong internet at home. However, it is important to note that this data is pre-pandemic. Now that more companies are shifting to work from home models, this classification area will likely need significant updates.

Graphic Visualization of Linear Regression Model



From the beginning we recognized that median download speed was not normally distributed and in fact there were quite a few outliers. The model shown here shows that in general as tract population increases, so does median download speed. Within each RUCA there did not seem to be significant variance between household median incomes. As such, we decided to look more deeply at tract population to understand the relationship to median download speeds.

Graphic Visualization of Linear Regression by RUCA



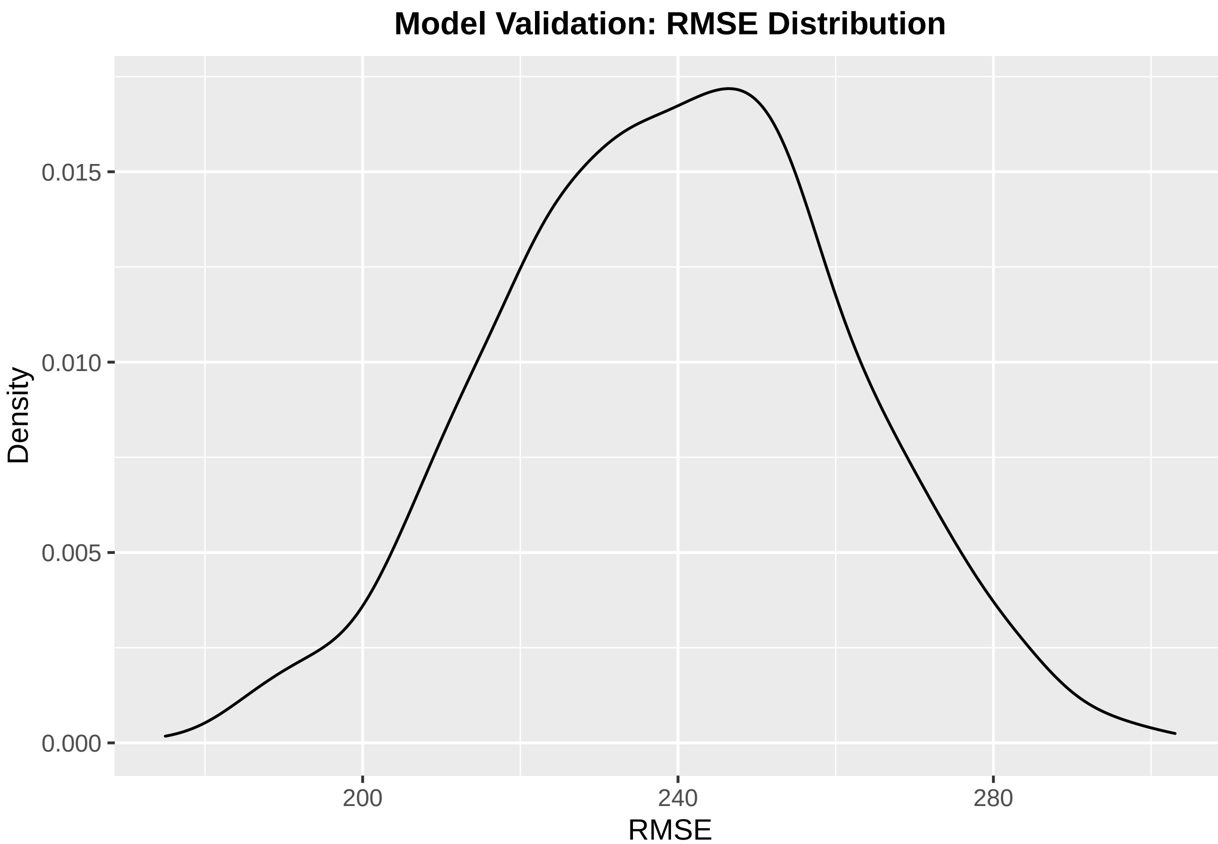
The final graph predicts median download speed for each RUCA category with tract population as the independent variable. It is important to note download speed cannot be negative, so the negative speed values for the corresponding tract population per RUCA are not applicable. Based on the linear model, Micro High Commuting has the fastest median download speed at each tract population values, whereas Micro Low Commuting has the slowest median download speeds at each tract population value.

Linear Model Validation and Quality: RMSE

This model is validated via a monte carlo simulation method with 1000 randomly selected samples. The table below displays the RMSE across the resampled models. Our mean RMSE from the 1000 "fits" is 239, which is lower than the RMSE from our regression model, though still not highly predictive.

.metric	mean	std_err
rmse	239.4224930	0.6905585
rsq	0.0480472	0.0006877

The distribution of RMSEs reflects a fairly normal distribution pattern, however, it is distributed over quite a wide range (~180-300). This shows that the model does not fit consistently to "new data", further validating that it does not predict broadband speed reliably given the factors chosen for the model.



Key Findings, Limitations, and Next Steps

This analysis examined potential relationships between RUCA codes of US Census tracts and median broadband download speeds in the state of Louisiana. Additional relationships explored include population density, percent Black and percent Hispanic, and median household income. Overall, findings revealed these relationships are not necessarily predictive of median download speed. Additionally, univariate analysis indicates much of the state of Louisiana continues to be underserved or unserved according to industry standards despite previous funding and legislation. Although the literature suggests a connection between access to broadband and race, this is not necessary the case when considering broadband speed in Louisiana. There did however, appear to be a relationship between commuting populations and broadband speeds. A high commuting population always had higher broadband speeds than low commuting populations.

When considering the overall analysis the presence of outliers was a limiting factor. The root mean squared error is sensitive to the high outliers in the positive skew. The chosen data points including advertised instead of actual download speeds also restricts the scope of the analysis. RUCA as a data point is also a limitation due to the fact the codes were generated from Census data from 2010. An additional limiting factor is the external validity of the study. The data in this project cannot be generalized beyond the state of Louisiana.

Future studies may investigate ways to focus funding in the most equitable method in areas that are currently unserved. This may benefit from an investigation into additional predictive variables such as the number of providers available in a tract, total dollars spent by providers, and average commuting time. Law makers in the state of Louisiana should reevaluate policies targeting only rural areas to better serve all citizens within the state.

References

Broadband USA (2021, November). *Indicators of Broadband Need.* Broadband USA Maps. https://broadbandusa.maps.arcgis.com/apps/webappviewer/index.html?id=e2b4907376b548f892672ef6afbc0da5

Cooper, T. & Tanberk, J. (2021, November 5). *Best and worst states for internet coverage, prices, and speeds.* BroadbandNow Research. https://broadbandnow.com/research/best-states-with-internet-coverage-and-speed

De, R., Pandey, N., & Pal, A., (2020, June 9). *Impact of Digital Surge During COVID-19 Pandemic: A Viewpoint on Research and Practice.* Elsevier Public Health Emergency Collection. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7280123/

Farrish, K. (2020, August 4). *Health Care and Education Suffer When There's No Internet Access.* Connecticut Health I-Team. http://c-hit.org/2020/08/14/health-care-and-education-suffer-when-theres-no-internet-access/

Rivkin-Fish, (2021, April 29). *Is the FCC's reverse auction fatally wounded or just bloodied?.* Benton Institute for Broadband & Society. https://www.benton.org/blog/fcc%E2%80%99s-reverse-auction-fatally-wounded-or-just-bloodied