

DECONSTRUCTING THE DIGITAL DIVIDE: THE GEOGRAPHY,  
DEMOGRAPHY, AND SPATIAL DEPENDENCE OF INTERNET STABILITY IN  
THE US

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Peter Cody Fiduccia

August 2022

© 2022 Peter Cody Fiduccia

DECONSTRUCTING THE DIGITAL DIVIDE: THE GEOGRAPHY,  
DEMOGRAPHY, AND SPATIAL DEPENDENCE OF INTERNET STABILITY IN  
THE US

Peter Cody Fiduccia, Ph.D.

Cornell University 2022

Internet access and connectivity has become a crucial issue of public policy across the globe. During the COVID-19 pandemic, as individuals and households transitioned to remote work and learning, usage of and strain on home networks increased dramatically. The ability to interact with the internet is quickly becoming recognized worldwide as a determinant of social, economic, and even physiological well-being. With the continuing increase in usage of telehealth, remote work & learning, and distance collaboration tools, the importance of internet access is underwritten by assumptions regarding the internet's stability of connection. A broadly accepted metric to ascertain two-way video & audio stability is known as latency. Being able to empirically and visually describe the geographic distribution of latency across spatial units is of critical importance to understanding where potential policy interventions or government assistance programs are most needed. Similarly, understanding the spatial landscape of latency reveals inequities between socioeconomic, racial, and regional populations. In order to create the most nuanced, empirically sound predictive models to understand factors that influence latency, local regression techniques must be brought to bear. In this paper, I combine a rigorous exploration of the literature with a variety of empirical tools to solve these challenging issues by examining latency across all census tracts in the United States. Quantitative techniques included in this examination are traditional

univariate, bivariate, and multivariable statistical methods, cartographic transformations, exploratory spatial data analysis, autocorrelation analyses, spatial demographic methods, local regression modeling, geographic interpolation, and kriging. I find that rural census tracts, and tracts with higher poverty rates, particularly those with populations other than non-Hispanic White, experience poorer internet stability. I provide identifiable visualizations for where latency is at its best and worst. I classify and specifically identify typologies of neighborhoods to explicitly show discrete groups of census tracts where policymakers can plan interventions. Finally, I present kriging as a methodological tool to predict previously unknown values of latency in order to better fill in the gaps of coverage areas and stability measurements.

## BIOGRAPHICAL SKETCH

Peter Cody Fiduccia was born and raised in Orange County, New York and attended Warwick Valley High School. A musician and actor by training, Peter believed a career in the arts would be the next logical step. Opting for a healthy sense of pragmatism regarding future career security, he enrolled in a Bachelor's degree in Business Administration program at Hartwick College in Oneonta, New York. A natural sense of curiosity for how organizations function, and subsequently how successful leaders operationalize their vision, led him to complete a Master's degree in Business Administration at Binghamton University School of Management focusing on organizational behavior. Peter then took a one-year hiatus from academia, during which time he worked in food & beverage at the Statler Hotel on Cornell University's campus. Gaining a new appreciation for the intricate workings of a complex, higher-education institution, Peter desired to understand the public-facing side of academia, and successfully completed a Master's degree in Public Administration at Cornell University's Institute for Public Affairs. Seeking to reach the pinnacle of novel research, Peter's lifelong dream was realized when he was accepted into a doctoral program in 2017. His time in academia has allowed him to have the great fortune of working with, and learning from, communities, individuals, and colleagues from all walks of life and corners of the globe. Peter moves onward from the academic community to life's next chapter with humility and gratitude for the many experiences and lessons that have been afforded him along the way.

Dedicated to my parents: for your unwavering love and encouragement throughout my journey and beyond. Your passion for discovering strange, new worlds inspires me to deeply explore our planet and its inhabitants; your long list of impressive, industry-defining achievements fuels my creativity at work; your selfless support constantly reminds me of the importance of family; your humor through all circumstances prompts me to never take myself too seriously; and your example for creating a home filled with warmth, support, and positivity excites me to build one of my own. With all the love I can muster - thank you. This achievement is as much yours as mine.

## ACKNOWLEDGMENTS

This project would not have been possible without the tireless support of my dissertation chair, Dr. John W. Sipple, and my dissertation committee members Dr. Peter Rich and Dr. Jeff Rzeszotarski. Thank you for continually pushing me to expand my areas of conceptual and empirical skill, and for being interested in all my questions along the way. To my brothers: without each of you, the stumbles along my path would have certainly resulted in fall damage from which I could not recover. Thank you for a lifetime of memories, jokes, side-splitting happy hours, one-in-a-millions, nail-biting final circles, that part when the witch came out, congratulatory hugs, contemplative nights staring at the embers, when data was there, catnip, teamwork, PS<sup>F</sup>C, and silly songs in rugbies. iO. To my personal and professional mentors: thank you for pushing me to expand my thoughts, my abilities, and my areas of understanding to achieve this, my highest of dreams. To Learning: thank you for bringing me new visions of humility, discovery, and connection. To Love: thank you for affirmation, allowing me to go easy on myself, and for jumping and ducking at the same time. To Life: thank you for reminding me that time is not a predator that stalks us all our lives, but rather a companion that walks with us along the way reminding us to cherish every moment – for it will never come again. Finally, to all those who have touched my life along the way: *Live Long and Prosper.*

## TABLE OF CONTENTS

Biographical Sketch.....	III
Dedication.....	IV
Acknowledgements.....	V
Preface.....	XI
Chapter 1: Life through Internet Lag.....	1
Chapter 2: Decoding the Digital Divide.....	27
Chapter 3: Spatial Dependence of Internet Stability in the US.....	56
Epilogue.....	84
Technical Appendix.....	87
Empirical Appendices.....	88
Bibliography.....	97

## PREFACE

Our world continues to become increasingly interconnected in a myriad of ways. We have the technology to instantly speak with another human being continents away. Advances in engineering provide transportation between previously unreachable landscapes. Using satellite technology, we can connect communities, survey land masses, and predict weather patterns across the planet. This plethora of advancement also allows us, as researchers, the tantalizing opportunity to explore the influence of technology on a host of social factors and outcomes. To be critical, in ideally the most constructive way possible, of the variability in power, agency, impact, and disparate access that comes along with rapid socio technical advancements.

Despite its meteoric rise to prominence in the 1990s, and its ubiquitous use since then, the internet has received quite a bit of attention over the past several years, mainly as a way by which we can connect, evaluate, and explore previously underserved neighborhoods, nations, and populations. Most recently, during the COVID-19 pandemic, we saw examples of the attention garnered by the World Wide Web: Starlink satellites have brought internet to the most remote areas of the world, and we all became a bit more tech savvy trying to figure out why working from home was causing our home networks to slow down. Finally, the pandemic caused an industry-changing shift to working from home, as well as a rare shift from in-person to remote education and healthcare services. This transition to nearly full-time living & working at home meant that more members of households were connecting to the internet. Even if a household was single occupancy, they were connecting to a network that, due to location, population density, or other infrastructure factors, was not designed to have so many people, in total, using it at once.

Across a wide cross section of communities and geographies, I began to hear frustrated stories from friends and colleagues about dropped conference calls, slow

downloads, choppy audio in meetings, or the dreaded ‘buffering circle’ while streaming a favorite show. I also heard anecdotes about young students having difficulty staying engaged while attending ‘Zoom school’ because of the vastly different learning environment that exists when trying to learn from home. Being connected to educators, administrators, and parents, in the K-12 education space, I began to think about the similarities between what we already know to be detrimental to a child’s learning experience and how the switch to remote learning caused by the pandemic might add another difficulty to the learning mix. I thought of a sociology of education course that compared the experience of a student from a higher socioeconomic status family to that of a lower SES student; the examples are many, but the story is nearly always the same. A student from a lower SES family may be more likely to live in a setting where space and privacy are at a premium, and distractions, other obligations, or the necessity to work to support the family are higher. If students, regardless of their SES, may already be having difficulties concentrating and engaging with their work after school, having already had the benefit of being in a dedicated classroom, this does not bode well for students having to spend the entire school day connecting via the internet.

This thought led me to start exploring the landscape of available data on internet connectivity. What followed was an informative and challenging journey through the sociological literature, public data repositories, technology journals, and industry meetings. I quickly realized that to assess the entire breadth of items related to the internet would be too tall a task, and one that would ultimately be too broad to be useful for policymakers. I did, however, end up coming across a dataset that would narrow the trajectory to what eventually became this project. In assessing the Speedtest.net dataset from the private company Ookla, which will be detailed in all chapters of this work, I was fascinated by the latency metric. As this project explains in detail, latency is a measure of stability with regards to two-way communication. When you engage in a

two-way interaction via the internet, whether that's a Zoom conference call, a Teams meeting, or a FaceTime on Wi-Fi, the amount of 'lag' you experience can be measured in milliseconds. If there is no lag, latency is extremely low, and you likely will never notice a problem or become frustrated with your technology; I consider this 'stable internet.' Lag as low as 15ms, however, begins to interrupt the smoothness of your two-way communication and can result in lip sync issues, dropped audio, or the dreaded 'reconnecting' status message; I consider this 'unstable internet.' The higher the value in milliseconds, the longer it takes for information to 'ping' between the two parties or devices.

What was most interesting about latency, was that so few people who were having trouble with their internet stability understood that latency was the problem - not bandwidth. Bandwidth, which dictates average upload and download speeds, is a measure of one-way communication. Either a device is sending information to a server, uploading, or gathering information from a server, downloading. When it takes a longer time than usual to download a file, or to send an email with a large attachment, these are bandwidth issues, not latency issues. The distinction is that latency only focuses on concurrent, two-way connections where both parties are sending and receiving information, while bandwidth is a one-way connection that either sends or receives but does not simultaneously interact with another party or parties. When I would hear anecdotal stories about students having trouble maintaining a connection to their remote classroom, or individuals having to stop their video just so they could have an audio connection without drops, these were all examples of high latency values - poor internet stability. The combination of incorrect nomenclature, little knowledge about the differences in stability and bandwidth, and the dire consequences to remote learners and workers if latency is high enough to cause significant disruptions, led me to explore latency nationwide.

This project, through its many iterations, set out to create the first nationwide empirical evaluation of internet stability. Though the dataset used for this project spans the entire globe, I chose to only focus on the United States. When I first examined the Ookla data, my questions were broad and spanned the realms of telehealth, network service provision, and even the utilization of Thiessen polygons to update and revise hospital catchment zones. As with most projects, however, these grand ideas quickly began to narrow as the gaps between what is known in the literature and what is critically important to explore for timely policymaking became more evident.

Chapter one began as an exploration of the entire Ookla dataset, within the US, with the goal of mapping the distribution of various internet-related variables by census tract. In addition to latency information, the Ookla dataset also measures upload & download speeds, the average number of devices within the survey area, and the number of times a speed test was run over the survey period. These other variables were of great interest, but as the sociological underpinnings of why stable internet is so important began to become clearer, latency took center stage. All other variables were not considered for analysis, and the chapter ended up focusing entirely on the spatial distribution of latency. In attempting to craft a story across all three chapters, this one focuses on where latency is better or worse, specifically with regard to urban and rural locales. Exploring the ‘where’ in social science often brings up the element of ‘to whom.’ People and places are inextricably tied together and, as such, one exploration quickly follows the other in tandem. During the course of analyses related to chapter one, it was clear there were differential effects within those who experience varying levels of internet stability. This motivated a shift in chapter two, and chapter one produced an empirically rigorous, yet broad, evaluation of the geographic distribution of internet stability in the US. In it, I found that 71% of the rural population in the US has poor internet stability. Furthermore, rural census tracts overall are much more likely

to have high latency, and living in any other region than the Northeast puts populations at a disadvantage for stable internet.

Chapter two was originally a selection focused entirely on telehealth access and hospital service provision. While working on another project related to rural healthcare, I was interested in the shift from in-person to remote healthcare due to the pandemic. Namely, the differential impacts this shift had on those in rural locales; rural residents often must drive longer distances to reach healthcare which places them at a disadvantage already. Some may think, then, that the pandemic would lessen the burden on rural populations as they could take healthcare visits from home without having to drive. This is, in fact, not true as rural communities in the US are often far less likely to have internet access and, even if they do, the technology is outdated--usually DSL lines or satellite instead of broadband or fiber--and access to a PC or other device at home occurs at lower rates. After the exploration of chapter one, I also confirmed the prima facie hypothesis that rural areas have less stable internet, making them more prone to disconnection and 'Zoom fatigue' after trying to remain engaged with a choppy connection for long periods of time. This is still an area of great interest, but the question surrounding 'who' has better or worse stability was still left unanswered from chapter one, leading to a shift in the focus of chapter two. Chapter two removed the exploration of telehealth services and instead focused entirely on detailed racial, poverty, and school-age population evaluations of latency across all census tracts. Using data from the US Census, I was able to evaluate latency across specific racial groups--more detailed than those traditionally covered by broader analyses--to find differential effects of latency on non-Hispanic White, non-Hispanic Black, non-Hispanic Asian, non-Hispanic American Indian & Other, and Hispanic/Latino populations. In addition, I wanted to explore the sub-effects of poverty and percent school-age population per tract. This would not only paint a more accurate racial picture, but one that included the

nuance of wealth and proportion of the tract population that were ‘going to school.’ Using clustered robust OLS models as well as marginal predictions, I showed with the highest levels of statistical significance that non-Hispanic White populations tend to experience the best stability across region and rurality. Furthermore, all races other than non-Hispanic White experience worse internet stability as poverty rates and the levels of school-age children increase.

Chapter three, from the very start, was to be a methodological exploration of latency using spatial analyses and local regression techniques. Many of my previous projects, including those co-authored with department faculty, focus on bringing the spatial context to bear. Place matters, and it directly interacts with relationships, markets, outcomes, accessibility, and distribution of resources. In all my work, I try to make the naturally occurring geospatial relationships between people, place, and policy front and center. At the least, it allows us to gain valuable perspective on where certain populations experience sociological phenomena. At the best, it allows us to create more nuanced, methodologically appropriate estimation models by accounting for the fact that ‘everything is related, but closer things are more related.’ This, the first law of geography by Waldo Tobler, governs the examination of relationships in space and calls for us to use, when conceptually appropriate, local regression modeling to mitigate spatial autocorrelation. A core assumption of linear regression is independence of observation; spatial regression allows us to maintain that independence of observation across space instead of only the empirical sample itself. Chapter three presents such an exploration using spatial error modeling to examine latency across all census tracts. Using spatial regression was initially planned for chapter two when it included telehealth analyses but was moved to chapter three when telehealth was tabled. Threads of chapter two did make it to chapter three, however, in the form of exploring how to impute unknown values. When chapter two was examining hospital catchment zones, I

thought of using Ookla data to see where it would be most likely, due to being far from a provider, someone might use telehealth. Naturally, no dataset is perfect and the Ookla data does not have universal coverage across all land areas of the US. It is irresponsible to assume that areas with missing data are unpopulated, so that data would have to be estimated. This necessity of estimation carried over to chapter three when I created the spatial weighting matrix for local evaluation of latency. If there is an unknown value, that tract may be designated as null, changing the composition of the neighborhood. If we can impute the unknown values with some degree of precision, we can both construct a more complete neighborhood while simultaneously providing an estimated 'landscape' of the particular variable in question. Such a technique for estimation originated in soil & natural sciences and is called point interpolation. Using the spatial relationships between points, we can estimate unknown values of variables at all locations between the known points. I use this technique, also known as kriging, to produce an estimated landscape of latency across certain test regions in the US. Chapter three verifies that latency is spatially autocorrelated across US census tracts, but such autocorrelation can be accounted for by using spatial error modeling, thereby explaining 10% more variance in latency. I also provide an example of kriging to give future researchers a baseline social science test case. The estimation landscape that was created was methodologically accurate but must be refined in order to be as refined as I believe it should be to be used as a tool for public policymaking in social science.

In the end, the three chapters of this work represent a contribution to the literatures of educational sociology, geospatial science, public policy in social science, telecommunications, and rural development. This project is a concentrated and targeted set of analyses that evaluate internet stability across geographies, racial groups, ruralities, and populations, and ends with a methodological framework for the estimation of unknown values in social science data. It is my most sincere hope that this

project can be used as a guide for understanding latency across the US, as a means by which policymakers can identify and implement interventions to those communities most in need, and as a first step toward broader adoption of spatial methodologies in everyday social science research.

## **Chapter 1: Life through Internet Lag**

### *The Geography of Latency in the United States*

#### **Introduction**

The COVID-19 pandemic continues to reveal new information about persistent, systemic disparities in our society. As the world transitioned to remote work & learning, and more individuals connected to their home internet networks, those networks became more susceptible to increased disruptions, instability, and lag. Employees and employers were forced to quickly expand their remote capabilities to account for the switch to asynchronous work. Simultaneously, a new instructional environment began to take shape, one in which teachers connect to students via the internet and the daily classroom experience is delivered via Zoom. In both cases, workplace efficiency and daily learning rely on stable video and audio transmission. As the potential for dropped connections and disruption increases, the switch from in-person to remote environments presents unique challenges for employees and students alike. While most attention toward internet connectivity is focused on access, upload and download speeds, internet latency, which measures the stability of two-way connections, represents a new, urgent measure of inequality in our society. Through a detailed geographic analysis of all continental US Census Tracts, this paper visualizes the structural inequalities of internet connectivity and finds that rural populations are disadvantaged with regards to quality internet connectivity. High latency, which results in unstable, disrupted connections, is most often found in areas of non-urban populations: 71% of the rural population in the US experiences poor stability. Rural areas with higher proportions of school-age populations experience worse stability than their urban counterparts, and those living in the South experience worse stability than the other three major regions of the country, with the Northeast being the most stable region by measure of latency. The portrait of internet stability presented here is

crucial to understanding future policy implications of access equality for remote employment and education.

This paper has three primary goals. First, to lay the groundwork, I review the literature to explore the history of computer-aided communication, internet accessibility, geography in relation to socioeconomic status, and the rural-urban disparity of access to digital resources. Second, I provide an empirical backdrop of the current distribution of latency across the entirety of the United States, by census tract. Third, I execute bivariate and multivariate analyses to ascertain directional relationships between elements of latency and geography. Finally, I discuss the implications of the geographic distribution of latency in the US and the important questions surrounding latency brought to the forefront by this foundational analysis.

### **Background & Literature Review**

#### *Access to the internet by geography: the 'digital divide'*

Since the late 1990s, scholars have studied various components of the internet, World Wide Web, and associated technicalities (Leiner et al. 2009; Rachfal and Gilroy 2019). Over the past decade, scholars have begun to examine more specific elements of digital connectivity as it relates to public policy (Skerratt 2010), socio-politics & socio-economics (DiMaggio et al. 2001; Khatiwada and Pigg 2010; Li and Ranieri 2013; Selwyn 2015; Stern and Wellman 2010; White and Selwyn 2013), and the myriad of access types & services available (Blank and Groselj 2015; White and Selwyn 2013). Across fields, the disparities in the distribution of Information and Communication Technology (ICT) systems is known as the 'digital divide,' with the earliest analyses in the 1990s (McConnaughey, Lader, and Chin 1998), and having recently been defined as "...the gap between individuals, households, businesses and geographic areas at different socio-economic levels..." by the OECD (Li and Ranieri 2013). Recent studies in the UK have noted the bifurcation in the digital divide

literature being that between the study of the socio-economic implications and the dedicated analysis of built infrastructure as it relates to the internet (Philip et al. 2017). Furthermore, within the infrastructure literature, the term *digital exclusion* has been used to describe areas and populations that experience either extreme delays or total non-adoption of ICTs “...through circumstances beyond [their] immediate control...” (Warren 2007). Rural locales have long been acknowledged as lagging behind their non-rural counterparts with regard to development of access, making non-urban places in the US at highest risk of being digitally excluded (Downes and Greenstein 2002). We also understand that usage patterns among rural and non-rural locales differ (Sanders 1998).

*Why does latency matter above and beyond internet speed?*

Much of the literature, and commercial advertisement, focuses on whether a household has access to the internet and, if so, the upload and download speeds of that connection (Hargittai 2004; Malone 2001). These upload and download speeds are what is known as bandwidth; this term is often described as ‘how big the pipe is,’ speaking to how much ‘internet can fit through.’ However, the other literature on human-computer interaction, organizational theory, and applied management have all documented the importance of interactive communication in order to coordinate work and resolve ambiguity (Bass and Avolio 1994; Monge et al. 1985). While bandwidth is certainly important for downloads and quality of video streaming [one-way videos such as YouTube], I choose to focus on latency which describes the time it takes for information to be sent from an individual’s computer to the nearest server [and possibly beyond] and back again. Latency is described colloquially as ‘ping,’ and this terminology can be linked to its namesake exemplar: sonar systems. When an active sonar system sends out an audio signal, the time it takes to ‘ping’ off a target and travel back to the host vessel is measured in seconds. Similarly, the time it takes for

your home computer to send information, that information to be received by the nearest hub or server, and information to be received again by your computer is measured in milliseconds. The longer this travel time takes, the more interruption, lag, or perception of ‘slow’ two-way communication you experience. Opposite to bandwidth, latency directly influences the quality of two-way internet interactions such as dyadic conferencing, multi-party sessions, audio/video chat rooms, and online classroom learning. It is crucial to understand and empirically measure and map the variability of latency to specifically understand the nuances of ‘digital deserts,’ where they exist, and what groups are differentially impacted.

### *Stability of Communication*

Even at the dawn of computer-aided communication, scholars stressed that video and audio conferencing should be as lifelike as possible, and that “...it should function like face-to-face communication” (Fish et al. 1992; Kiesler, Siegel, and McGuire 1984; Short, Williams, and Christie 1976). Specifically, Human-Computer Interaction [HCI] and Computer Science [CS] research surrounding the software platforms necessary for multi-party conferencing--an example of which would be Skype or Zoom--cite the need for smooth connections that minimize delays and gaps in order to provide users with a meaningful, engaged experience (Cutler et al. 2002; Junuzovic et al. 2011; Yamashita et al. 2013). Most recently, scholars and politicians alike have advocated for access to the internet to be considered a guaranteed right as a form of active citizenship, rather than a privilege afforded to certain groups (Reglitz 2020; Tully 2014). This question of rights, of course, is a complicated, multi-faceted question that I do not explore in this paper.

### *The influence of COVID-19 on Remote Work & Learning*

Internet access in rural locales has been an issue on the forefront of policy professionals for the past decade, and researchers have also focused on the disparities

of access as it relates to high-speed internet (Drake et al. 2019; Whitacre and Mills 2007, 2010). The historical implementation of technology access in rural areas was achieved through the building of a nearly full-coverage landline telephone network across the US (Belinfante 2001). In recent decades, this coverage also allowed for rural areas to access the internet through DSL, but this internet was far slower than the technology available to urban customers, by comparison (Lin and Liou 1988). The Federal Communication Commission [FCC] has stated “Accessing the internet has become a prerequisite to full and meaningful participation in society” (Kruger 2014; Rachfal and Gilroy 2019). Regardless of policy statement or historical underpinning, the reality is that rural areas are less profitable for broadband providers and more costly for municipal governments to service; leading to a significant difference between the speed and quality of internet connection between geographic areas (Hollman, Obermier, and Burger 2021). This, again, places non-urban regions and their populations at increased risk for disruption and disconnection during periods of remote work & learning.

#### *Local Analyses: The Spatial Context*

The literature surrounding place and education reinforces my supposition that geography matters: where a family lives and, subsequently, where a child attends school can have a profound and varied impact on educational attainment (Crowder and South 2011; Pasculli et al. 2008). Research indicates that neighborhood context, including poverty rates, educational attainment, and family composition can contribute to increasing socioeconomic segregation, have direct effects on childhood development, and can be a determinant factor in outcomes (Owens, Reardon, and Jencks 2016). Bischoff & Reardon illustrate the example of two children in socioeconomically disparate neighborhoods, and this relates directly to our discussions on social / cultural capital (Reardon and Bischoff 2011). This variation between areas within a community

or school district is not a newly researched concept. Some aggregate literature on education, policy, and geography, others relate spatial theories, access, equity, and the educational differences between rural and urban districts, while others still speak to the importance of communities' understanding the geographical makeup of their own locale, as well as surrounding areas, to make informed policy decisions on neighborhood planning and educational resources (Gulson and Symes 2007; Hoglebe and Tate 2012). This analysis sets the stage for the importance of spatial contexts in detailed multivariate work. Though I do not execute spatial regressions in this paper, the underlying conceptual understanding of space and place is central to illustrating the variance of latency across the US. It is important to explain why, particularly with such a robust national dataset, we would want to first explore the scope of, and then later account for, local variation.

#### *How geography intersects with internet access and SES*

Nearly 15% of counties in the US have one-quarter of their working-age populations classified as low educational attainment (Telford 2019). Out of these low-education counties, 75% of them are rural. Moreover, their geographic distinctiveness underscores the importance of the spatial context analytical operations. Over 70% of the low-education counties are in the South region, and nearly 80% have had high poverty rates for three or more decades (Smith and Trevelyan 2019). Nearly three-quarters of these rural, Southern, low education counties are locales in which populations are at least 20% African American or Hispanic (US Census Bureau 2020). We already understand that rural communities, regardless of demographic makeup, are at a disadvantage for having stable internet. We also understand that educational achievement of non-white populations, regardless of geography, lags behind White and Asian populations by up to two grade levels (Reardon et al. 2019). This body of previous work indicates that, on average, rural locales are more likely to experience

educational disparities, lagging behind their urban counterparts. Additionally, the ‘digital divide’ makes clear the internet access challenge also faced by rural areas, exacerbating existing inequalities. Finally, for those rural areas with high minority populations--particularly in the US South--the internet and education access inequality gap is particularly true. All told, the case for an investigation of internet access as a driver of education inequality is salient: particularly in the ‘Zoom era,’ where students and teachers are using remote learning as the primary mode of instruction.

Apart from studies done in Europe over the past year, there is little research on the effect of remote learning on working and schooling outcomes (García and Weiss 2020). Moreover, of the few studies which examine internet access, proliferation, and/or development, the data available from the federal government [FCC] are not granular enough and must be aggregated to the state level (Lehr et al. 2006). This paper is one of the first empirical studies to critically examine the distribution of internet latency by geography and demography at the Census Tract level across the entire United States. It is also an important empirical foundation on which to build further analyses with additional datasets. The newest research examined actual speeds in the homes of participants using custom-built hardware units in combination with the Ookla speed test platform in order to measure specific latencies & metrics on-site at participants’ homes (Hollman et al. 2021). Though this is, by far, the most granular examination to date of latency, the sample size is not large enough to make broader analyses effective at the national scale. I build on this research by using the full Ookla dataset to critically examine geography and urbanicity.

### **Data and Research Design**

The research question underpinning the project focuses on latency and is divided as follows. First: *are populations living in rural areas disadvantaged by disproportionately worse internet connections?* The null hypothesis posits there is no

difference whether one lives in a rural or urban census tract on average latency. I estimate rural tracts will have worse latency [H1]. Second: *are populations of school-age children disadvantaged by worse internet stability?* The null hypothesis posits that average latency will be the same regardless of age across the tract's population. I estimate there may be no discernable global effect, but that rurality will interact with the school-age population, to the detriment of rural tracts [H2]. Third: *Is there a region or state effect that disadvantages populations toward worse latency and connectivity?* The null hypothesis posits there will be no difference between regions of the US or state-level effects with regard to average latency. I estimate there will be a geographical effect by region or by state [H3].

This framework positions latency as a 'disamenity,' and by measuring the disamenity precisely, I am able to discuss rates of higher exposure and where they occur. Most importantly, the paper visually identifies tracts, states, and regions of the country that, as a result of high latencies, exhibit higher risk for missed connections, lags, and drops in audio/video, all of which can negatively influence the experience and progress of remote workers and learners.

### Ookla Speed Test

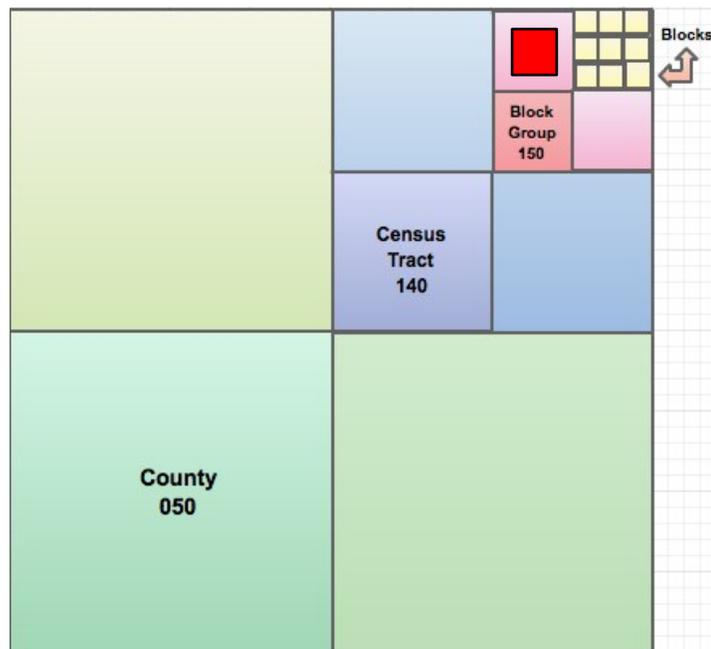
National-level data describing internet speeds, access, and/or connectivity are sparse. ISP data is not ideal because there is an inherent conflict of interest as some internet service providers [ISPs] will present data from their own networks that only paints them in a favorable light. Independent information gathered from state agencies does exist, but the levels of analysis or collection methodologies are often disparate, which would make uniting the data for a national analysis inefficient. I use a novel dataset from the technology company Ookla. Ookla's services include the widely utilized Internet Speed Test. This test, run through an internet browser on the user's desktop, tablet, or mobile device, measures upload & download speeds, as well as latency to the

device executing the test. The dataset also measures how many tests are executed, as well as how many unique devices exist, within the collection zone. The data are presented in survey areas at the 1x1km level, in squares or ‘tiles.’ The files are provided open-source and free to the public as part of the company’s program Ookla for Good. They are updated on a semi-annual basis, but the timeline for publication is not known. Future analyses using longitudinal data may gain use from downloading on a set timeline, yearly.

Creating the Dataset: Combining Ookla and Census Data

While the data covers the entire globe, I focus on the continental United States in this analysis. Raw Ookla data is presented in the 1x1km tiles. Within the dataset, translated to a shapefile using a spatial analysis software program [QGIS], are the unique IDs, as well as values for upload, download, latency, # of devices, and # of tests. The raw data was parsed to the continental US [excluding Alaska, Hawaii, and Puerto Rico]. In total, there are ~1.6M tiles across the country. Using the state of Delaware as a test case for unit of analysis, I identified that 98% of the tract polygons contained internet speed data. Though counties

were at 100% coverage, the sample size of using just counties would be too low. In addition, the number of tiles needed to cover an entire county would call for a level of aggregation that would result in too much aggregation. The image on the right shows the



Examples of Census Boundary Units

approximate scale of 1x1 tiles (red square) and standard Census coverage units. Conversely, block groups only had a 15% coverage rate--due to many of the groups being below 1x1 km in area--which would have resulted in undesirable overlap issues. Therefore, I decided on tracts rather than groups, block groups, or counties. First, I created centroids for each Ookla tile and then spatially joined them to the Census tract that contains them. This approach enabled me to construct a novel tract-level dataset with information about internet latency that could be merged with Census and ACS data on population characteristics. This procedure entailed some instances of imperfect overlap<sup>1</sup> between spatial units, but fortunately did not result in much data loss. Of the 72,043 total populated census tracts of the continental U.S., 71,093 [99%] contained one or more of the 1.7 million matched Ookla tiles.

This was a major component of the project prior to performing any descriptive or predictive analyses. The final sample size for this project is 71,093 US census tracts. This is after parsing non-populated tracts and tracts without Ookla fixed broadband test information [which, together, totaled less than 0.5% of the entire original sample]. Ookla tiles, which are aggregated to the tract level, total 1.7 million observations for the continental US.

## **Methods**

The goals of this paper are to provide descriptive, evaluative information about the relationship between internet latency, urbanicity and population by age. To accomplish this, I use a range of common methodological tools to characterize these patterns, from t-tests and bivariate correlations to OLS prediction models. I pair descriptive statistics with an exploratory spatial data analysis (ESDA) using maps and formal measures of geographic clustering. A brief, technical methods appendix

---

<sup>1</sup> As with any involved aggregation and/or joining operation, shapefile boundaries are not always precisely aligned - particularly when using data from multiple publishers. In this case, there were edge overlaps & overruns between the Ookla tiles and the tract boundaries. When selecting toolbox operations in QGIS to merge these layers, some tiles were outside the boundaries of tracts and, as such, were excluded from the aggregation.

describes these approaches and their interpretation in more detail. My goal in using these measures is to limit the influence of the modifiable areal unit problem [MAUP] in the measurement of internet latency (Wong 2004) and to explore patterns of local variance. In spatial analyses, the MAUP is a potential bias that can be introduced to a geographically oriented dataset simply by arbitrary alteration of bounding areas or boundary lines; the most famous example of which is gerrymandering. By ‘slicing’ a particular area of population in different, arbitrary ways, one could come up with different combinations of political majorities without any change in the underlying data. By choosing spatial units that strike a balance between overlapping [tiles are larger than blocks] and hyper-aggregation [too many tiles would be needed to cover a county], I mitigate the MAUP as best as possible.

### Measures

The primary dependent outcome is average latency, a continuous variable measured in milliseconds, from the Ookla internet speed data, as described in the introduction above. Average latency is first ascertained from the raw Ookla tiles at the 1 km x 1 km level. The information is then collected and aggregated up to the tract level, creating a tract average latency which is the final measure of stability. As the research questions specifically explore populations, namely rural as compared to urban, all multivariable analyses are weighted by population per tract. The first predictor measure is rurality, a dichotomous indicator variable derived from Census measures of the percentage of a tract with rural population. Any tract greater than 55% rural population as defined by the Census Bureau data is given a value of ‘1’ in the dichotomous classification of ‘rural tract.’ The second predictor measure is percent of school-age children in the tract, a continuous variable determined from Census population data. The third predictor measure is region, a non-ordinal categorical variable denoting which of the four regions of the United States each census tract falls. Interaction effects are

measured between rurality \* percent school-age as well as rurality \* the squared version of percent school-age, to test whether there is a non-linear relationship between percent school-age and the dependent variable mean latency. These measures address the three primary components of my research questions: rural, age, and geographic disparities with regard to latency. Table 1 shows an overview of the primary measure of interest, average latency, across the four US Census regions as well as the number of tracts per region by rural and urban classification.

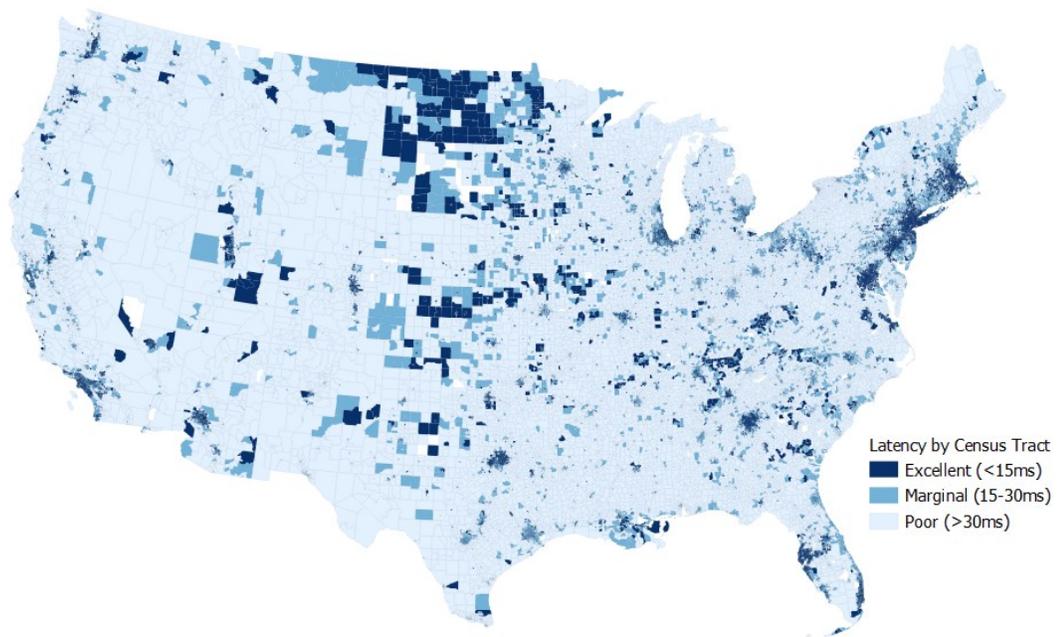
Number of Census Tracts in Sample (Average Latency in Parentheses)					
	NE (20ms)	MW (34ms)	S (38ms)	W (28ms)	Total
Urban	10927	12813	19710	13718	57169
Rural	1932	4053	6273	1666	13924

*Table 1 – Census Tracts by Region*

## **Results**

### *Descriptive Spatial Visualizations*

Figure 1 is a map depicting the distribution of latency across the US, by census tract. Darker areas are lower [more stable] latency, while lighter areas are higher [less stable] latency averages.



*Figure 1 – Average Latencies in the United States by Census Tract*

Though Figure 1 highlights broad trends across the US, viewing census tract data at a national scale introduces the risk of obscuring regional nuances. To overcome this problem of visualization at scale, figure 2 highlights the well-known geographies of Washington, D.C., Virginia, and Delaware areas on the East Coast and highlights the variation of latency between urban and rural tracts. This begins to paint the picture of how populations in different areas of rurality experience varying levels of internet stability.

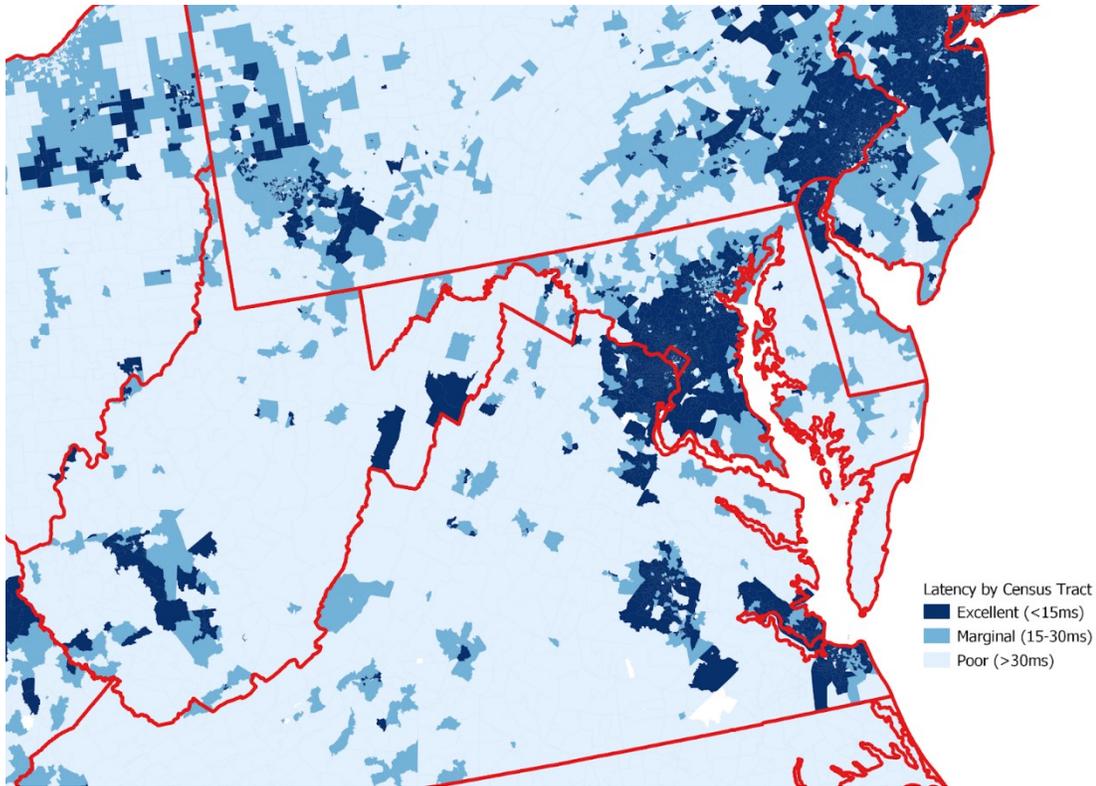


Figure 2 – Average Latencies in Washington D.C., Delaware, and Virginia

Figures 3 through 6 show latency in the US by Census Regions and rurality to highlight the variation of latency across the country. Moreover, these figures clearly illustrate the spatial dichotomies that exist between rural and non-rural locales. The figures show latency via choropleth visualization, with darker colors indicating more stable (lower latency) internet across individual tracts. Urban tract boundaries are highlighted in red, while rural tracts do not have a boundary color. This is meant to starkly illustrate that the vast majority of rural tracts experience poor latency. In the Northeast, Figure 3, we can observe that the major urban centers [Boston, New York, Philadelphia, Newark, Pittsburgh], on average, experience excellent stability - despite some urban tract outliers that have marginal latency figures [such as the cities along the I-90 corridor in upstate New York]. The light blue choropleths are mostly rural tracts and make up the vast majority of the land area in the Northeast.

## Latency by US Census Tract Region 1: Northeast

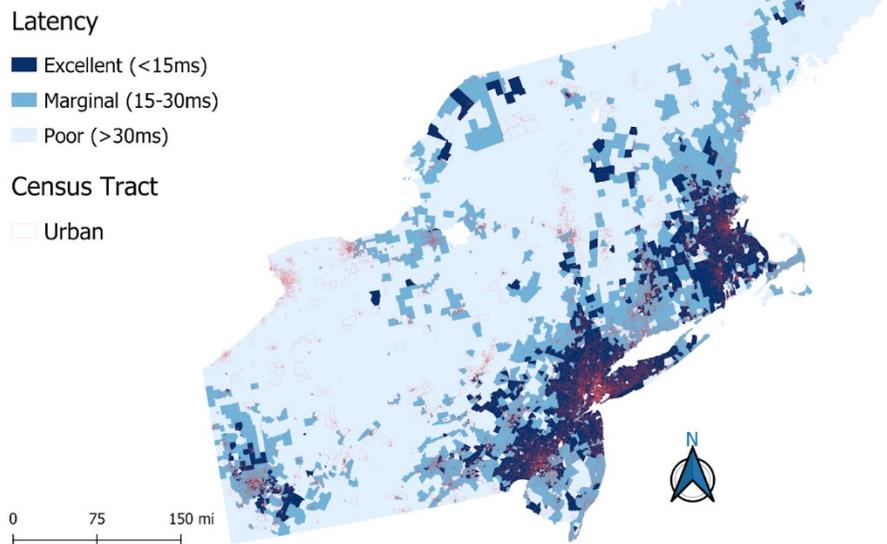


Figure 3

In the Midwest, Figure 4, we again visualize a similar overall pattern but with one major difference: North Dakota [top left portion of the map]. In the late 1990s North Dakota’s telephone cooperatives and local private companies came together to combine resources and invest in fiber broadband internet (Kienbaum, 2020). Fiber-optic networks, though more expensive to install initially than mobile broadband towers, depreciate at a much slower rate and, as such, are far more “future-proof” than traditional cable or DSL [phone] internet lines (Communities, Horrigan, and Satterwhite 2012; Katz, Avila, and Meille 2011). According to the 2020 FCC data, over 80% of North Dakota by land area is connected via fiber networks. This state level investment can be concretely visualized in Figure 4 where the majority of vast, rural census tracts have low latency, extremely stable internet connectivity.

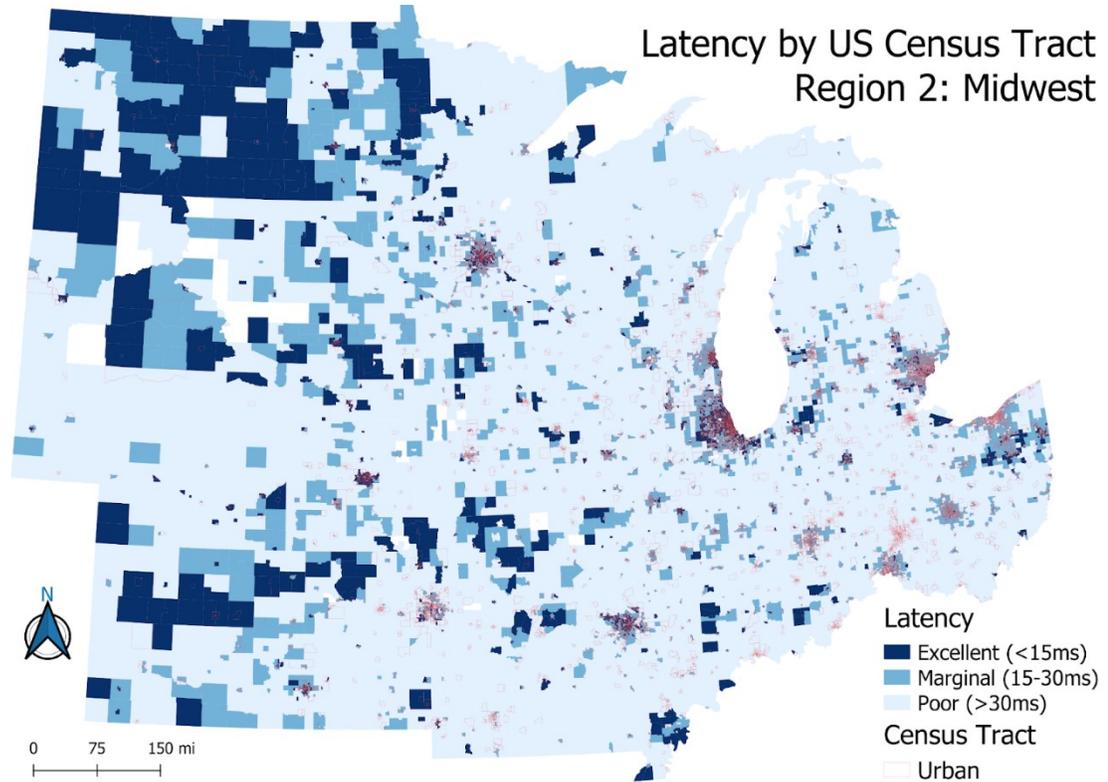


Figure 4

Moving to the South region, we observe the clearest representation of rural communities' disparity regarding internet stability in Figure 5. Though urban centers still return with low latency, high stability internet, the overwhelming majority of rural census tracts in the South are in the lowest category of stability. Even states that are not anecdotally thought of as 'deep south' states, like Maryland, Delaware, or Virginia experience the same difficult disparity between urban and rural locales.

## Latency by US Census Tract Region 3: South

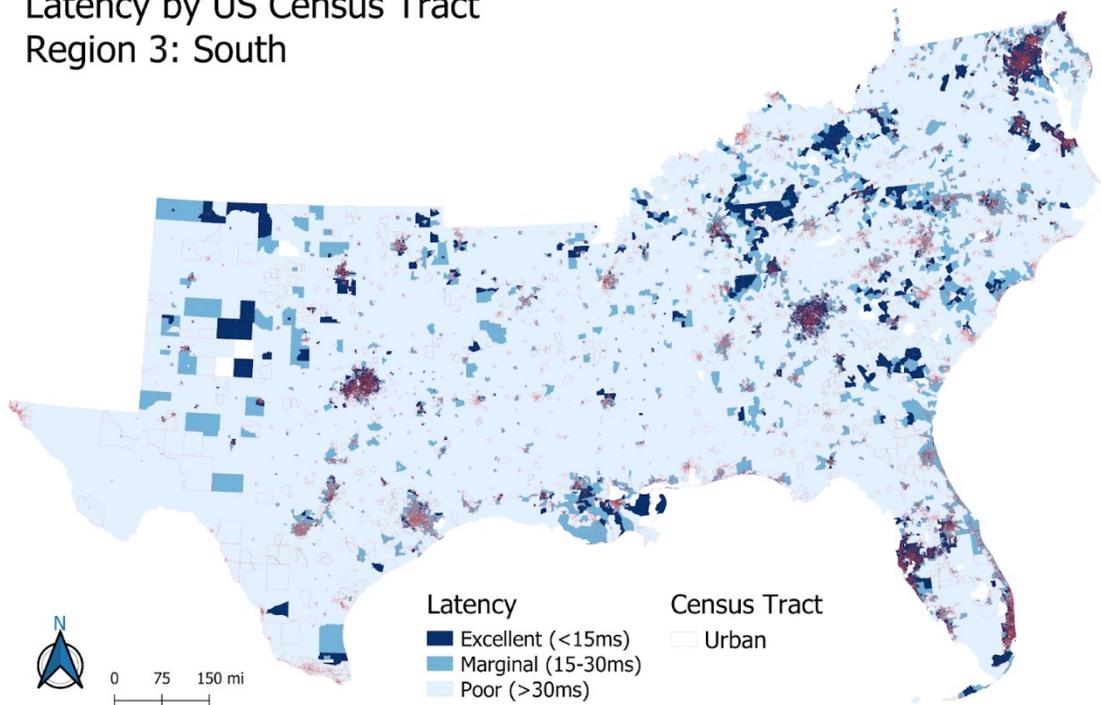


Figure 5

Finally, we observe the West region in Figure 6 and see, once again, a similar pattern emerges - to the detriment of internet stability in rural communities. There are far fewer urban census tracts [red outline] aside from the coastal cities of California and urban centers of Oregon and Washington. What areas of high stability we do see outside of urban centers are primarily located on or near US Military and/or Federally owned and operated sites.

# Latency by US Census Tract Region 4: West

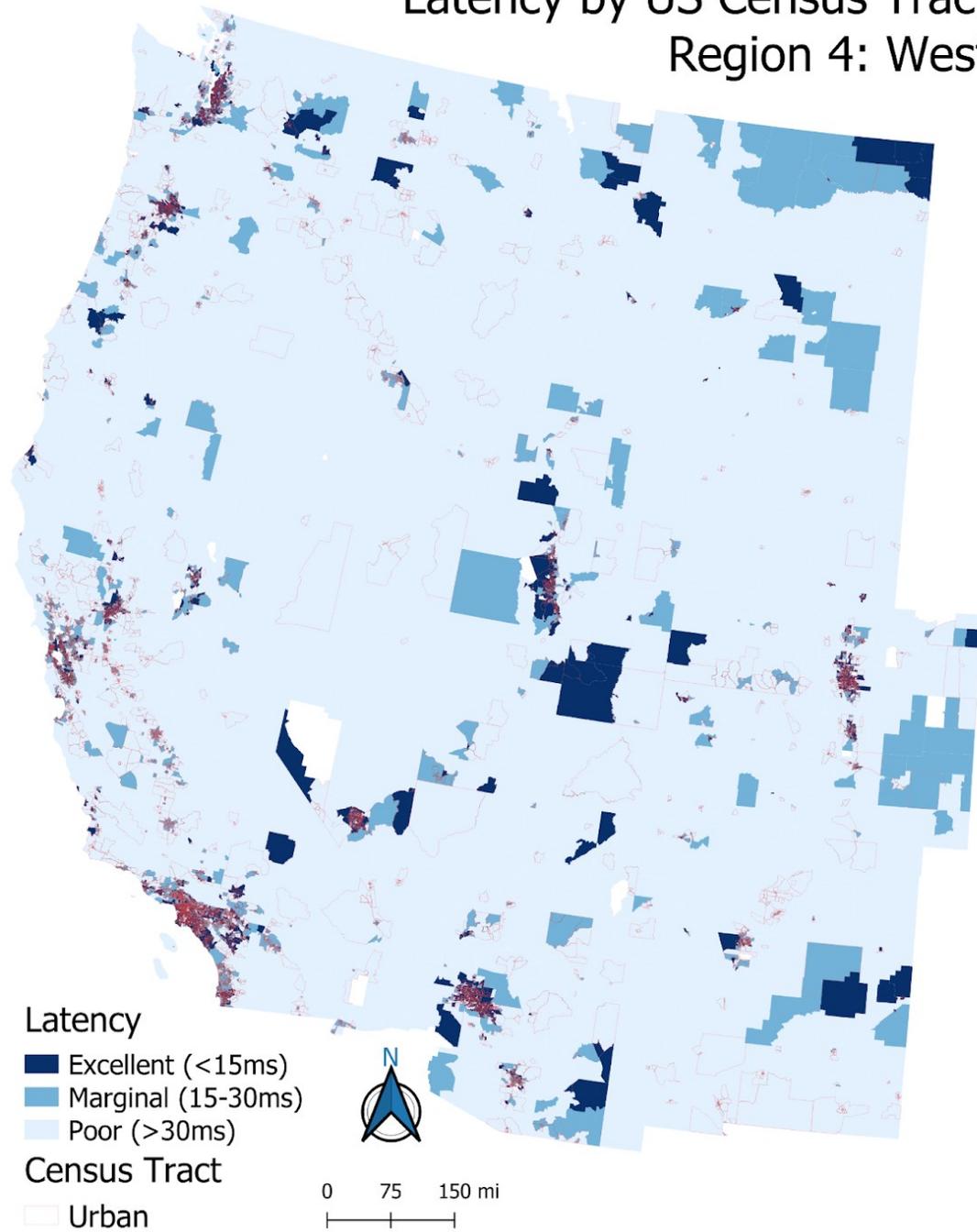


Figure 6

The exploratory spatial data analysis presented in these four figures is not meant to be exhaustive, but powerfully illustrative in setting up the quantitative analyses in the following section. The disparities in urban and rural locales with regard to latency [stability] are crucial to view at scale so that we can better understand and contextualize the empirical information presented in the results section.

### Univariate

The unit of analysis for the dataset is census tracts in the continental United States, with a total sample size of 71,093 tracts. Any tract greater than 55% rural population as defined by the Census Bureau data is given a value of '1' in the dichotomous classification of 'rural tract.' There are 57,169 urban tracts and 13,924 rural tracts representing 76% and 24% of the sample, respectively. The average latency of all tracts in the sample is 28ms. The majority of census tracts fall between 0 and 40ms, some tracts are in the extreme latency ranges of 50-100ms, and a small number of tracts are outliers in the 100+ms range. There are 225 census tracts that do not have any school-age children, representing 0.3% of the sample. On average, census tracts in the United States have a population of 23% school-age children.

### Bivariate

Correlation analyses in Table 2 show a significant, positive relationship between average latency and percent rural population within a tract (+0.51\*\*) [H1]. There is a significant, slightly negative relationship between average latency and percent school-age children within a tract (-0.01\*\*) [H2]. A two-sample t-test for average latency by rurality was performed, indicating a statistically significant variance between rural and non-rural census tracts at 80ms and 20ms, respectively [H1]. Robustness checks were performed, and all returned with p-values of < 0.01. Appendix 1 shows the t-test for rurality. A one-way ANOVA was performed to test for significant variance in means between regions of the country with regard to average latency. This returned

statistically significant, with an average latency for Northeast, Midwest, South, and West returning at 20ms, 34ms, 38ms, and 28ms, respectively [H3]. The statistical appendix contains tables illustrating the ANOVA. Though bivariate analyses do provide insight into both initial hypotheses, the relationships cannot control for the combination or interaction of multiple explanatory factors such as rurality, age composition, and categorical region. Thus, a multivariable framework must be implemented.

	Avg Latency	% Rural
% Rural	+0.51*	
% School-Age	-0.01*	-0.04*

*Table 2 - Correlation Between Primary Variables of Interest*

### Multivariable

Two multivariable regression models were executed: a standard, ordinary least squares model and a cluster-robust model including state fixed-effects. In the OLS model, above and beyond the effect of school age population and region of the country, being in a rural census tract results in an increase of average latency [H1 - Confirmed].

Above and beyond the effect of rurality and region, an increase in school-age population results in an increase in latency [H2 - Insignificant], but this result was not statistically significant. Several interaction terms were also investigated including the relationship between rurality and the squared term of school-age population, as there is a non-linear relationship between average latency and school-age population. This interaction term returned as having a positive coefficient--higher latency--and statistically significant. Holding rurality, school-age population, and interaction effects constant, and using the Northeast as the control region, the effect of being in the

Midwestern, Southern, and Western regions of the US resulted in 8ms, 13ms, and 11ms of increased average latency, respectively [H3 - Confirmed]. Using this structure, 24% of the variance in average latency is explained through the model. Table A-1 shows the results of the standard OLS regression model in the appendix.

Results from the clustered robust, state fixed-effects model are shown in Table 3. Controlling for state effects, above and beyond the effect of tract school-age composition, being in a rural census tract results in, on average, a 108ms increase in latency as compared to urban tracts ( $\beta=108.5$ ;  $SE=3.62$ ) [H1 - Confirmed]. Controlling for state effects, for each percentage increase in school-age population, latency increases by 18.5ms, above and beyond the effect of rurality and interaction effects, and this result is statistically insignificant at the 0.05 level, not the 0.001 level [H2 - Confirmed]. Accounting for state-level effects, the interaction between rurality and tract school-age composition, squared term, again returns a significant, positive relationship [H2 - Confirmed]. Within-state and between-state explanations of variance in mean latency is measured at 22% and 31%, respectively [H3 - Confirmed]. Overall, the model explains 23% of the variance in latency, like the OLS model.

<b>Clustered, FE Model</b>	<b>Effect on Average Latency(ms)</b>		
<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	18.36	16.25 – 20.48	<b>&lt;0.001</b>
Rural Tract	108.5	101.4 – 115.6	<b>&lt;0.001</b>
<i>%school-age</i>	<i>18.52</i>	<i>0.484 – 36.56</i>	<b>0.044</b>
Int: Rural x %school-age <sup>2</sup>	500.48	367 – 633	<b>&lt;0.001</b>
Observations	70,868		
Groups	49		
Within R <sup>2</sup>	0.22		
Between R <sup>2</sup>	0.31		
Overall R <sup>2</sup>	0.24		

*Table 3 - Clustered, Robust State Fixed-Effects OLS Predictive Model*

## **Discussion**

It is important to put these empirical findings in context, and critical to the discussion to note that higher latency is **not** a desirable outcome. Correlation and t-test results confirm my Hypothesis 1 that rural tracts experience worse latency, on average, than urban tracts. The findings reject the null hypothesis that rurality makes no difference with regard to latency. From the bivariate tests we can also clearly understand that increasing rurality puts students and families at a significant disadvantage for poor connection stability, where the average latency is over three times higher in rural census tracts. In practice, this can translate into audio-video misalignment, lip sync issues (Junuzovic et al. 2011), of several seconds and/or video & audio drops of anywhere from two to four times per minute, depending on additional home network

factors beyond latency. It is important to note that getting a precise understanding of how frequently syncing or dropping issues occur is incredibly challenging and these are only estimates based on available information and self-reported user experiences. Multivariable results show us that populations living in rural tracts experience latencies that are nearly four times higher than their urban counterparts, when controlling for region, school-age population, and interaction effects.

The results fall in line with the literature reviewed above regarding built telecommunications infrastructure as well as *prima facie* anecdotal evidence of rural disparity, the latter having been previously unanswered with robust empirical means at the tract level. There is a single, clear takeaway from this study: rural areas are disadvantaged when it comes to reliable, stable internet connections. Moreover, absent population-age and rurality effects, I empirically verify that populations living in any region other than the Northeast experience worse internet stability than their counterparts. This paper is a significant contribution by way of providing an empirical analysis of connection stability [by means of latency]. It not only builds on the literature's elements surrounding infrastructure and the digital divide, but augments these efforts by providing a replicable, easily discernible portrait of internet stability across the entire United States by a granular spatial unit. The findings are further solidified after subsampling robustness checks and accounting for non-linearity in the school-age population predictor variable using a square term in the clustered robust fixed-effect model.

### *The Scope of Results*

I frame the scope of this paper in the introduction but wish to reiterate the importance within the discussion of results. I choose to limit this paper to an empirically robust yet conceptually broad, geographic exploration of latency in the United States to frame a new and urgent technological disamenity - one which is particularly salient during a

time when policy debate is centered around investment in broadband infrastructure. The purpose of this scope is twofold: First and foremost, to allow for latency [and therefore ‘stability’] to be presented in a clear, concise way as a new metric for equity in work and school during the age of remote interaction. Second, because the initial results presented in this paper signal the need for a much more granular, robust empirical investigation of this new metric by more delineated racial groups, additional age structures, and adjustment for clustering and dispersion using spatial regression methodologies. All these considerations I address in a forthcoming companion paper.

### **Limitations**

#### *Ookla Data*

An important limitation to using this dataset is the collection methodologies used by Ookla. This is known, in the information technology and computer science fields, as ‘black box’ collection; specifics on how the data get to the 1x1km tile level are known only to Ookla. I recognize and acknowledge that I undertake this analysis without detailed information on the collection black box. This does raise concerns about aggregation, multiple device re-testing, and data smoothing on the part of Ookla. Given that this is the only large-scale, granular, publicly-accessible dataset on broadband internet speeds, however, it is worth analyzing despite those limitations. More importantly, the information contained within this analysis by way of the Ookla data can and should be regarded as the ‘best case scenario.’ Even considering the limitations presented by the black box phenomenon, we can assume that the data output is not meant to paint service or speed test providers in a poor light. Therefore, it is possible that the results are, in some locations, more unequal than shown here due to the unknown aggregation methods undertaken by Ookla.

### **Looking Forward**

The central result from multivariable analyses is that rural census tracts experience

worse levels of latency than their urban counterparts. The results confirm the hypothesis that rural areas are disadvantaged with higher likelihoods of internet instability while supporting and augmenting the recent studies on the inequality of education access via remote learning (Haeck and Lefebvre 2020; Murat and Bonacini 2020). In addition to the disadvantage already faced by rural tracts, rural populations that have greater proportions of school-age population face an even greater stability disadvantage than their urban counterparts. These results are important because they provide empirical evidence of geographic inequalities with regard to internet stability, building off recent work examining internet access variability over space (Zahnd, Bell, and Larson 2021), and specifically calling attention to school-age learners during a time when education has shifted to remote means. As we continue to work through the potential long-lasting effects of the pandemic, questions about differences in learning progress have already begun to arise. This study can serve as empirical verification that certain ruralities experienced worse internet stability, adding a new, robust measure to explorations of setbacks (Allington and McGill-Franzen 2003; Cooper et al. 1996), workplace conferencing now taking place at home (Chiariotti et al. 2019; Rao et al. 2019), and utilization of internet-capable devices at home (García and Weiss 2020). Furthermore, the results mark the first robust, dedicated analysis of internet stability [latency] at a national scale using a novel, merged and cross walked source. In the new age of remote work and learning, discussions of inequality must include the contexts of access to stable internet connections, particularly in rural locales. I also acknowledge the need for the robust collection of individual / household-level speed & latency measurements so that specific, granular analyses can be performed without the unknown aggregation concerns currently present. Finally, the literature surrounding internet access as a means of digital inequity should more completely reflect latency as a measure of disamenity rather than simply analyzing upload &

download speeds and built infrastructure. I will continue contributing to this essential work with two forthcoming analyses: one expansion of the preliminary geographic exploration focused on detailed racial, age-structure, and sub-regional spatial analyses of latency, as well as a methodological contribution examining geospatial techniques by which researchers can mitigate gaps in data collection imputation through local kriging.

---

## **Chapter 2: Decoding the Digital Divide**

*Racial, Socioeconomic, and Population Inequalities of Internet Stability in the U.S.*

### **Introduction**

As communities across the world continue to grapple with the myriad of economic, social, and cultural impacts of the COVID-19 pandemic, the ‘return to normalcy’ has been signaled in part by students and workers returning to in-person interaction. For students in the United States, as in many other countries, the shift to remote learning was one that was fraught with challenges. These challenges are not, however, experienced equally across households of varying race or wealth. Income inequalities, neighborhood segregation, and wealth disparities exist across racial groups, with White and Asian populations having multiple times the wealth of Black, Hispanic, and/or Native American households. Long has public education, particularly in the United States, been hailed as ‘the great equalizer’ - providing opportunity to those who may not start life in the middle or upper socioeconomic classes to access the resources that increased wealth brings.

When major shocks to that system--like the COVID-19 pandemic--occur, and disrupt the delivery of, or participation in, traditional education, this can further exacerbate a wide variety of extant racial and wealth inequalities. Specifically, when students need to switch from in-person to remote learning, the success of this transfer is entirely precluded on the availability and stability of internet access. Understanding the well-founded link between educational outcomes and future success, already existing pre-pandemic gaps may only be widened due to populations’ differential access to stable internet connections.

The variability in broadband infrastructure across the United States is a well-studied example of disparities between populations and ruralities, known as the digital divide. Existing inequalities in this system, too, will only be made more evident by a

shock such as the COVID-19 pandemic. Communities with less access to information and communication technologies prior to the shift to remote work & learning only face additional challenges to connect and stay connected.

This paper measures latency, a metric of internet stability, across all census tracts in the United States to ascertain which populations are at higher risk of unstable internet. I find that Black, Hispanic, and Native American populations experience less stable internet--measured by latency--than their White and Asian counterparts. Census tracts with higher poverty rates also have worse stability, an effect which is experienced at higher levels by non-White, non-Asian populations. In census tracts with increasing school-age population, across races and ruralities, instability increases as well, above and beyond racial effects.

## **Background & Literature Review**

### *Characterizing the Digital Divide*

The past twenty years of sociological research has seen a growth in the exploration of inequities, exclusion, and differential impacts of digital resources across geographies and groups (Blank & Groselj, 2015; DiMaggio et al., 2001; Hassani, 2006; Khatiwada & Pigg, 2010; Stern & Wellman, 2010; White & Selwyn, 2013). Other fields including telecommunications and geographic sciences have also begun to explore the notion of the digital divide (Helsper, 2012; Howard et al., 2010; Malecki & Demaray, 2003). Across all fields, there is a strong consensus that however the divide is defined it will be a divisive barrier to inclusion: “If societies are today partly, and will in the future be more or less completely, structured around the internet, then the demands of economic efficiency as well as social and political equity require that no social group finds itself excluded from participation” (Sparks, 2013). At its most broad, the digital divide can be defined as the disparity between individuals, households, businesses, communities, or locales with regard to both their access to internet resources as well as

information and communication technologies for a wide range of activities (Srinuan & Bohlin, 2011). More granular levels of differentiation include identifying the built infrastructure that allows for internet connectivity and socioeconomic divergences, both of which are briefly summarized here to give context to the myriad of challenges faced by those communities or populations that may find themselves experiencing one or more elements of the digital divide.

The elements of built infrastructure [or lack thereof] that reinforce the digital divide are complex, but can be summarized through the understanding that rural telecommunications infrastructure lags behind urban areas, resulting in large numbers of households having restricted use to information and communication technologies [ICTs] (Howard et al., 2010; Peters, 1999). This quickly becomes an issue of complex actors & markets, privatization, public-private partnerships, and policy. The UK, and Europe by-and-large, has been far more efficient at adopting new built infrastructure than the US, with the largest state-sponsored program being the Broadband UK [BDUK] initiative (Philip et al., 2017). This comparison is made with a caveat close to my spatial-analyst's-heart: spatial context. The sheer difference in geographic size between the UK and the US, or the US and any European nation [with whom the US is often compared regarding education / social outcomes] underwrites the vastly different challenge faced by policymakers in administering programs. Take the example of New York state, which is ~2x the size of the UK. Over a geographic area roughly half the size, the UK is administering policy using federal-level resources as opposed to state-level resources in New York. This is very much a back of the napkin comparison, but the point remains a worthwhile one: to roll out broadband infrastructure updates across a nation like the US takes significantly more resource, time, and policy planning than a nation like the UK. That isn't to say, however, that US states couldn't learn valuable lessons from nations in Europe.

The other broad-level aspect of the digital divide that is worth exploring is the social and economic effect of disparities in digital inclusion. Blank and Groselj provide the most accessible bifurcation of the phenomenon by asserting three levels of the divide: access alone, methods of access, and the sociological benefits afforded by access (Blank & Groselj, 2015).

#### *Disparities in Access to the Digital Ecosystem*

The notion of ‘digital disparities,’ or the difference in experience with and/or access to electronic resources & instruction is not a new topic of research. From the late 1990s to the early 2000s there was a dedicated effort to reduce the digital divide and, arguably, it was successful in the United States - in the school building (DeBell & Chapman, 2006; Goolsbee & Guryan, 2006). Disparities between race and socioeconomic status regarding computer access in school “...were largely eliminated by 2003” (Clotfelter et al., 2008). At first, a potential solution was to fund schools’ or states’ purchases of hardware [laptops for students], but this did not prove to have a significant effect on learning outcomes when tested (Shapley et al., 2007). Though the proliferation of technology, at home, work, and school, has grown exponentially over the past decade, the perception may be quite different from the empirical reality. Even as recently as 2021, more than 40% of families, particularly working-class, do not have a laptop or desktop, or do not have access to broadband internet (Khilnani et al., 2020). As Goudeau and colleagues add, even with access to a computer and the internet, there are several factors that can create inequalities for students or work-from-home employees across social class: if there is only one shared computer in the household; if there is a space within the household that can be dedicated to working and learning; whether the school or instructor relies on more printed than electronic material (Goudeau et al., 2021). Even teachers’ and administrators’ understanding of the challenges to reliable connectivity faced by students reflects the many challenges

that were exacerbated by the pandemic. In a 2020 US study, only 50% of teachers nationwide were confident that their students had reliable access to internet access at home, with 15% of teachers indicating that “approximately 50 percent” of their students had reliable internet access (Stelitano et al., 2020). There are also significant racial differences within the digital resource ecosystem. While over 80% of White Americans report owning a desktop or laptop, less than 70% of Black and Hispanic adults own such a form of technology, and the self-reported proportions are the same when asked whether the household has access to the internet (Atske & Perrin, 2021). In the same study, 75% of Hispanics report accessing their internet only through their smartphone rather than via a desktop or laptop and lack traditional broadband service at home (Atske & Perrin, 2021). Another interesting disparity that becomes important as we peel back the layers of socioeconomic and racial digital inequities is for what purpose different groups use technology. Families from upper and middle-class SES backgrounds are more likely to use their digital resources [computers, internet, etc.] for work and education, while those from lower SES families are more likely to use digital resources for entertainment, gaming, or social media (Harris et al., 2017). Additionally, students from upper- or middle-class families are more likely to use digital resources for academic activities rather than for leisure activities in working class households (ibid). Finally, social reproduction theory indicates that schooling, and the resources therein, is not univariately administered or experienced across class and culture (Bourdieu & Passeron, 1977). This extends to, and is further constrained by, the move to remote instruction as now the opportunity for socialization, ability to interact with peers, access to extracurricular activities, or even external options like music lessons / sports are all severely curtailed if not unavailable, further exacerbating disparities already experienced during in-person schooling by lower SES families (Lareau, 1987, 2011). These numerous factors reinforce the supposition that families

in lower-income situations, especially those in rural areas [with already variable access to the internet, generally] are at an initially higher risk for experiencing challenges related to remote work & schooling (Blank & Dutton, 2014; Drabenstott, 2001; Philip et al., 2017; Selwyn, 2015; White & Selwyn, 2013).

*Racial Inequities in Educational Outcomes, Wealth & Opportunity*

Beyond the disparities of access to, and uses of, technology faced by households of varying SES, the literature has made apparent the differences in outcomes for racial groups across the US, particularly with regard to schooling. If we are to explore the differentially disruptive influence of unstable internet [and thus, by proxy, a students' ability to remain successfully engaged with course material via remote instruction], we should also understand the existing disparities faced by non-White school-age children. *Note: this paper focuses on non-Hispanic, non-White racial groups as: non-Hispanic Black, non-Hispanic Asian, non-Hispanic American Indian, and Hispanic/Latino.* Through evidence that spans over thirty years, we understand that non-Hispanic Black, non-Hispanic Native American, and Hispanic/Latino students often score below non-Hispanic White and non-Hispanic Asian students on various standardized tests, have lower rates of post-graduate enrollment, take up advanced placement and I.B. classes at a lower rate, graduate from primary school at a lower rate, and are more frequently disciplined using more 'severe' actions (Darity et al., 2009; Francis & de Oliveira, 2019; Hemphill & Vanneman, 2011; Okonofua & Eberhardt, 2015; Shedd, 2015). We also understand this achievement and outcome gap to not only be explained by disparities in socioeconomic status, but similarly linked to the racial variance therein. Students from families or households with higher socioeconomic statuses are more likely to live and learn in 'educationally enriching environments' with better resources and technologies (Diamond, 2006; Shedd, 2015). Furthermore, the presence of a computer in the home is one important factor that is

found to contribute to the average ~62% lower math scores recorded by non-Hispanic Black students as compared to non-Hispanic White students (Clotfelter et al., 2008). As research continues to explore disparities faced by students with regard to remote learning, we must include latency, both in terminology and in understanding, in order to fully grasp the challenges at play; particularly as we start to see the stark realities of racial inequities. During COVID-related school closures in 2019-2020, 10% of non-Hispanic White students were *without access* to online instruction - as compared to 30% of Hispanic/Latino and 40% of non-Hispanic Black students (Dorn et al., 2020).

Household wealth is also an important factor that contributes to the persistent access and opportunity gaps between racial and ethnic groups (Hanks et al., 2018). Wealth of a household can be related to the general access to educational opportunity (Conley & Hamlin, 2009), access to and benefits from informal networks (Hovardas, 2016), ability to invest in educational technology and tools, and providing an insulating effect to students from potential broader level societal and economic shocks. Census data shows the observed disparities between the average non-Hispanic Black, and Hispanic/Latino to the average non-Hispanic White working male at 0.68 to 1 and 0.63 to 1, respectively (Census Bureau, 2004; Proctor et al., 2016). At the household level, this racial disparity persists: the median net worth of a non-Hispanic White household averages greater than \$150,000 while non-Hispanic Black and Hispanic/Latino households' net worth average ~\$17,000 and \$20,000, respectively (McIntosh et al., 2020). The disparity in opportunity and outcomes also extends to neighborhood wealth, the segregation of which has also persisted over time (Faber & Sharkey, 2015; Owens, 2016; Reardon & Bischoff, 2011). We also know these outcomes to be spatially segregated by neighborhood, but this is not a specific element of the literature or research methodology I approach in this paper.

*Latency as a Necessary Metric for Understanding Digital Inequities*

Much of the literature, and commercial advertisement, focuses on whether a household has access to the internet and, if so, the upload and download speeds of that connection (Hargittai, 2004; Malone, 2001). While bandwidth is certainly important for downloads and quality of video streaming [one-way videos such as YouTube], latency, opposite to bandwidth, directly influences the quality of two-way internet interactions such as dyadic conferencing, multi-party sessions, audio/video chat rooms, and online classroom learning. It is crucial to understand and empirically measure and map the variability of latency to specifically understand the nuances of the digital divide, and what groups are differentially impacted. Even at the dawn of computer-aided communication, scholars stressed that video and audio conferencing should be as lifelike as possible, and that “...it should function like face-to-face communication” (Fish et al., 1992; Kiesler et al., 1984; Short et al., 1976). Specifically, Human-Computer Interaction [HCI] and Computer Science [CS] research surrounding the software platforms necessary for multi-party conferencing--an example of which would be Skype or Zoom--cite the need for smooth connections that minimize delays and gaps in order to provide users with a meaningful, engaged experience (Cutler et al., 2002; Junuzovic et al., 2011; Yamashita et al., 2013). Understanding where latency presents in lower or higher values--thus in more or less stable internet connectivity, respectively--is critical to the framing of internet stability as a disamenity, one which differentially impacts populations of varying race, poverty, and age composition (Fiduccia, 2022). Knowing how income and wealth influence outcome, in both education and other societal constructs, the ability for wealthier individuals and households to select for either ruralities or areas with built internet infrastructure, and more stable internet connectivity by proxy, is a persistent and interconnected phenomenon. It should be noted, though, that the COVID-19 pandemic may reveal positive elements to the racially-differential realities of pandemic

restrictions. For example, non-Hispanic Black and Hispanic/Latino workers were disproportionately hit by employment restrictions and were more likely to experience pandemic-related unemployment (Gould et al., n.d.; Williams et al., 2020). Some research hypothesizes that the forced unemployment of these groups may end up resulting in more time able to be spent with their children / students, thereby resulting in potentially better outcomes for non-Hispanic Black and/or Hispanic/Latino students than in a non-pandemic environment (Francis & Weller, 2022).

The interconnected elements of the literature above work to frame the research questions within this study with the goal of understanding the variable influence of internet instability on remote work & learning in the US.

### **Data and Research Design**

The research questions underpinning the project focus on the specific demographic and racial inequities across latencies and is divided as follows. First: *are different racial populations disadvantaged by disproportionately worse internet connections?* The null hypothesis posits there is no difference in average latency between racial groups across US census tracts. I estimate non-White populations will experience worse latency [H1]. Second: *are populations in tracts with higher poverty rates disadvantaged by worse internet stability?* The null hypothesis posits that average latency will be the same regardless of poverty rate across individual tracts. I estimate that tracts with higher poverty rates will experience worse latency, and that this relationship will be different by racial categories [H2]. Third: *are populations in tracts with higher proportions of school-age children disadvantaged by worse internet stability, when accounting for detailed racial population groups?* The null hypothesis posits that average latency will be the same regardless of proportion of school-age population across the tract. I estimate there may be no discernable global effect, but that specific racial subpopulation effects and rurality will interact with the school-age

population, to the detriment of rural tracts and non-White, non-Hispanic population [H3].

This framework expands upon the geographic description of latency as a 'disamenity,' and directly identifies which racial groups and geographies experience worse stability across the United States. The specificity by which I identify the locales at higher risk for missed connections, lags, and drops in audio/video result in a useful geographic metric for policymakers and researchers. The detailed racial analysis also provides additional empirical evidence for disparities between majority and minority populations in the US.

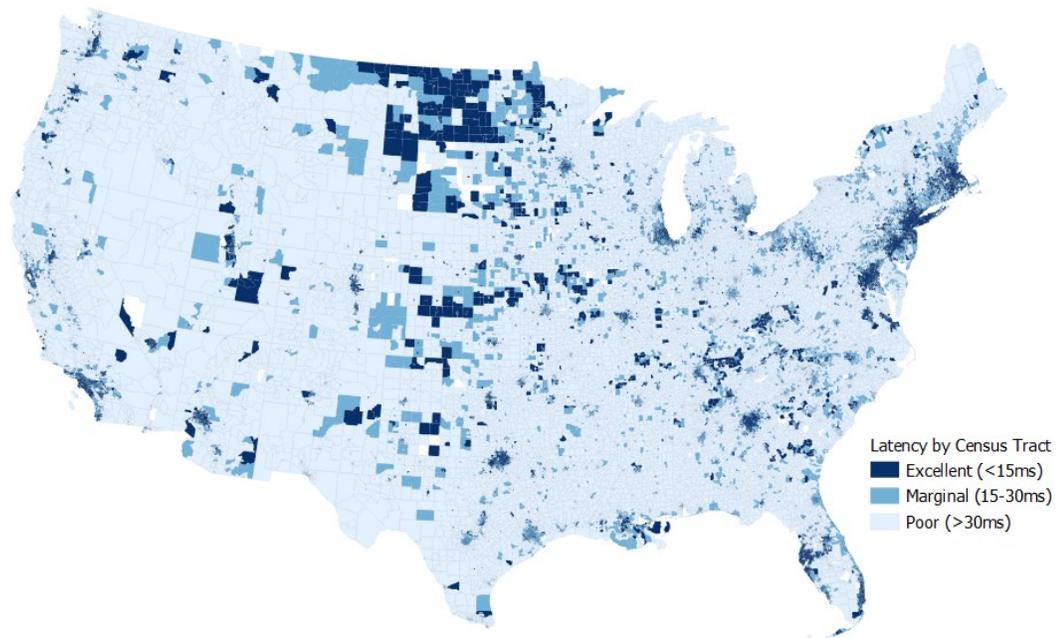
#### *Ookla Speed Test*

National-level data describing internet speeds, access, and/or connectivity are sparse. ISP data is not ideal because there is an inherent conflict of interest as some internet service providers [ISPs] will present data from their own networks that only paints them in a favorable light. Independent information gathered from state agencies does exist, but the levels of analysis or collection methodologies are often disparate, which would make uniting the data for a national analysis inefficient. I use a novel dataset from the technology company Ookla. Ookla's services include the widely utilized Internet Speed Test (<https://www.speedtest.net/>). This test, run through an internet browser on the user's desktop, tablet, or mobile device, measures upload & download speeds, as well as latency to the device executing the test. The dataset also measures how many tests are executed, as well as how many unique devices exist, within the collection zone. The data are presented in survey areas at the 1x1km level, in squares or 'tiles.' The files are provided open-source and free to the public as part of the company's program Ookla for Good (<https://www.ookla.com/ookla-for-good>). They are updated on a semi-annual basis.

### *Creating the Dataset: Combining Ookla and Census Data*

While the data covers the entire globe, I focus on the continental United States in this analysis. Raw Ookla data is presented in the 1x1km tiles. Within the dataset, translated to a shapefile using a spatial analysis software program [QGIS], are the unique IDs, as well as values for upload, download, latency, # of devices, and # of tests. The raw data was parsed to the continental US [excluding Alaska, Hawaii, and Puerto Rico]. In total, there are ~1.6M tiles across the country. Using the state of Delaware as a test case for unit of analysis, I identified that 98% of the tract polygons contained internet speed data. Though counties were at 100% coverage, the aggregation of tiles would be too high. Conversely, block groups only had a 15% coverage rate--due to many of the groups being below 1x1 km in area--which would have resulted in undesirable overlap issues. Therefore, I decided on tracts rather than groups, block groups, or counties. First, I created centroids for each Ookla tile and then spatially joined them to the Census tract that contains them. This approach enabled me to construct a novel tract-level dataset with information about internet latency that could be merged with Census and ACS data on population characteristics. This procedure entailed some instances of imperfect overlap between spatial units, but fortunately did not result in much data loss. Of the 72,043 total populated census tracts of the continental U.S., 71,093 [99%] contained one or more of the 1.7 million matched Ookla tiles.

This was a major component of the project prior to performing any descriptive or predictive analyses. The final sample size for this project is 71,093 US census tracts. This is after parsing non-populated tracts and tracts without Ookla fixed broadband test information [which, together, totaled less than 0.5% of the entire original sample]. Ookla tiles, which are aggregated to the tract level, total 1.7 million



*Figure 7 - Average Latency by US Census Tract*

observations for the continental US. Figure 7 shows the final latency measurements via exploratory spatial data analysis across all census tracts.

### **Methods**

The goals of this paper are to provide detailed empirical evaluations of internet stability across the United States while simultaneously accounting for differences in racial groups, poverty rates, and concentration of school-age population. To accomplish this, I use a range of common methodological tools to characterize these patterns, from t-tests and bivariate correlations to OLS prediction models. I pair descriptive statistics with robust multivariate regression models and post-estimation marginal prediction plotting. A brief, technical methods appendix describes these approaches and their interpretation in more detail.

### **Measures**

The primary dependent outcome is average latency, a continuous variable measured in milliseconds, from the Ookla internet speed data. Average latency is first ascertained

from the raw Ookla tiles at the 1 km x 1 km level. The information is then collected and aggregated up to the tract level, creating a tract average latency which is the final measure of stability. The first predictor measure is rurality, a dichotomous indicator variable derived from Census measures of the percentage of a tract with rural population. The second predictor measure is percent of school-age children in the tract, a continuous variable determined from Census population data. The third predictor measure is poverty rate, a continuous variable constructed by gathering tract-level ‘population in poverty’ data from the 2019 American Community Survey estimates. The fourth measure is region, a non-ordinal categorical variable denoting which of the four regions of the United States each census tract falls. It is important to note that, in this analysis, the region measure is used primarily as a control and robustness check, as previous work has confirmed the relative advantage afforded to tracts located in the Northeast region of the US. Interaction effects are measured between rurality + percent school-age, rurality + the squared version of school-age, and rurality + poverty rate, as there is a non-linear relationship between percent school-age and the dependent variable mean latency. Post-regression marginal estimations on latency are conducted by two predictor variables, poverty rate and percent school-age population within the tract. These estimations are executed and plotted by rurality [urban vs rural tracts], regionality [each of the four census regions], and racial population weights, described in the following section.

#### *Racial Categories & Population Weights*

Previous work provided a detailed, geographic portrait for the incidence of latency across the US by census tract and accounted for variance by region and rurality alone. It is through this work that we understand the danger in assuming access alone is a sufficient enough metric to ascertain reliable connectivity and, furthermore, that there are systematic disadvantages to being located in rural tracts, or tracts outside of the

northeast. The background and literature review of the current study indicates the myriad of well-known and documented reasons as to why the disamenity of latency cannot be treated as equal across racial groups. As such, I define six mutually exclusive categories to account for specific population weights per tract: non-Hispanic American Indian; non-Hispanic Asian; non-Hispanic Black; non-Hispanic White; Hispanic or Latino; non-Hispanic Other. For the regression models and estimated marginal effects + plots I do not include non-Hispanic Other, both for ease of visualization and due to the relatively low sample size within that group. An important consideration to the usage of these racial population groups is that of analytics weights in OLS regression. The use of analytic weights over frequency or sampling weights is chosen because the original Ookla data, in 1x1km tiles, is already presented as an average value of various metrics. Multiple tiles are then aggregated up into the surrounding spatial unit, census tracts. This specifically calls for the use of analytics weights in the OLS regression models due to the product-of-means construction of the primary dependent variable of interest, latency. As an additional method of estimating predictive effects on latency, including racial categories, region, rurality, and either poverty rate or percent school-age population in a single model, I run a scaled [+1x for each racial subgroup], clustered-robust OLS regression model with population-based analytic weights; one with poverty rate and one with percent school-age population as included indicator variables. The full versions of each of these regressions can be found in the technical appendix.

## **Results**

### *Univariate*

The unit of analysis for the dataset is census tracts in the continental United States, with a total sample size of 71,093 tracts. Any tract greater than 55% rural population as defined by the Census Bureau data is given a value of '1' in the dichotomous

classification of 'rural tract.' There are 57,169 urban tracts and 13,924 rural tracts representing 76% and 24% of the sample, respectively. The average latency of all tracts in the sample is 28ms. The majority of census tracts fall between 0 and 40ms, some tracts are in the extreme latency ranges of 50-100ms, and a small number of tracts are outliers in the 100+ms range. Urban tracts, when analytically weighted by total tract population, have an average latency of 20ms. Rural tracts, on average, have a latency of 76ms. Single-variable differences by racial analytic weights return the following results: the mean latency for an average non-Hispanic American Indian in an urban tract is 29ms, and 114ms in a rural tract. The mean latency for an average non-Hispanic Asian in an urban tract is 16ms, and 62ms in a rural tract. The mean latency for an average non-Hispanic Black in an urban tract is 20ms, and 102ms in a rural tract. The mean latency for an average non-Hispanic White in an urban tract is 21ms, and 72ms in a rural tract. The mean latency for an average Hispanic or Latino in an urban tract is 20ms, and 80ms in a rural tract. There are 225 census tracts that do not have any school-age children, representing 0.3% of the sample. On average, census tracts in the United States have a population of 23% school-age children. There are 199 census tracts that do not have a calculable poverty rate, representing 0.2% of the sample. On average, the poverty rate for all census tracts in the US is 15%. Figure 8 shows the initial broad-level relationship between rurality, race, and latency; the interpretation of which leads to the necessity of a more detailed, multivariable analysis.

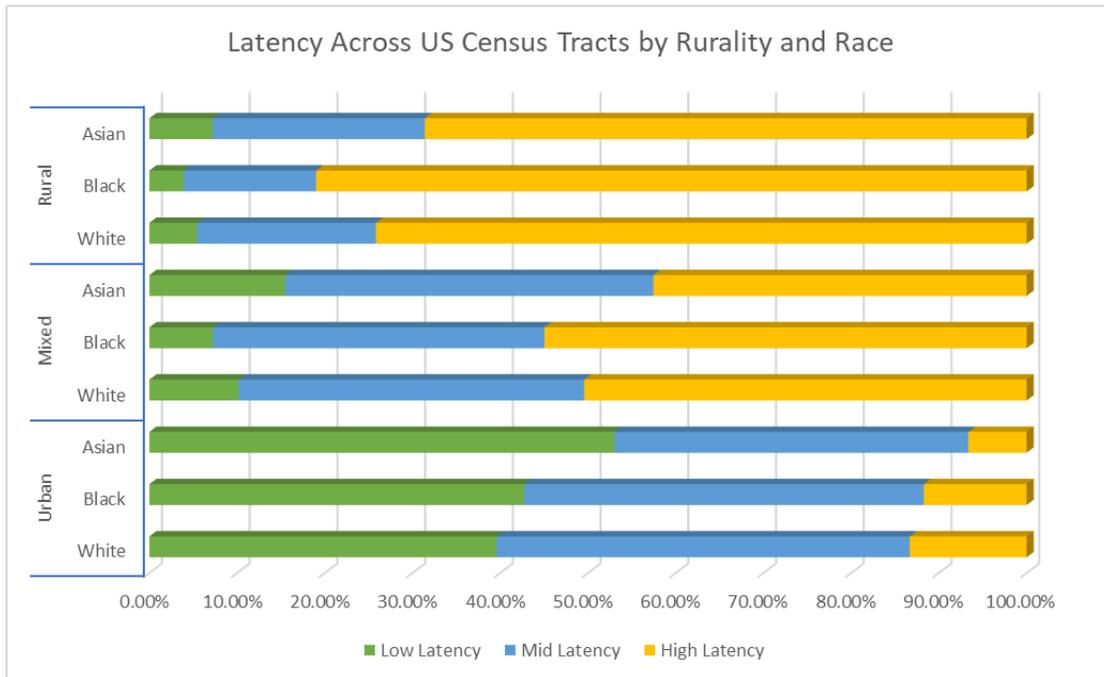


Figure 8 - Latency by Rurality and Broad Racial Categories

Bivariate

Correlation analyses were performed between three main variables of interest: average latency, poverty rate, and percent school-age population by urban/rural category. Across all racial a-weights, significance and direction were constant: In urban tracts, significant, positive relationships were found between all three variables. In rural tracts, a significant positive relationship was found between latency and poverty rate, while a significant negative relationship was found between latency and percent school-age population. Weighted by total population, a significant, slightly positive relationship between average latency and poverty rate in urban tracts (+0.094\*\*) and a significant, positive relationship in rural tracts (+0.21\*\*) is observed [H2]. By tract rurality, weighted by total population, I find a significant, slightly positive relationship between average latency and percent school-age children in urban tracts (+0.067\*\*) and a significant, slightly negative relationship in rural tracts (-0.049\*\*) [H3]. Though bivariate analyses provide insight into two hypotheses, bivariate tests alone cannot ascertain the detailed

racial disparities within latency [H1], and I cannot control for the combination of or interaction with multiple explanatory factors including rurality, racial population weights, and region. Thus, a multivariable framework must be implemented. Table 4 and Table 5 show correlation tables for urban and rural tracts, respectively, for the relationships observed above.

Urban Tracts N=57,169	Avg Latency	Poverty Rate
Avg Latency		
Poverty Rate	+0.094**	
% School-Age	+0.067**	+0.172**

*Table 4 – Correlation Between Latency, Poverty, and % School-Age Population in Urban Tracts*

Rural Tracts N = 13,924	Avg Latency	Poverty Rate
Avg Latency		
Poverty Rate	+0.210**	
% School-Age	-0.049**	+0.160**

*Table 5 - Correlation Between Latency, Poverty, and % School-Age Population in Rural Tracts*

### Multivariable

To approach the question posed by H1, the results of separate multivariable OLS regressions, with analytical weights of each primary racial population, are presented. Table 6 displays the results of these regression models. Across all census tracts in the US, above and beyond the effect of rurality, poverty rate, or region, the average Non-

Hispanic White experiences a latency of 6ms higher than all other races; the average non-Hispanic Asian experiences a latency of 12ms higher than all other races; the average non-Hispanic Black experiences a latency of 11.5ms higher than all other races; and the average Hispanic-Latino experiences a latency of 10.5ms higher than all other races.

Race-Weights	NH-AmInd	NH-Asian	NH-Black	NH-White	Hisp/Latin
% School-Age	49.90***	9.144***	8.057***	5.013*	33.04***
Rural Indicator	57.63***	107.0***	190.7***	77.23***	160.6***
Rur x Sch-Age <sup>2</sup>	-627.2***	544.0***	986.6***	12.67	1,022***
Midwest	-14.29***	6.211***	6.653***	11.48***	5.165***
South	3.383**	4.294***	8.803***	13.87***	10.09***
West	38.28***	4.390***	3.645***	11.22***	7.600***
Constant	2.680	10.86***	11.91***	10.50***	4.945***
<i>Observations</i>	<i>68,412</i>	<i>69,781</i>	<i>70,489</i>	<i>70,857</i>	<i>70,834</i>
<i>R-squared</i>	<i>0.207</i>	<i>0.216</i>	<i>0.297</i>	<i>0.253</i>	<i>0.269</i>

Table 6 – Weighted OLS Regressions Predicting Latency w/ Region, Rurality, and School-Age Population

As stated in the methods section, two variants of OLS regression models are run to ascertain the combination of effects that rurality, race, region, poverty, and school-age population may have on latency. I report here on the scaled, clustered-robust OLS regression model with population-based analytic weights: one including interaction with poverty rate and one including interaction with percent school-age

population. I report the most germane findings here, while full regression tables can be found in the technical appendix. Predicting mean latency in relation to tract-level poverty rate and percent school-age population, both models explain approximately 27% of the variance in latency overall. Rural census tracts experience, on average, latency of +49ms higher than urban tracts. Each unit increase in poverty rate results in, on average, a +34ms increase in latency, above and beyond racial and interaction effects. Each percentage point increase in school-age children within a tract results in, on average, a +40ms increase in latency, above and beyond racial and interaction effects. Interacting rural census tracts and poverty rate, above and beyond racial effects, returns a positive, statistically significant result with regard to latency. Interacting rural census tracts and percent school-age population, above and beyond racial effects, does not return a statistically significant result.

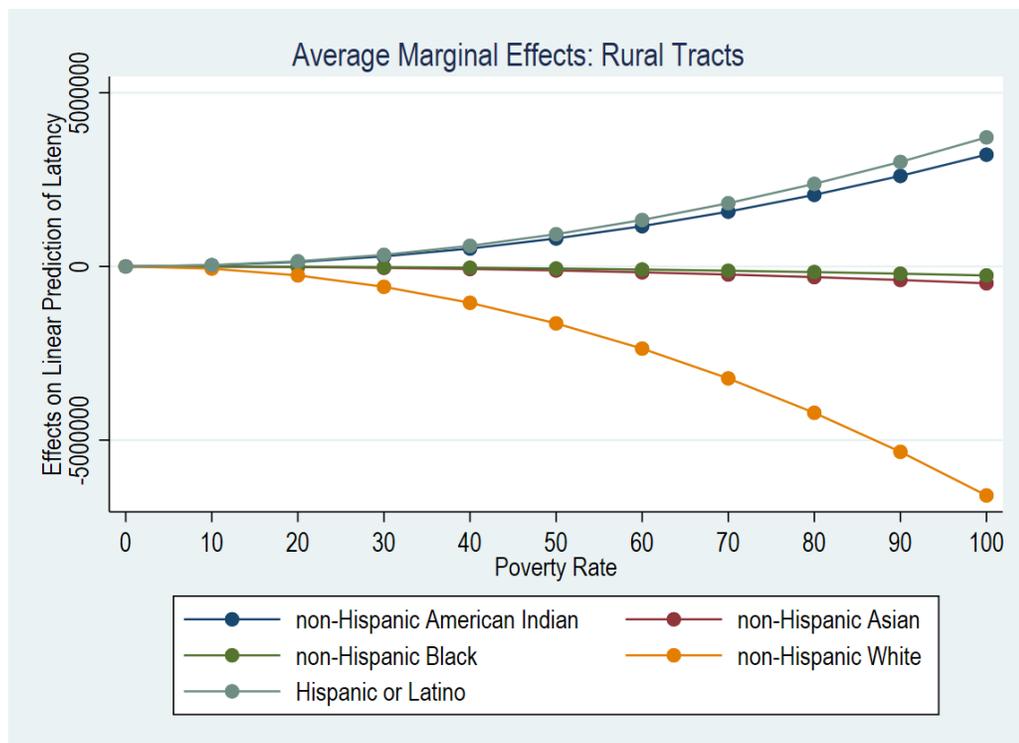


Figure 9 - Marginal Effects on Avg. Latency, Rural Tracts by Poverty Rate

Beyond individual effects and dyadic interactions, results of the model can be interpreted using marginal effects on the linear prediction of latency over separate increases in poverty rate and percent school-age children. In rural tracts, as poverty rate increases, non-Hispanic American Indian and Hispanic/Latino racial groups are expected to experience worse latency than non-Hispanic Asian and non-Hispanic Black populations, who themselves are expected to experience latencies that remain relatively constant across poverty rates. Non-Hispanic White populations, conversely, as poverty rate increases, are expected to experience better latencies as compared to all other racial categories. Figure 9 shows the results of predicted marginal effects on latency by racial categories, as poverty rate increases, in rural tracts.

In rural tracts, as percent school-age population increases, all racial groups except non-Hispanic American Indian experience proportionally worse latency, with non-Hispanic Black and Hispanic/Latino experiencing the worst predicted effects on

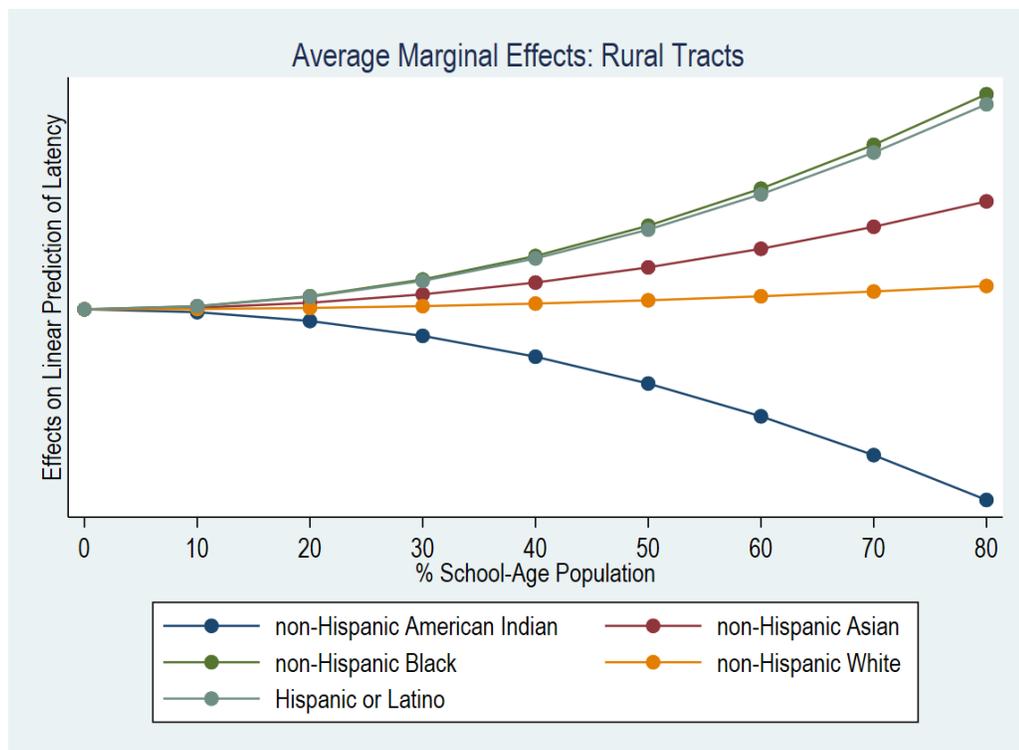


Figure 10 - Marginal Effects on Avg. Latency, Rural Tracts by Percent School-Age Population

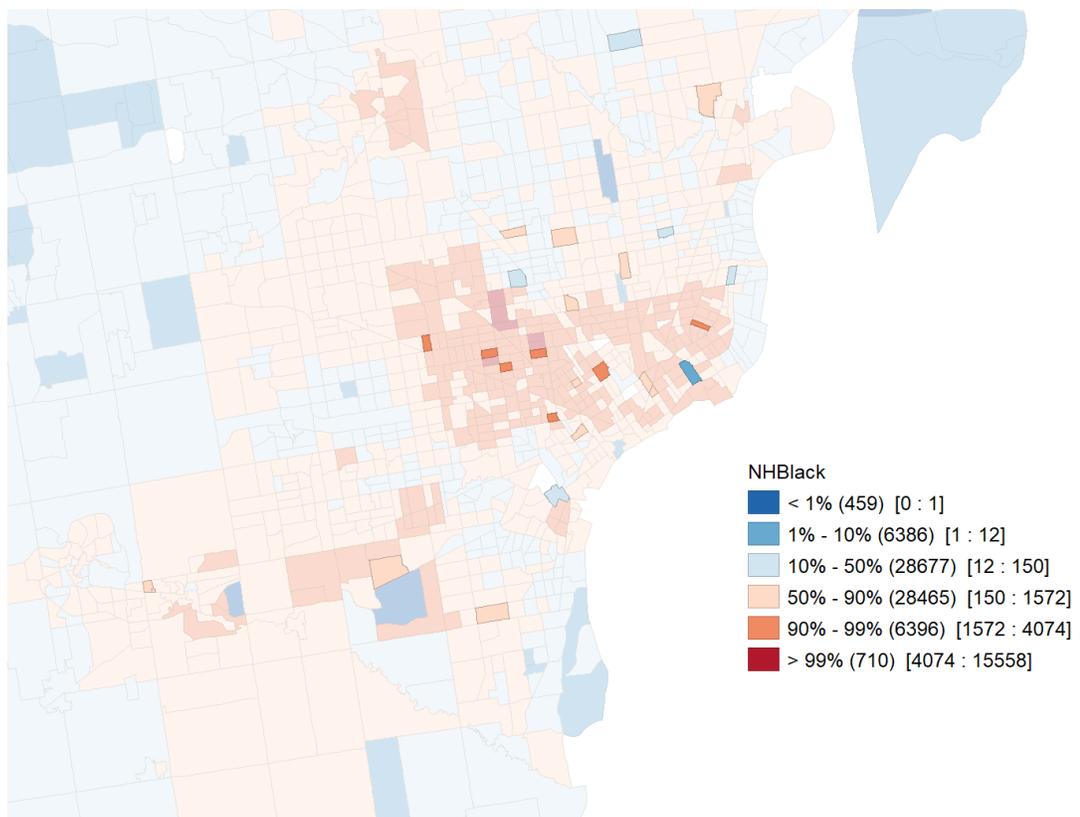
latency. Figure 10 shows the results of predicted marginal effects on latency by racial categories, as percent school-age increases, in rural tracts.

### **Real-World Evidence: Micro Case Studies of Latency & Race**

To directly illustrate how latency disparately impacts non-White populations, even in urban areas, I ‘zoom in’ to two well-known urban centers: Detroit, Michigan, and New Orleans, Louisiana. By examining the latency within these urban census tract groupings, and then observing the tracts’ percentile of non-White population, in this case non-Hispanic Black, I can visually and empirically document the disparity in internet stability. *Note: for all following figures, there are certain census tracts that are highlighted, while remaining tracts are more translucent. This is done intentionally to bring certain tracts to the forefront; highlighted tracts are those that experience high latency [poor internet stability]. In addition to experiencing high latency, the highlighted tracts may also be neighbored by low latency tracts. In chapter three I term these groupings as ‘Regions of Opportunity.’ These are neighborhoods of tracts for which, in theory, it would be inexpensive and efficient to expand or otherwise implement interventions to lower latency [as the surrounding area already experiences low values].*

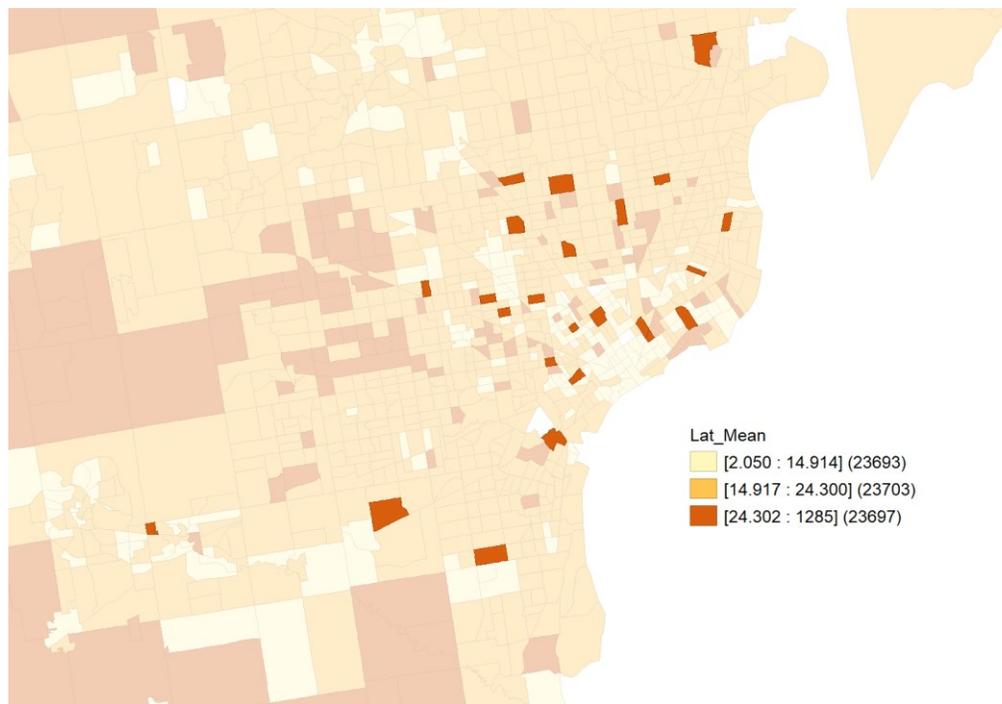
Detroit, Michigan

The first micro case study observes the metropolitan area around Detroit, Michigan. Figure 11 shows the percentile non-Hispanic Black population in each census tract. Starting at the city center, at the center rightmost area of the map, many of the tracts are between greater than 50% non-Hispanic Black population. Suburbs, in translucent light and dark blue, are 10% and less non-Hispanic Black population. The highlighted tracts in the figure, by and large, are >50% or >90% non-Hispanic Black.



*Figure 11 - % Non-Hispanic Black Population, Detroit, MI. Tracts Highlighted: Regions of Opportunity*

Moving to Figure 12, which depicts the same tracts but now shows average latency, we can see that all the tracts are >25ms, which is likely to produce jittery two-way communication [anything greater than 30ms is considered poor stability]<sup>2</sup>. By simply viewing these relationships using exploratory spatial data analysis, we can visually



*Figure 12 – Average Latency, Detroit, MI. Tracts Highlighted: Regions of Opportunity*

observe that the areas of Detroit that experience the worst latency are, by a wide margin, populated by primarily non-White populations. The typologies for neighborhood stability are defined in chapter three, but the highlighted tracts are ‘Regions of Opportunity’ which represent the ‘low hanging fruit’ of connectivity: those tracts that are experiencing poor stability, but all of the neighboring tracts directly contiguous to them experience good stability. It is within reason, then, to assert that, barring an unknown geographic or infrastructure-related barrier, these tracts in the Detroit area

---

<sup>2</sup> Due to limitations of the GeoDa software’s choropleth classification schema, I was unable to manually classify the latency ‘bands,’ which would have allowed for a band to begin at 30ms and above.

could be the targets of efficient, low-cost intervention as compared to distant areas of more than one tract with poor stability.

### New Orleans, Louisiana

A similar story plays out in the metropolitan area surrounding New Orleans, Louisiana. Even in the southern region of the US, in a state with a much higher average non-Hispanic Black population than Michigan, we observe the same troubling trend of stability disproportionately impacting areas of non-White population. Figure 13 shows the percentile of non-Hispanic Black population in census tracts of New Orleans and the nearby locales. Most tracts within the city limits, excluding some recently developed areas downtown, and nearly all tracts in the rural areas surrounding New Orleans are at >50%, >90%, or >99% non-Hispanic Black population. Highlighted tracts in this figure again represent the ‘Regions of Opportunity’ where the highlighted tract experiences poor stability but is surrounded on all sides by tracts that experience good stability. It is interesting to note here that some of the ‘good stability’ tracts are indeed populated by majority non-Hispanic Blacks, but the fact still remains that those tracts with the worse stability do not have a majority non-Hispanic White population.

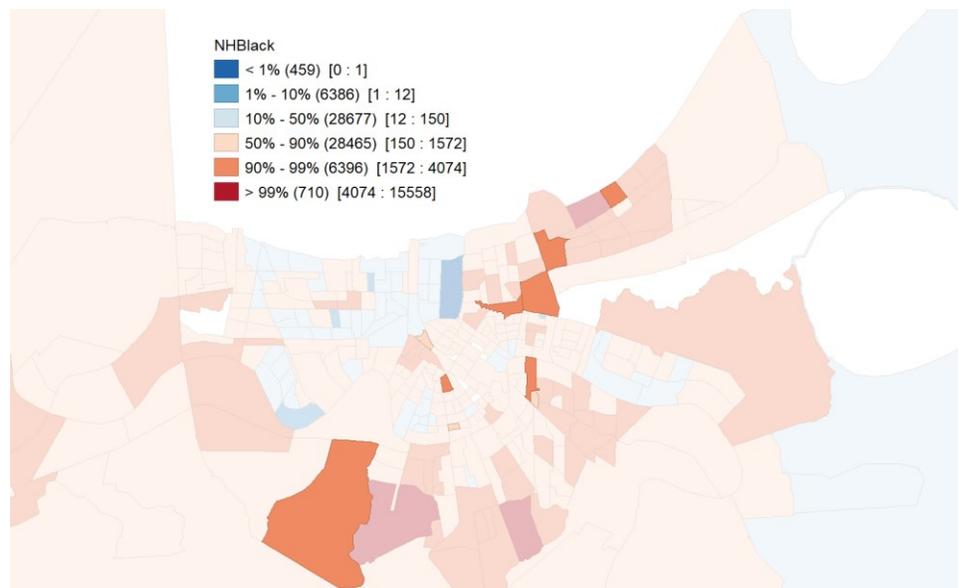
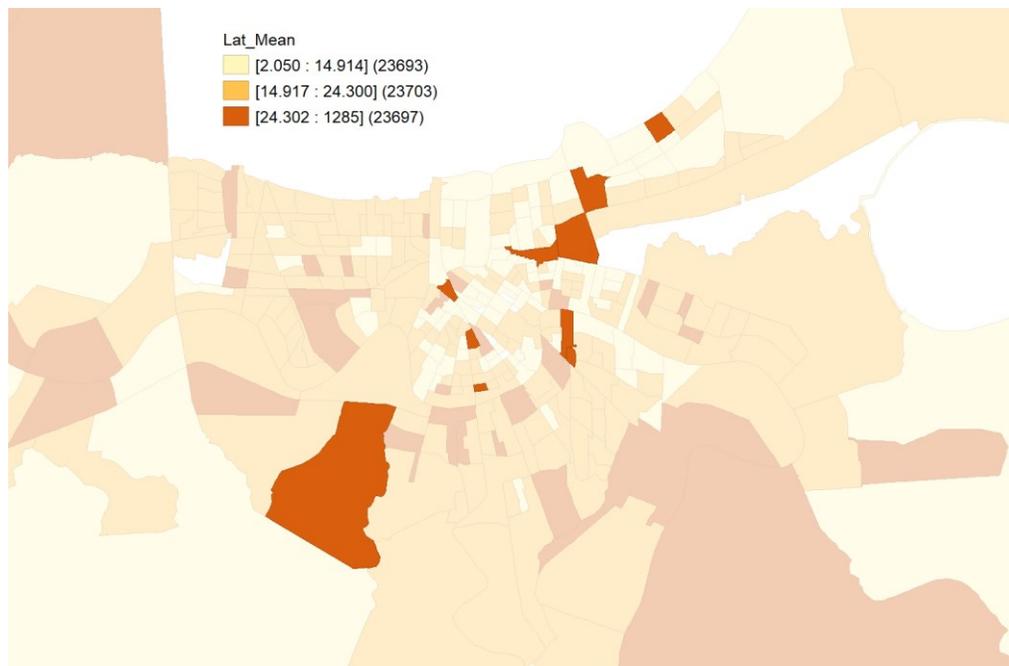


Figure 13 - % Non-Hispanic Black Population in New Orleans, LA. Tracts Highlighted: Regions of Opportunity

Observing average latency in Figure 14, we see the full effect of living in the southern region of the US as highlighted in Chapter 1. The vast majority of the tracts in this area, above and beyond racial composition, experience middling or worse stability. The highlighted tracts, though, are all those that experience poor stability. By mapping and empirically confirming those tracts that experience poor stability, policymakers have an interactive method with which they can identify areas for direct intervention. Like the example in Detroit, observing the Regions of Opportunity in New Orleans provides an example of urban areas that also experience poor latency. Such a finding dovetails into the results presented in the multivariable analyses of this chapter that show a non-linear effect of latency with regard to poverty and rurality. Not all urban areas universally experience stable internet and, in fact, those urban tracts that have predominately non-White populations tend to experience worse stability than their urban, White neighboring tracts.



*Figure 14 - Average Latency in New Orleans, LA. Tracts Highlighted: Regions of Opportunity*

## **Discussion**

As previous research indicates, higher values of latency are **not** desirable outcomes and, as such, only when negative coefficient directions or lower values of latency are reported should the results be interpreted to be beneficial (Geelhoed et al., 2009; Kryczka et al., 2013). While univariate results do confirm variance in latency between racial categories, they do not indicate the direction of the hypothesized relationship exactly. Non-Hispanic non-White racial groups do not have unilaterally worse latency outcomes. Instead, latency is similar between all groups but non-Hispanic Asians in urban tracts. In rural tracts, however, I do show that non-Hispanic White and non-Hispanic Asian groups are less disadvantaged than non-Hispanic Black, non-Hispanic American Indian and Hispanic-Latino populations. This confirms hypothesis one in that there are statistically significant differences in latency between racial groups. Bivariate results give us the first indication that tracts with higher rates of poverty are disadvantaged regarding latency; both localities return positive correlations with latency. Multivariable analyses further confirm hypothesis two: in both modeling variants of the OLS regressions, as well as in the marginal results, increases in poverty rate results in higher latency, putting tracts with higher levels of poverty at greater disadvantage. Furthermore, above and beyond any racial effects, a unit increase in tract level school-age population results in higher latency. This offers a confirmation of hypothesis three in and of itself, but it is further confirmed when controlling for racial differences. For all groups except non-Hispanic American Indian, which returned insignificant, the interaction between each race and percent school age was statistically significant, even at subsampling levels. Due to the numerous interaction effects including squared terms to account for non-linear univariate relationships, the most useful interpretation of the poverty rate and percent school-age regression models comes from the marginal results.

Margin plots for urban tracts are included in the technical appendix but I only discuss results for rural tracts here, as we know them to experience worse latency even before subcategorizing on race (Fiduccia, 2022). Figure 15 shows the linear prediction of latency, for all racial categories, as poverty rate increases, for rural census tracts. Even in rural tracts, many of which are majority non-Hispanic White, there is a striking disparity between White and non-White populations' experience of internet instability. Figure 16 shows the linear prediction of latency, for all racial categories, as percent school-age population within the tract increases, for rural census tracts. Though we know the interaction between percent school-age population and non-Hispanic American Indian race, specifically, to be insignificant, the predictions on latency shown in this figure serve to paint a concrete portrait of the digital divide in the rural tracts of the United States (Hollman et al., 2021; Lehr et al., 2006).

As the population of school-age children increases and so, too, do the number of individuals and households connecting to the internet in rural areas, we can see evidence of higher latency. Non-Hispanic Whites, similar to the poverty rate effects, experience the 'best' stability by comparison, but are still at higher risk of unstable connections as compared to their urban tract counterparts. This speaks directly to the concern educators, families, and researchers have about remote learning and the differential outcomes that may be present due to inequitable platforms, access, and connectivity (Francis & Weller, 2022; Hobbs & Hawkins, 2020; Kane et al., 2006; Murat & Bonacini, 2020).

In the technical appendix, multiple figures are presented which diagram the relationship of predicted marginal effects on latency by individual OLS regressions using racial analytic weights, as percent school-age and poverty rates increase, by region of the US and rurality. These marginal results are useful in diagramming the average, race-specific effects on latency. The results replicate effects noted in the main

combined regression. On the aggregate, the individual, race-analytic-weighted marginal results reinforce the finding that rural tracts experience worse internet stability than urban tracts. Additionally, for every race other than non-Hispanic American Indian, northeast tracts experience more stable internet. Finally, across all regions and ruralities, non-Hispanic White populations tend to start at a more stable level of internet [lower latency value] than all other races.

### **Looking Forward**

This paper expands upon the literature surrounding the digital divide (Clotfelter et al., 2008; Hollman et al., 2021; Lehr et al., 2006), particularly scholars who have begun to examine the differential effects of the switch to remote learning on race and socioeconomic status (Francis & Weller, 2022; Murat & Bonacini, 2020), by providing an empirical evaluation of latency by race across all census tracts in the United States. I find that non-Hispanic non-White populations experience worse internet stability than White populations, that this effect is worsened in rural tracts, that poorer tracts experience worse stability, and that tracts exhibit worse internet stability as the percentage of school-age population increases, though the last effect is likely observed in parallel with increased overall population density.

Additionally, this paper provides a quantitative baseline for future research aimed at examining how COVID-19, particularly the switch to remote instruction, influenced the achievement & outcome gaps in the US. As scholars continue to explore the effect of the pandemic on schooling, questions surrounding the efficacy of non-synchronous and remote learning are already being asked (Hobbs & Hawkins, 2020). When households are queried about how their students experienced / remained connected to the classroom, any examination of participation in ‘Zoom-classrooms’ must take into account the stability of the student’s connection to the instructor - a metric directly impacted by the latency of the household. This paper serves as an

empirical companion for qualitative studies, a foundation for community-specific analyses, and a contribution to the literature for education in the age of the COVID-19 pandemic.

Finally, the measurements of social and economic phenomena, such as latency, lend themselves to be examined through the spatial context. After all, we know the first law of geography to be true: all things are related, but those things that are closer are more related (Tobler, 1970). In a companion analysis, I examine the spatial clustering of latency to showcase hot spots, statistically significant clusters, and spatial patterns within the latency landscape. The goal of this work is to, one, build a more nuanced predictive model for latency and, two, utilize spatial kriging methodologies to impute unknown and unmeasured values more accurately in the hopes of providing policymakers better, more efficient information.

---

## **Chapter 3: Spatial Dependence of Internet Stability in the US**

### *An Exploration Using Local Regression and Kriging*

#### **Introduction**

Access to the internet is quickly becoming recognized worldwide as a determinant of social, economic, and even physiological well-being. With the continuing increase in usage of telehealth, remote work & learning, and distance collaboration tools, the importance of internet access is underwritten by assumptions regarding the internet's stability of connection. A broadly accepted metric to ascertain two-way video & audio stability is latency, also referred to as ping. Being able to empirically and visually describe the spatial distribution of latency across geographies is of critical importance to understanding where potential policy interventions or government assistance programs are most needed. Similarly, understanding the geographic landscape of latency can reveal inequities between socioeconomic, racial, and regional populations. In this paper, I seek to explore and clearly identify neighborhoods that experience varying levels of internet stability, and advocate for the continued usage of spatial methods in social science; providing more nuanced methods by which we can predict latency via local regression and kriging analyses. I provide a detailed literature review outlining historical and recent research surrounding the internet as a resource, local cluster analysis, and kriging in the social sciences. I find that internet stability is indeed clustered across certain neighborhood of census tracts in the United States, that we can reliably increase the explanation of variance in latency using local regression modeling, and that both ordinary and universal kriging methods can provide insight into the predicted values of latency where measurements currently do not exist. I conclude by calling for additional research into the sociodemographics of the latency neighborhood clusters, as well as more empirically robust kriging models to further our ability to impute unknown values in future research on internet stability.

## **Background and Literature Review**

### *Access to, and Adoption of, the Internet as a Resource*

Since the mid-1990s and the significant increase of internet adoption across the US, many describe this shift in quick access to information as the catalyst for the global information economy (Webster 1995). This use of this information to bring about broad social and economic change is characterized as the ‘informatization of social life,’ a phenomenon which grants, to all, access to information that was previously only available to the few (Cairncross 2002). Even in the early 2000s scholars noted the potential of the internet to positively influence the economic, social, and developmental outcomes of rural communities as digital connectivity reduces the gap usually experienced by geographic distance and ‘remoteness’ (Leatherman 2000). Economic development in rural areas is the primary lens through which internet access has been viewed over the course of its expansion from urban centers (LaRose et al. 2007), which was the case regardless of rurality during the late 1990s (Montes 2003). Despite the promises of increased economic viability to both rural and urban areas alike, we know that internet access is not universally, or even equally, distributed across the nation (Douthit et al. 2015; Grubestic and Murray 2002; LaRose et al. 2007).

The expansion and proliferation of internet access is often described as a market-driven process, one which is largely separate from US federal policy providing any kind of universal financial backing for the build-out of infrastructure (Downes and Greenstein 2002). Formally, the Telecommunications Act of 1996 is meant only to ‘promote competition,’ and ‘enhance universal service,’ but does not strictly delineate the means by which the federal government does so, least of all in rural areas (Alter et al. 2007). This policy places the *de facto* responsibility of broadband access expansion on private firms, relying on the commercial sector to increase universal access by

means of business-level investment in infrastructure and connectivity (Maxwell 2005). In many cases, the ability for companies to utilize existing pole, ground line, or copper phone line wire routes to upgrade to higher-speed internet means foregoing the more expensive investment in wireless transmission technologies [which are much lower latency, on average, than fixed broadband delivery systems (Katz, Avila, and Meille 2011)] (Malone 2001; Skerratt 2010). One logical way of exploring the diffusion of internet infrastructure would be to examine telecommunication companies' decision-making processes for where to expand service. Unfortunately, but not surprisingly, this information is closely guarded by companies, wherein many indicate that divulging this information would give competitors an unfair advantage to understanding market demand; so much so that even the Federal Communications Commission [FCC] has been unable to solicit the information from private firms (Fishman 2021). Therefore, understanding where new service areas are being planned or by what factors new markets are determined is extremely challenging (Chaudhuri, Flamm, and Horrigan 2005). Research in the mid-2000s attempted to break the code on the most challenging component of all--price--but since this data, too, is not readily available to the public, the most researchers were able to determine was that the factors are "...numerous and might vary by situational contexts" (Black et al. 2009; Khatiwada and Pigg 2010). Khatiwada and colleagues executed a comprehensive, county-level evaluation of current service provision across the US in 2010 and made note of the unique and varied challenges with using FCC-provided data, particularly when trying to get to any unit more granular than county (Khatiwada and Pigg 2010).

Regardless of the means behind private firms' choices of service area expansion, it has been widely recognized that access to the internet is a social and economic resource, recently one which can have an influence on healthcare, particularly during eras such as the COVID-19 pandemic (Zahnd, Bell, and Larson

2021). Classified as a “super determinant of health,” internet / communications infrastructure can directly influence individuals’ and communities’ abilities to access healthcare related resources (Fishman 2021; of Health, Services, and Others 2020). In the recent pandemic, internet access was clearly linked to the ability to access telehealth, remote education,, work-from-home environments, and even online vaccine information platforms (Loeb, Atkins-Jackson, and Brown 2021). Previous research also identifies broadband access as a means to access routine medical care, particularly specialists, for rural communities as well as a means for individuals to avoid the dangers of exposure to COVID-19 during the height of the pandemic period (Daniel Xu 2020; Stratmann and Baker 2020). Zahnd and colleagues build off of Khatiwada’s original work exploring the FCC data by including data from the American Community Survey (ACS) data, which includes broadband subscription information; this, in conjunction with FCC data, allows them to classify areas of “realized broadband access” that delineate infrastructure-based access and potential access based on areas of subscription availability (Zahnd et al. 2021). Recent studies expand upon both works by presenting latency, a measure of internet stability, as a further evaluation of reduced access and equity for rural communities (Fiduccia 2022).

Finally, the COVID-19 pandemic has widened economic, achievement, and resource gaps, one of which being the disparities in technological access between various groups [the digital divide] (Vogels et al. 2020). Due to economic strains caused by the pandemic, non-Hispanic Black and Hispanic/Latino communities were reported to be twice as likely to have cut or canceled their internet service, exacerbating the already present digital divide (ibid). The access issue is even more apparent for non-Hispanic American Indian communities, with less than half of those living on reservations with computers having access to an internet connection classified as ‘high-speed,’ as compared to ~80% nationally (Wang 2018).

Additionally, recent studies have explored the influence of income segregation, poverty, and neighborhood effects further exacerbated the pandemic for populations of color, in turn worsening digital inequities (Francis and Weller 2022). This further positions internet access and, by extension, internet stability, as a spatially-oriented social disamenity. Therefore, we must continue to explore the spatial diffusion of internet technology in order to solidify the foundation for geographic analyses.

*Spatial Diffusion of the Internet: Neighborhoods of Instability*

Apart from the existence or nonexistence of a particular resource within a singular geography, the variance in resources, or social phenomena, over space is crucial to understanding nuanced, local effects. Spatial processes and structures have been widely evaluated for many years in the urban planning, soil science, and geographic science literatures, but only recently have social science constructs been analyzed using advanced geospatial methods (Anselin 1995; Fotheringham, Charlton, and Brunson 2001). Geographic Information Systems (GIS) software and techniques have given rise to now-familiar social science constructs & terminologies such as neighborhood effects, peer effects, social capital diffusion, and network effects, all of which rely on the empirical evaluation of location within the predictive models and descriptive constructs (Tobler 1970); *ibid*). Location within space has a direct influence on the diffusion of technologies and phenomena (Anselin, Syabri, and Kho 2010). As a social and economic resource, the internet is concentrated in some places more than others, primarily because of the inground, built infrastructure in the US. Research in the late 1990s determined the earliest evaluation of broadband access disparities by showing that the Pacific Northwest and New England subregions had the highest density of internet activity via domain registration data (Moss and Townsend 1998). As mentioned previously, the ‘felt distance’ between disparate locations is purportedly reduced because of an increased ‘interaction speed’ (Grubestic

and Murray 2002). Though we do not fully understand the exact mechanics behind the structures, price is also directly variable by location (Chaudhuri et al. 2005). Data from the National Telecommunications and Information Administration [NTIA] found positive correlations between internet usage, educational attainment, household income, age, employment status; in addition to a noticeable gap in usage between non-Hispanic White and all other racial groups (McLoughlin 2005).

As has been done by both the FCC and previous researchers, using simple geographic distributions, identifying regions of broadband access can be done by defining contiguous zip codes or counties that share similar levels of broadband availability (Black et al. 2009; LaRose et al. 2007). Outlined in the methods section in more detail, I analyze the landscape of internet stability in this paper using spatial autocorrelation, as some researchers in recent years have used this method to analyze the general availability of broadband by zip code (Grubestic 2006; Grubestic and Murray 2002) as well as census tract (Zahnd et al. 2021). By ascertaining the interdependence between values of the same variable at different locations, a ‘neighborhood of instability / stability’ can be constructed and observed (Cliff et al. 1981).

#### *Kriging for Social Science Analysis*

The dataset utilized in this study from the Ookla company is quite robust, but not universal in its coverage. Since the data comes from the firm’s own service provision, it contains more direct observational data than a random sampling. The limitation therein lies with the impetus of an individual running a test; if nobody within the 1x1km<sup>2</sup> tile area decides to run a speed test, there will be ‘no data’ for that tile. Subsequently, if a census tract happens to contain only a single tile, and there are no tests, there is ‘no data’ for that tract. Though this is rare, in order to ascertain a complete sample we could impute those values using established geostatistical tools

such as interpolation (Auchincloss et al. 2007; Bader and Ailshire 2014; Doelling et al. 2013; Mooney et al. 2020). One method within the interpolation toolbelt is kriging, which estimates a value and a margin of error for unobserved or unknown locations (Cressie 1988). One limitation to ordinary kriging, however, is its inability to precisely measure ‘between-location variation’ or when there are correlates with the dependent variable of interest in a multivariable model (Mooney et al. 2020). One body of literature and associated methodology that attempts to mitigate the limitations of ordinary kriging is ‘land-use regression.’ Land-use regression can estimate unknown values of a variable of interest where covariates have been measured, but the variable itself has not (Ross et al. 2007). The primary limitation of this method, however, is that land-use must be strongly tied to the model or variable of interest; a direct result of this method being a derivation of the physical disorder and owner-occupied-homes areas of study (Hoek et al. 2008). Universal kriging is an interpolation technique that can incorporate spatial correlation measurements between both known observations as well as their covariates in the same model (Clougherty et al. 2013; Hengl, Heuvelink, and Stein 2004; Stein and Corsten 1991). This paper will utilize centroid data from the Ookla tiles to be the first implementation of universal kriging on a social-phenomenon dataset such as internet stability.

### **Data and Research Design**

The research questions underpinning the project focus on the spatial relationships within latency, as well as the methodological efficacy of universal kriging, and are outlined here. First: *is latency spatially autocorrelated across US census tracts?* The spatial null hypothesis posits that latency values across census tracts will be distributed randomly with no clustering or dispersion. I estimate there will be positive spatial autocorrelation, indicating clustering [H1]. Second: *does a local regression model account for more overall variance in the latency measure than*

*a traditional OLS model?* The spatial null hypothesis posits that there will be no difference in explanatory power of the model between a spatially autoregressive model and an ordinary least squares regression. I hypothesize that the spatial error model will have a higher overall predictive power than the OLS [H2]. Third: *is universal kriging a more effective means of predicting or imputing unknown values of latency with census tracts?* This is a purely methodological exploration of the kriging subtypes [ordinary and universal] based on advances in spatial empirics and I estimate it will be a more nuanced way to impute values than simple non-spatial averaging [H3].

This framework further expands on latency as a geographically diffused and clustered ‘disamenity,’ allowing for the exploration of internet stability via spatial means. Spatial analyses are meant to provide more nuanced information than a-spatial methods, primarily by linking to a direct geographic coordinate position, accounting for influences of neighboring locations, or describing various attributes of variables with respect to location (Chaney and Rojas-Guyler 2016). By analyzing internet stability using spatial methods we are better able to identify three key elements above and beyond a-spatial analyses. 1) Which tracts of higher or lower latency are clustered together 2) Where the regression model predicting latency is a better or worse fit across tracts 3) Whether we can better predict unknown values of latency as compared to a-spatial methods.

### *Ookla Speed Test*

National-level data describing internet speeds, access, and/or connectivity are sparse. ISP data is not ideal because there is an inherent conflict of interest as some internet service providers [ISPs] will present data from their own networks that only paints them in a favorable light. Independent information gathered from state agencies does exist, but the levels of analysis or collection methodologies are often disparate, which would make uniting the data for a national analysis inefficient. I use a novel dataset

from the technology company Ookla. Ookla's services include the widely utilized Internet Speed Test. This test, run through an internet browser on the user's desktop, tablet, or mobile device, measures upload & download speeds, as well as latency to the device executing the test. The dataset also measures how many tests are executed, as well as how many unique devices exist, within the collection zone. The data are presented in survey areas at the 1x1km level, in squares or 'tiles.' The files are provided open-source and free to the public as part of the company's program Ookla for Good. They are updated on a semi-annual basis.

#### *Creating the Dataset: Combining Ookla and Census Data*

While the data covers the entire globe, I focus on the continental United States in this analysis. Raw Ookla data is presented in the 1x1km tiles. Within the dataset, translated to a shapefile using a spatial analysis software program [QGIS], are the unique IDs, as well as values for upload, download, latency, # of devices, and # of tests. The raw information covers the entire globe, so it was parsed to the continental US [excluding Alaska, Hawaii, and Puerto Rico]. In total, there are ~1.6M tiles across the country. Using the state of Delaware as a test case for unit of analysis, I identified that 98% of the tract polygons contained internet speed data. Though counties were at 100% coverage, the aggregation of tiles would be too high. Conversely, block groups only had a 15% coverage rate--due to many of the groups being below 1x1 km in area--which would have resulted in undesirable overlap issues. Therefore, I decided on tracts rather than groups, block groups, or counties. I first created centroids from the Ookla tiles, then identified how many tracts did or did not contain data from those centroids. A wide variety of spatial operations were then performed to link the Ookla tiles / variables to the census tract in which they were contained. I then linked state and county identifying information to the Ookla centroids.

Taking into account the various transformations and operations, it is important to note that the resultant dataset used for analysis is novel. Using a unique data source, I create a novel compiled dataset by merging and cross walking disparate, quantitative & geographical raw files. This was a major component of the project prior to performing any descriptive or predictive analyses. The final sample size for this project is 71,093 US census tracts. This is after parsing non-populated tracts and tracts without Ookla fixed broadband test information [which, together, totaled less than 0.5% of the entire original sample]. Ookla tiles, which are aggregated to the tract level, total 1.7 million observations for the continental US.

#### *Local Analyses: The Spatial Context*

The literature surrounding place, health, internet access and other sociological phenomena reinforces my supposition that geography matters: where a family lives and, subsequently, where a child attends school [one example] can have a profound and varied impact on outcomes. Research indicates that neighborhood context, including poverty rates, educational attainment, and family composition can contribute to increasing socioeconomic segregation, have direct effects on childhood development, and can be a determinant factor in outcomes (Owens, Reardon, and Jencks 2016). Bischoff & Reardon illustrate the example of two children in socioeconomically disparate neighborhoods, and this relates directly to our discussions on social / cultural capital (Reardon and Bischoff 2011). This variation between areas within a community or school district is not a newly researched concept. Some aggregate literature on education, policy, and geography, others relate spatial theories, access, equity, and the educational differences between rural and urban districts, while others still speak to the importance of communities' understanding the geographical makeup of their own locale, as well as surrounding areas, to make informed policy decisions on neighborhood planning and educational resources (Gulson and Symes

2007; Hogebe and Tate 2012). Crowder and South posit the notion of an “extralocal” neighborhood effect as the “...areas surrounding an individual’s neighborhood of residence.” This conceptualization of a neighbors-of-neighbors effect gets at the heart of the spatial statistical assumption: It is not in the most appropriate standing for social science researchers to assume a model will predict outcomes uniformly across a geography – which is exactly what we do when we run global analyses such as traditional OLS regressions (Crowder and South 2011; Pasculli et al. 2008). I detail the methodological considerations to spatial regression approaches in the methods section, but it is important to set the stage for why, particularly with such a robust dataset, we would want to account for local variation.

## **Methods**

### *Exploratory Spatial Data Analysis [ESDA] & Local Regression*

Local variation is examined through the usage of various spatial analysis techniques. Broadly, descriptive mapping, or data cartography, is used to produce maps of a state, district, or many districts with choropleth visualizations; a map in which geographically bounded areas are shaded based on the variability in the variable of interest. Descriptive mapping is meant to visually represent data on a scale which provides a geographically comparative perspective. Furthermore, global spatial autocorrelation (Jerrett, Gale, and Kontgis 2010; Legendre 1993; Moran 1948) is used to determine whether a large, significant, underlying pattern is present in data spanning a set unit of analysis. Local Indicators of Spatial Autocorrelation [LISA] (Anselin 1995) analysis is used to further narrow the frame of reference. LISA is used to determine the local clusters originally identified in the global analysis and subsequently display them graphically. The neighborhood is determined prior to running a spatial regression, the final and most precise step in spatial analyses. Spatial regression is meant to account for local variation within the context of the prediction

equation and, as such, provides more nuanced results than traditional OLS (Fotheringham 2009; Fotheringham et al. 2001). Local regression mitigates the effect of clustering or dispersion, otherwise characterized as spatial heterogeneity or nonstationarity, an example of both is shown in Figure 15. The final regression, and any changes in explanatory power or coefficients, remedy as much spatial autocorrelation as possible, thereby accounting for local variance.

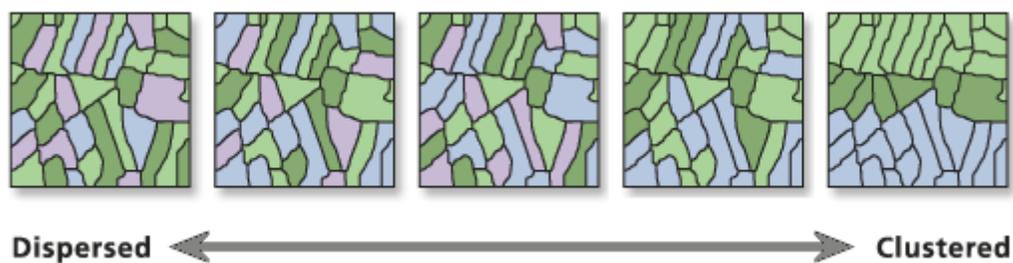


Figure 15 – Examples of Spatial Autocorrelation: Dispersion, Random, Clustering

With regard to the identification of spatial clustering, specifically H1, the global and local spatial clustering measures [Moran’s I] can be used to identify contiguous census tracts for which the local indicator of spatial autocorrelation [LISA] for average latency is significant (Anselin 1995). As has been used in previous spatial explorations of sociological phenomenon, I construct the spatial weights matrix using second-order queen’s contiguity (Kolak et al. 2020). This means each census tract, combined with any number of census tracts that are directly contiguous and the next set of physically contiguous census tracts, is defined as a ‘neighborhood.’ By identifying these spatially-defined neighborhoods, I then execute the global Moran’s I, local Moran’s I, significance clustering maps, and the local regression model. In this paper, I contextualize and define the five primary forms of neighborhoods output from LISA clusters in the following manner:

1. High-High: This is where a census tract returning high values of latency [poor stability] is surrounded by other census tracts of high latency. I term these **‘Regions of Unstable Connection.’**
2. Low-Low: This is where a census tract returning low values of latency [excellent stability] is surrounded by other census tracts of low latency. I term these **‘Regions of Stable Connection.’**
3. Low-High: This is where a census tract returning low values of latency [excellent stability] is surrounded by census tracts of high latency. I term these regions **‘Regions of Disparity’** because the island of excellent stability is, for whatever reason, unable or unwilling to diffuse the low latency internet to surrounding high latency tracts.
4. High-Low: This is where a census tract returning high values of latency [poor stability] is surrounded by census tracts of low latency [excellent stability]. I term these **‘Regions of Opportunity’** because there may be opportunities for the nearby, low latency tracts to expand access or infrastructure, to reduce instability of the single high latency tract.
5. Statistically Insignificant: This is any tract that does not return statistically significant at the 0.05, 0.01, or 0.001 level.

### Kriging

In order to ascertain the viability of advanced imputation methods, both ordinary kriging and universal kriging are utilized in the analysis of latency across multiple spatial areas. Ordinary kriging is utilized to form a baseline measurement of two areas: New York state and several states in the Western Census Region of the United States. Kriging analyses are performed using the mean latency measure as the variable of interest (Wahba 1975, 1990). Since kriging is a point interpolation tool, the primary dataset utilized for the rest of the paper, a polygon shapefile of all census tracts in the US, cannot be used. Therefore, a dataset consisting of the point centroids of each

1x1km Ookla tile is used instead, significantly increasing the granularity of the data while simultaneously providing a richer landscape from which to predict unknown values. This format enables the tool to assess direction and distance between each centroid point, then estimate a semivariogram to create a prediction raster surface. For the purposes of this paper, a circular semivariogram is utilized. In short, various ‘forms’ of semivariograms are utilized for different theoretical and/or empirical assumptions of spatial autocorrelation over distance. Since the spatial regressions in this paper use first and second order queen’s contiguity weighting matrices, I make the assumption that autocorrelation increases at a short distance between census tracts, then peaks, then decreases as distance continues to increase. This forms the basis for the distance ‘limit’ for the second order queen’s matrix and informs the decision to use a circular / spherical semivariogram, which also assumes a drop in autocorrelation over an assumed distance. Much like the inverse distance weighting function, the measured values closest to the location of the unmeasured area of interest have the most influence on the calculation. While IDW uses a simple distance weight, however, kriging utilizes the matrix created by the spatial semivariogram.

## **Results**

### *Global Autocorrelation*

First, a test for nonstationarity and spatial heteroskedasticity is performed by way of the global Moran’s I. A first-order queen’s contiguity matrix is tested first, returning a value of +0.46, indicating a significant positive clustering effect. Using the second-order contiguity weighting, significant, positive autocorrelation was again observed at +0.35, shown in Figure 16, indicating the presence of strong spatial clustering effects that should be mitigated in order to provide more nuanced regression results [H1].

Two positive autocorrelation values from both first and second order matrices indicate

that similar values of latency tend to cluster together. In order to directly identify where these clusters are, and their characteristics, a local Moran's I is performed.

LISA Neighborhood Typology Analysis

The results from the local indicators of spatial

autocorrelation [LISA] analysis are shown in Figure 17. Following the schema noted above, Regions of Unstable Connection are found in predominantly rural areas, with a total of 9,372 tracts comprising this typology of neighborhood. Regions of Stable Connection are found in predominantly urbanized, developed areas, with a total of 29,092 tracts comprising this typology of neighborhood. Regions of Disparity are primarily found on the intersectional fringes between Regions of Unstable Connection and Regions of Stable Connection, with a total 3,707 tracts comprising this typology of neighborhood. Finally, Regions of Opportunity are primarily found within or on the very edges of Regions of Stable Connection, with a total of 509 tracts comprising this typology of neighborhood. Specific geographic trends of the neighborhood typologies listed above are detailed in the discussion. Having noted the characteristics of the LISA analysis, however, a multivariable analysis can be performed to predict influences on latency.

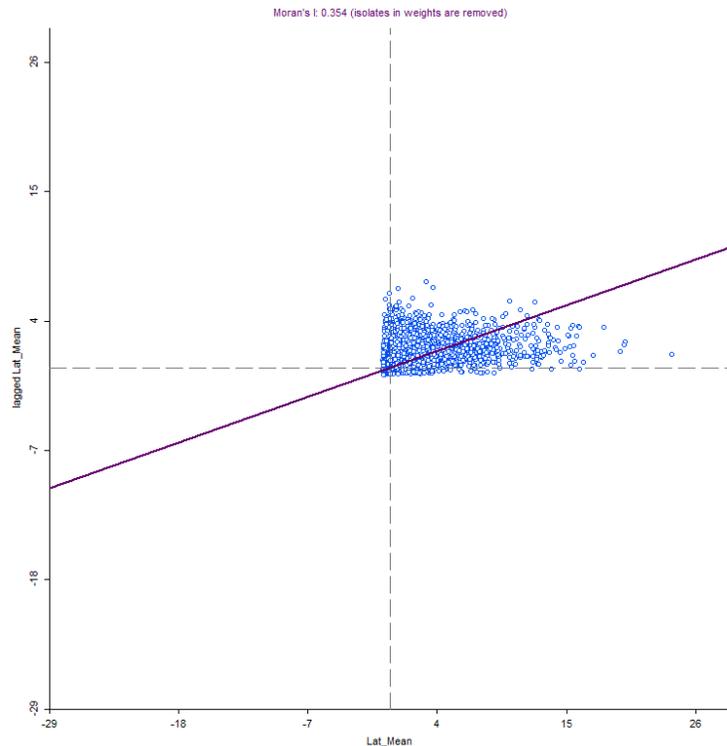


Figure 16 – Moran's I Analysis of Mean Latency

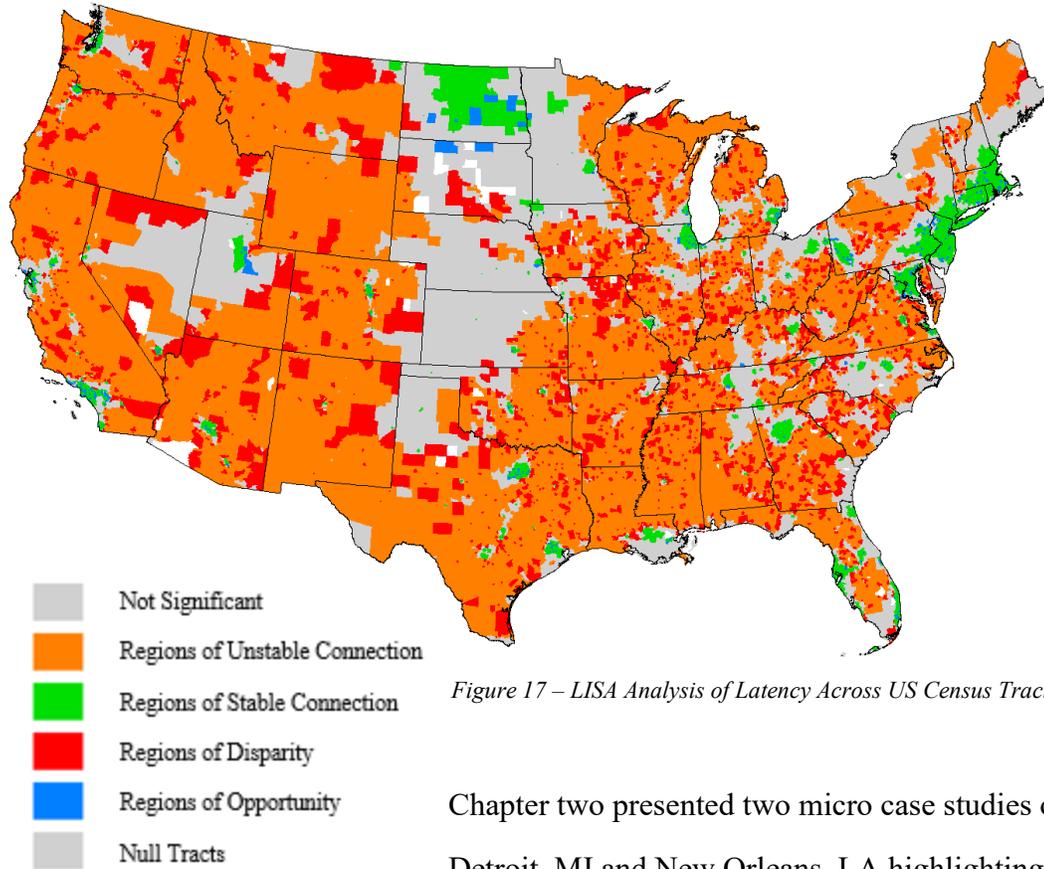


Figure 17 – LISA Analysis of Latency Across US Census Tracts

Chapter two presented two micro case studies of Detroit, MI and New Orleans, LA highlighting the ‘Regions of Opportunity.’ As observed in Figures 18 and 19, both of these urban, metropolitan areas present central tracts of high latency surrounded by neighbors of low latency. These can be thought of as the low hanging fruit for policy intervention, namely because all surrounding tracts have either current infrastructure or other already-in-place system that is resulting in low latency. Since the tracts within the Regions of Opportunity are also primarily non-White populations, this also serves an equally important access & equity argument: those populations already at risk are primarily non-White. By focusing on these regions first, access for those traditionally underserved will be directly influenced in the most efficient manner possible.

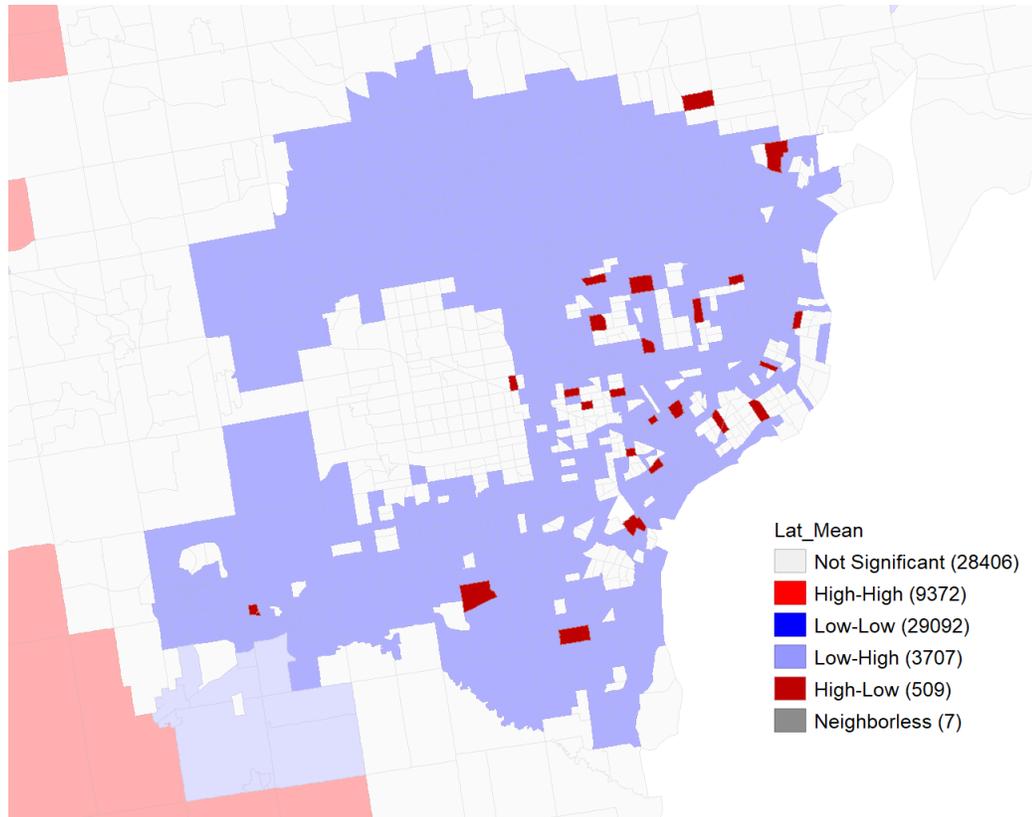


Figure 18 – LISA Neighborhood Analysis, Detroit, MI

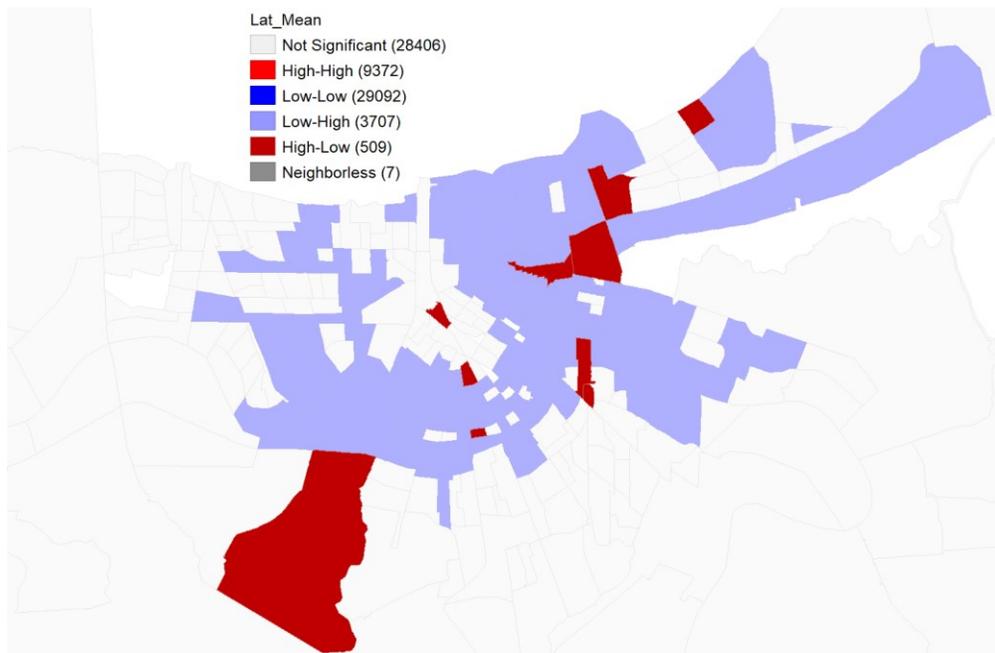


Figure 19 – LISA Neighborhood Analysis, New Orleans, LA

### *Non-Spatial and Geospatial Multivariable Analyses*

It is important to note that there is a tradeoff between a-spatial models and spatial models with regard to predictor variable specificity using the spatial software [GeoDa] in this study. While the error model adds nuance by accounting for variance across space [spatial moving average], it does not allow for the setup of more complex modeling intricacies such as analytic weights, as were used in clustered robust fixed-effect models and scaled OLS models (Fiduccia 2022). Output from the traditional OLS model is shown in the appendix, while output from the spatial error model is shown in Table 7. Accounting for spatial variance, we observe both changes in predictor coefficients & significance as well as a noteworthy increase in explanatory power for mean latency: from 0.27 in the OLS model to 0.38 in the spatial error model [H2]. From the spatial error model: On average, without accounting for specific racial population weights, those neighborhoods with an increasing percentage of school-age children experience lower latency. The constant can be interpreted as the following: An urban census tract [+ two contiguous orders of tracts surrounding it] with 0% school-age children, a zero-poverty rate, in the Northeast, with primarily non-Hispanic white population has an average latency of +19ms. All regions of the US, as compared to the northeast, experience higher levels of latency: the Midwest, south, and west regions experience +7.7, +9.6, and +5.9ms of additional latency as compared to the northeast, respectively. Post spatial-error model, a Moran's I statistic was run on the residuals of the regression model to test for mitigation of autocorrelation, which returned at 0.01, shown in Figure 20. This change, from 0.35 to 0.01, verifies that the coefficients and significance in the error model are free from spatial variance and do not violate the assumption of independence of observation.

Dependent Variable Average Latency (in ms)	OLS R <sup>2</sup> = 0.27	Spatial Error R <sup>2</sup> = 0.38
Predictor Variables	Coefficients + Significance	
Poverty Rate	11.0 ***	1.6
% School-Age Children	-9.6 ***	-16.5 ***
% Non-Hispanic Black	(-) ***	(-) **
% Non-Hispanic Am. Ind.	(+) ***	(+)
% Non-Hispanic Asian	(-) ***	(-) ***
% Hispanic / Latino	(-) ***	(-) ***
Rural Tract Yes/No	34.9 ***	28.6 ***
Majority Non-White & Rural	28.8 ***	22.9 ***
Interaction: Rural x Poverty	(+) ***	(+) ***
Midwest Region	8.0 ***	7.7 ***
South Region	10.5 ***	9.6 ***
West Region	10.5 ***	5.9 ***
Constant	15.0 ***	20.0 ***

Table 7 – Comparison Between OLS and Spatial Error Models

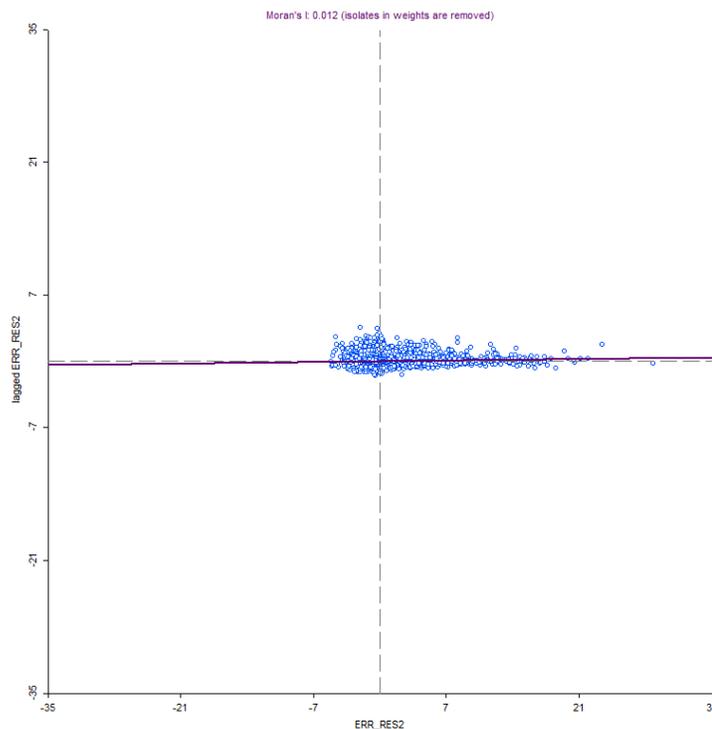


Figure 20 – Post Spatial Error Model Moran's I on Residuals

Ordinary & Universal  
Kriging

Results from the kriging analyses are presented in the form of figures, as the prediction made by the semivariogram is output in a raster [image] layer in ESRI ArcMap and visualized in QGIS. Figure 21 shows the ordinary kriging results from the

analysis of New York, with county and census tract boundary lines. Based on the results, estimations for areas where latencies will be highest are in areas where, based on initial ESDA, we would expect poor stability: the Adirondack mountains, southwestern New York, and the most rural areas of the center of the state. Parts of areas that, anecdotally, we would think would experience low latency such as Orange

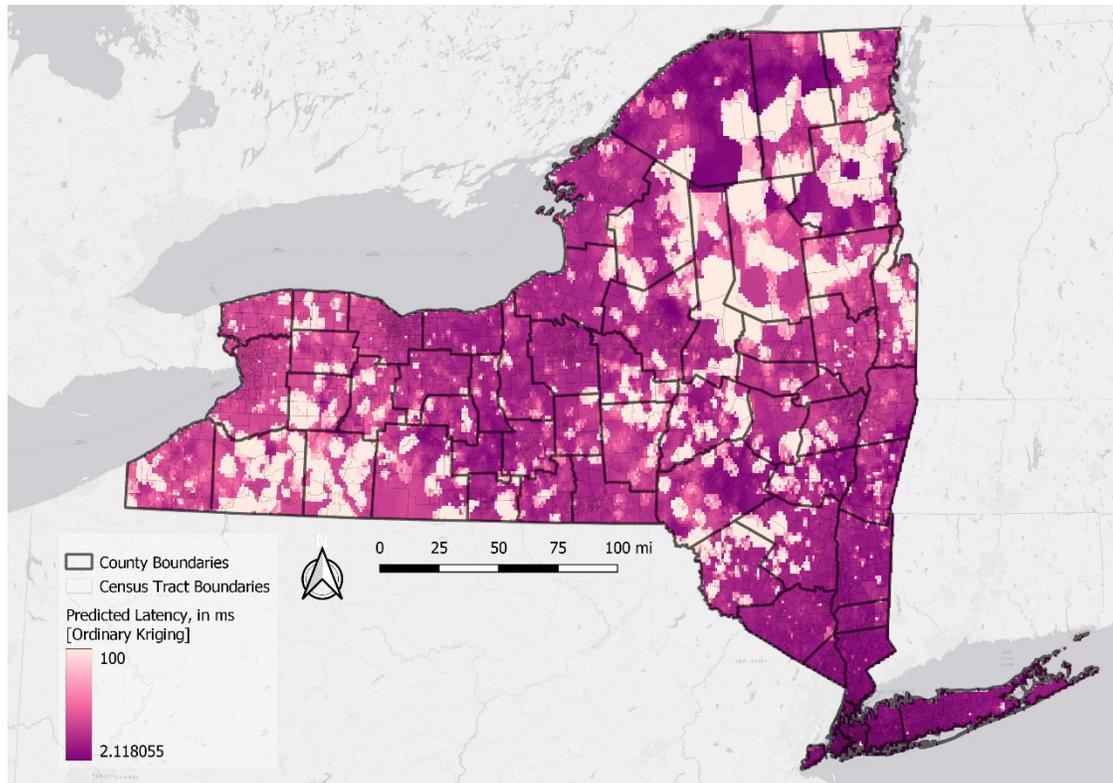


Figure 21 – Ordinary Kriging Raster Output: New York State, US

County [a relatively high-wealth, semi-urban area], and Saratoga county [a similarly wealthy semi-rural area] that have high-latency tiles on the fringes predict high latency raster landscapes. This is consistent with ordinary kriging, as it will predict higher latencies for areas that have fewer Ookla tile centroids, and by weighting the predictions to the nearest other tiles higher--which, often times, are also high latency--creates a noticeable ‘instability zone’ where a portion of the raster is highly unstable. It is important to mention that the raster prediction does have complete coverage,

while the point data from Ookla is limited to where the 1x1km tiles are located. In this way, ordinary kriging does provide at least a dataset of broader usability when seeking to explore unknown values.

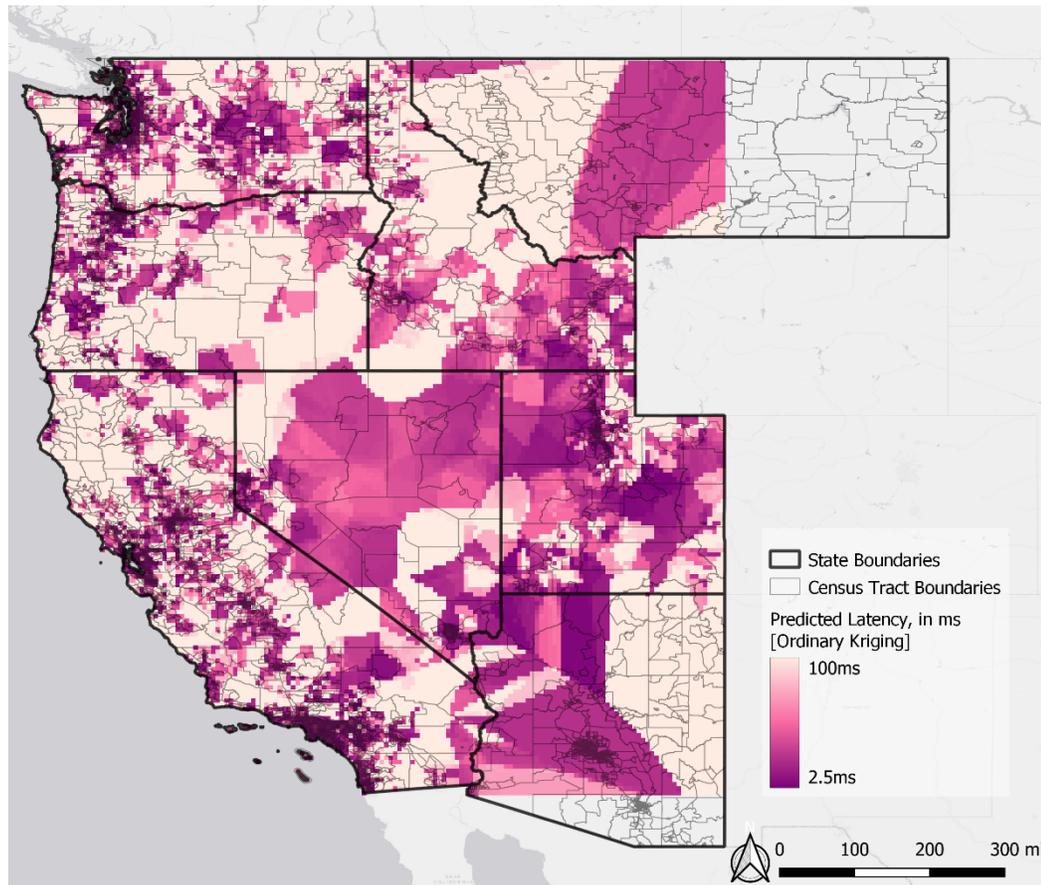
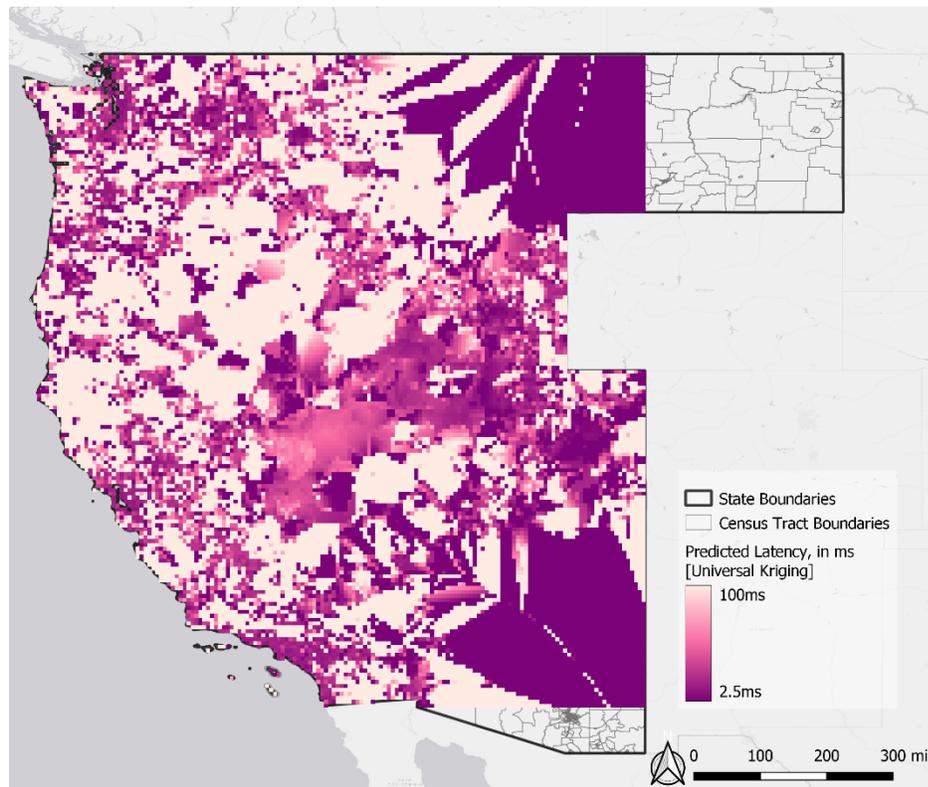


Figure 22 – Ordinary Kriging Raster Output: Western US

Observing the ordinary kriging results from the western US, shown in Figure 22, the patterns and usability are similar to the example of New York. Areas where the Ookla tiles either present high latency, or are sparse in number, or both have high latency and are physically disparate from other tiles, display as raster areas of high latency. Observing the middle of Nevada which is, for the most part, largely uninhabited due to military testing and Air Force ranges, we observe the ordinary kriging predicting high latencies with patches of mid latency. This is a result of the

nearest two or more Ookla point data being on both ends of the latency range: one or more is very high and one or more is very low. This does provide a ‘smooth’ landscape with which to observe latency, but it is not the complete picture.



*Figure 23 – Universal Kriging Raster Output, Western US*

When universal kriging is implemented, however, we can observe a noticeable difference in how the algorithm predicts latency across the same area of the western US, particularly in the sparsely-populated areas of Nevada. Figure 23 shows the universal kriging results of predictions on latency. Notice the overall enhanced specificity and granularity of the prediction, as well as the areas that are predicted to have the highest latencies. Areas such as the military testing and Air Force ranges have noticeably starker variation in their latency raster, something we may want when seeking to predict unknown values. I classify H3 as partially confirmed, and the

implications of the difference between the universal and ordinary kriging results are detailed in the discussion section.

### **Discussion**

By observing the global Moran's I measurement of latency across both the first and second order contiguities, hypothesis 1 is confirmed: latency is clustered amongst US census tracts in multiple iterations. Though, based on prima facie evidence alone, this may not be surprising, it is important to empirically confirm that tracts of similar latencies tend to cluster together, tying back to arguments of internet as a social resource and market-based effects (Bacher-Hicks, Goodman, and Mulhern 2021; Douthit et al. 2015; Francis and Weller 2022). More importantly, by categorizing certain clusters into neighborhood typologies, I offer a visual and empirically confirmed landscape of internet stability in more than a singular, tract-by-tract way. This augments literature surrounding spatial relationships of sociological phenomena, spatial analyses as a means to more thoroughly understand social science data, and reinforces my previous argument that rural and fringe locales experience worse stability than urban and most highly-populated tracts (Anselin et al. 2010; Chaney and Rojas-Guyler 2016; Fiduccia 2022; Grubestic 2006; Miller 1990). For policymakers seeking to identify areas for more efficient intervention, my typology of 'Regions of Opportunity' are census tracts of high latency surrounded by those of low latency; theoretically, this would be the lowest cost to implementing infrastructure or provision changes as both exist in direct spatial proximity to the tract with poor stability.

Using local regression modeling, specifically spatial error models, I have also confirmed my second hypothesis: that we can gain a more nuanced exploration of the predictors of latency by choosing local models over OLS. Non-spatial models predicted about 26% of the variance in latency, while first and second order spatial error models predicted ~38% of the variance. I have also noted, however, that this gain

in explanatory power comes with limitations and should be used in conjunction with other, more robust OLS methods in order to be as broadly methodologically sound as possible. Since some spatial models are currently incapable of simultaneously executing local neighborhood weights and analytic weights [as was performed in previous analyses for race-specific models], it constraints the predictive factors that can be included. Using multiple methods to explain a phenomenon, however, seems like a natural progression, as previous work tells us that we have to account for various populations, racial variance, economic factors, and social constructs when analyzing any social science variable of interest (Befort, Nazir, and Perri 2012; Chen and Mallory 2021; Diamond 2006; Hodson, Dovidio, and Gaertner 2002; Shedd 2015). Additionally, we know that the location of said populations, or individuals, is also of critical importance to understanding the broader landscape and proliferation of social science constructs that may be unequally distributed (Anselin et al. 2010; Bischoff 2008; Chaney and Rojas-Guyler 2016; Cliff et al. 1981; Reardon and Bischoff 2011). Therefore, I do not present the spatial model as a replacement for a scaled, clustered-robust OLS model, but local regression is crucial to understanding how relationships between variables of interest change when accounting for the fact that, in this case, census tracts do not have physical borders. Individuals and communities straddle and cross these conceptual ‘lines’ every single day, and to base analyses off of models that fail to account for this fluidity is not methodologically sound.

### **Kriging as a Tool for Social Science Research**

#### *Limitations*

My attempt to impute previously unknown values of latency across geographic units shows that both ordinary and universal kriging are useful in providing policy analysts and decision-makers another useful tool in the kit, but with limitations. First, using

universal kriging by default allows the range of the dependent variable of interest to be expanded beyond its original bounds. For certain metrics, change in annual income, for example, this wouldn't necessarily be a problem as the variable could have positive and negative values. For a metric like latency, however, where negative values cannot exist by definition, universal kriging providing negative values as an output to the raster model does not add specificity or nuance to the imputation landscape. This negates the added value of including covariates in the universal kriging model. Second, for this paper, I was limited by the technical capacity of the tools at hand. Using ArcMap and QGIS does allow for much more detailed ESDA but limits the empirical means by which you can specify the kriging models. Finally, as with any spatial interpolation, there must be a conceptual backing for why estimating the 'landscape' between points is both necessary and useful. In this instance, it fit both criteria: necessary because there were areas of tracts that did not include values from the raw Ookla data, and useful because the raster landscape can be used to ascertain whether populations in those areas experience good or bad latency without having to collect new data.

### *Refinements & Future Directions*

There are two ways by which we can further specify and refine both the ordinary and universal kriging models. In ordinary kriging, the model forming the raster landscape can be augmented by including additional point data, if available. For instance, if other years' data from Ookla were available for download, it could be added to the spatial program and used as an additional set of points from which to estimate the prediction landscape. In universal kriging, we can further specify the model by identifying covariates germane to the question at hand. In this project, for example, I would have liked to include geographically related covariates that may have an effect on predicted latency such as population density within the tract and/or distance from each point to

the nearest municipal centroid. There are also covariates that may influence latency on the broader, theoretical end that could be pulled from ACS data, such as average number of computing devices per household, percentage of population with a home computer, and/or proportion of the population working from home pre-pandemic.

The kriging analysis undertaken in this paper is but a first step in attempting to predict, map, and understand unknown values in latency. Based on recent research that has begun to analyze social science data using kriging techniques, it is a new and useful methodological tool to impute previously unknown values, specifically where data may be difficult to measure (Mooney et al. 2020). In the first example using New York state with ordinary kriging, it shows that a relatively regional-scale landscape can be interpolated, rasterized, and predicted to ascertain the unknown values of latency not present in the original Ookla data. This can be useful in ascertaining service areas or locales where interventions such as 5G service hotspots or wireless transmission infrastructure may be most useful. This usage can also be attributed to the example of ordinary kriging in the western US. When the western US example was analyzed using universal kriging, a more detailed and nuanced raster model emerged. In areas where there were no measured values and the algorithm interpreted new values, a more granular output was produced with universal kriging. This can further refine the identification of ‘instability zones’ or where exactly interventions may be most useful. As with any advanced statistical model, however, there is the risk that universal kriging can be overfit in some circumstances (Box 1976). Additionally, the mathematical construction of the universal kriging methodology allows for the predicted values to vastly exceed the original values minimums and maximums (Parmentier et al. 2011). This can be mitigated by specifically and carefully identifying the covariates, but I only had the technical software capacity to identify a single, universal covariate in this paper. Consequently, the output for universal kriging

had negative values of latency and latency predictions higher than 1000ms, both of which were suppressed for ease of visualization in the final raster layer.

### **Conclusion**

This paper uses a variety of spatial visualization and analysis tools to explore, specify, and predict internet instability over geographic space, specifically across census tracts in the United States. Preliminary analysis of the dataset showed that there was spatial autocorrelation present using global Moran's I. Local Indicators of Spatial Autocorrelation [LISA] analyses confirmed that latency, the variable of interest, is indeed significantly clustered across the nation. LISA measures then allowed me to identify and describe 'neighborhoods of instability' across the US, as well as create typologies for various degrees of internet stability across the defined spatial regions. Predicting latency using a spatial regression [error] model showed that we can explain more variance, but with limitations based on the mathematical differences between OLS and spatial models. OLS models allow for the inclusion of analytic and racial-category weights, as shown in Chapter 2, but spatial models do not. This, in most cases, is a trade-off worth making if only because in lieu of analytic or race weights, the specificity of the model is enhanced by the ability to include variance over space: the result of the spatial weighting matrix / neighborhood construction.

Despite the limitations, I show the value of local regression, provide researchers and policymakers a clear and defined portrait of areas of opportunity, disparity, stability, and instability regarding internet instability, and make a case for the use of kriging in the examination of unknown data measures. I also plan to augment this research by implementing kriging using R, in which the user-generated packages can offer options and interfaces for more tightly specifying covariates, a technique that can mitigate the range concern. Finally, I continue to implore scholars

across the academy to continue to use, and expand the use of, spatial methods to more thoroughly understand the complex policy landscape in social science.

---

## EPILOGUE

Over the course of the previous three chapters, I have explored multiple elements of internet stability across the US. Moreover, I hope to have provided policy analysts and decision makers with clearly identifiable geographic areas where latency is poor, populations who are particularly at risk for unstable internet, tools to employ for spatial analysis of latency, and methods by which unknown values can be estimated and mapped. Chapter one answers the question of ‘what’ and ‘where,’ chapter two explores ‘who,’ and chapter three begins to lay out a method by which we can ascertain ‘how’ to solve the stability disparity.

The implications in these chapters are understood considering traditional inequalities that face the populations highlighted here - particularly non-White, rural, and/or impoverished groups. Exploration of the literature in each chapter also speaks to the many underlying sociological explanations for why services and opportunity may be better in some areas, or for some populations, than others. This project also brings the terminology and understanding of what latency is, and how it influences two way connections, to the forefront of the literature and policy discussion. We have all heard the anecdotal evidence surrounding the difficulties of poor internet connection during remote work and learning. ‘Zoom fatigue’ is real, and for students sitting in a remote classroom trying to concentrate on an important mathematical formula, or participate in a literary reading, dealing with audio and video drops can not only be frustrating, but can lead to loss of progress and differential outcomes. We take stable two-way communication for granted: when we speak to one another in person, by and large our brains process audio and video in real time, resulting in an expectancy for certain conversational, physical, and emotional cues. These cues are subsequently missing when high latency causes a lapse in sync or a drop in connection. As mentioned at several points throughout the chapters, the locations and populations who are most

susceptible to poor internet stability are also those who already face additional challenges and inequities; adding high latency to the mix only makes a bad situation worse. The 7th grader in a rural locale trying to connect to a remote classroom in a house with only one computer, in a shared living room away from quiet and privacy, now may have to deal with the added difficulty of choppy or dropped connection. In the case of remote education, latency and its side effects quickly become an added cost of engagement to an already disengaged student.

The most interesting and nebulous exploration presented in this project is the imputation analyses offered by spatial kriging methods. There are both technical and conceptual considerations to be further understood, but being able to drop a ‘blanket’ of values across the area in question, thereby realizing potential values for previously unknown areas, is a useful tool no matter the phenomenon. Kriging is, however, another example of where quantitative methods and qualitative understanding must work together: having no knowledge of the underlying theoretical relationships between constructs, or simply the community makeup of a certain location, and blindly accepting the results of the kriging analysis is neither sound research or a useful result. Kriging relies on many highly complicated mathematical and geographic relationships, but the results will mean very little if they cannot be interpreted in context. In calling for additional refinement on covariates for universal kriging and latency, I recognize that these covariates may not be easily-measurable or already-defined quantitative measures and, in fact, may need to be based in the world of policy implementation, survey results, or other more qualitatively-leaning information.

As researchers seek to challenge, refine, and exceed the findings I’ve presented in these chapters, I share several areas of exploration that I tabled in the hopes of finding simplicity and efficiency in my offering. For those more technically minded, there is an opportunity to explore the methods by which Ookla collects and aggregates their data.

This remained a black box for the entirety of my project, but may be revealed in the near future. In either case, longitudinal analysis is also a strong recommendation: understanding the change in latency over time, combined with known changes in infrastructure or telecommunications policy, may provide deeper insight into the disparities between racial and geographic groups. Demographic decomposition analyses can provide more insight into the exact nature of state-level policy decisions with regard to latency, particularly by urban and rural distinctions. Economic analyses surrounding the demand for internet services over time may also provide a lens through which market effects of latency disparities can be better understood. Finally, simply changing the spatial unit of analysis may enable researchers to determine other impacts of latency I have not discovered here: congressional district boundaries, US Census place polygons, or even legislative shapefiles. In any or all cases, I hope this project provides researchers and others the literature, empirical tools, and exploratory data visualizations needed to bring latency / internet stability to the forefront of policy discussions. As we continue to rely on technology to connect our world, and the varied populations therein, it becomes increasingly more evident that it must be equitably distributed as well as critically maintained. Just as critical to participation in our global society as having access to the internet is, so, too, is being able to rely on a stable connection to retain the ability to participate.

## TECHNICAL APPENDIX

This paper utilizes established quantitative statistical techniques to perform descriptive and predictive analyses on the dataset. Summarization operations are used to describe single variable characteristics. To ascertain dyadic relationships between variables, pairwise correlation operations are employed. Two-sample t-tests as well as standard analysis of variance are utilized to ascertain differences in means between categorical samples of the data, namely rural and non-rural census tracts. To attain prediction effects on the dependent variable of interest, ordinary least squares regression is employed with post-regression diagnostics to test for normality. Since the question of regionality is also at play, a clustered robust fixed-effect model is also utilized which groups at the state level. Post-regression marginal analysis is performed to both act as a robustness check for the model as well as account for any nonlinear relationships. At all stages of quantitative analysis, robustness and sample-validity checks are performed using variance comparison tests, robust equal-variance tests, and Levene's tests using a random sample at the 10% and 1% levels - all of which returned statistically significant. Local variation is examined through the usage of various spatial analysis techniques. Broadly, descriptive mapping, or data cartography, is used to produce maps of a state, district, or many districts with choropleth visualizations; a map in which geographically bounded areas are shaded based on the variability in the variable of interest. Descriptive mapping is meant to visually represent data on a scale which provides a geographically comparative perspective. Furthermore, global spatial autocorrelation (Jerrett, Gale, and Kontgis 2010; Legendre 1993; Moran 1948) is used to determine whether a large, significant, underlying pattern is present in data spanning a set unit of analysis.

## EMPIRICAL APPENDICES

*Appendix Table A-1*

<b>OLS Regression Model</b>	<b>Effect on Average Latency(ms)</b>		
<i>Predictors</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	11.5	10.43 – 12.66	<b>&lt;0.001</b>
Rural Tract	72.19	69.67 – 74.70	<b>&lt;0.001</b>
% School-Age Pop	-5.35	-11.14 – 21.84	<b>0.525</b>
Int: Ruralx%School-Age	1228	1061 – 1395	<b>&lt;0.001</b>
Midwest Region	8.12	7.15 – 9.08	<b>&lt;0.001</b>
South Region	13.21	12.31 – 14.10	<b>&lt;0.001</b>
West Region	10.57	9.59 – 11.56	<b>&lt;0.001</b>
Observations	70,642		
R <sup>2</sup>	0.2326		

Appendix Table A-2

Tract Type	# of Obs	Mean Latency (in ms)
Non-Rural	57,169	20.4
Rural	13,924	78.8

P < 0.001

Appendix Table A-3: OLS Regression Output (GeoDa)

```

-----
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set          : Paper3_Shapefile
Dependent Variable : Lat_Mean  Number of Observations:70868
Mean dependent var : 32.2026  Number of Variables   : 13
S.D. dependent var : 47.8505  Degrees of Freedom    :70855

R-squared        : 0.271173  F-statistic           : 2196.9
Adjusted R-squared : 0.271049  Prob(F-statistic)     : 0
Sum squared residual:1.18263e+008  Log likelihood        : -363472
Sigma-square     : 1669.08   Akaike info criterion : 726970
S.E. of regression : 40.8544   Schwarz criterion     : 727089
Sigma-square ML  : 1668.77
S.E of regression ML: 40.8506

-----
Variable          Coefficient      Std.Error      t-Statistic      Probability
-----
CONSTANT          15.0087          0.657009       22.8439          0.00000
_povertyra        11.0667          1.5255         7.25449          0.00000
_pctschool        -9.65089         2.54208        -3.79646         0.00015
_region2          7.97656          0.486625       16.3916          0.00000
_region3          10.4659          0.456028       22.9502          0.00000
_region4          10.4566          0.511931       20.4259          0.00000
NHBlack           -0.000470093     0.000199383    -2.35773         0.01842
NHAmInd           0.00680278       0.00102761     6.61999          0.00000
NHAsian           -0.00467096      0.000395299    -11.8163         0.00000
HispTotal         -0.00052505      0.000156663    -3.35146         0.00081
_ruraltrac        34.8655          0.75616        46.1086          0.00000
_majnrur          28.7856          2.04797        14.0557          0.00000
INT_rur           159.226          5.09213        31.269           0.00000
-----

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER  12.111323
TEST ON NORMALITY OF ERRORS
TEST          DF          VALUE          PROB
Jarque-Bera   2          19564375.1947  0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST          DF          VALUE          PROB
Breusch-Pagan test  12          153988.7430   0.00000
Koenker-Bassett test  12          3740.0813     0.00000
===== END OF REPORT =====

```

Appendix Table A-4: Spatial Error Regression [Queens 2nd Order] Output (GeoDa)

```

-----
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set          : Paper3_Shapefile
Spatial Weight    : Paper3_Queens_Second
Dependent Variable : Lat_Mean  Number of Observations:70868
Mean dependent var : 32.202617 Number of Variables   : 13
S.D. dependent var : 47.850487 Degrees of Freedom   :70855
Lag coeff. (Lambda) : 0.640262

R-squared        : 0.375084 R-squared (BUSE)      : -
Sq. Correlation  : -      Log likelihood      : -358984.061680
Sigma-square     : 1430.85 Akaike info criterion : 717994
S.E of regression : 37.8266 Schwarz criterion   : 718113

-----
Variable          Coefficient      Std.Error      z-value      Probability
-----
CONSTANT          19.921           1.0962         18.1729      0.00000
_povertyra        1.57391          1.73338        0.908003     0.36388
_pctschool        -16.4539         2.75476        -5.97288     0.00000
_region2          7.7172           1.21376        6.35811      0.00000
_region3          9.58115          1.12295        8.53211      0.00000
_region4          5.89714          1.26452        4.66355      0.00000
NHBlack          -0.000585532     0.00024959     -2.34598     0.01898
NHAmInd          0.00219704       0.0012234      1.79585      0.07252
NHAsian          -0.00239461      0.000503041    -4.76027     0.00000
HispTotal        -0.00063735      0.000210704    -3.02486     0.00249
_ruraltrac       28.55            0.779053       36.6471      0.00000
_majnrur         22.8508          2.03981        11.2024      0.00000
INT_rur          111.839          5.19695        21.5201      0.00000
LAMBDA           0.640262         0.00659629     97.064       0.00000
-----

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST
Breusch-Pagan test          DF      VALUE      PROB
                            12      159974.8907  0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : Paper3_Queens_Second
TEST
Likelihood Ratio Test      DF      VALUE      PROB
                            1       8976.0429  0.00000
===== END OF REPORT =====

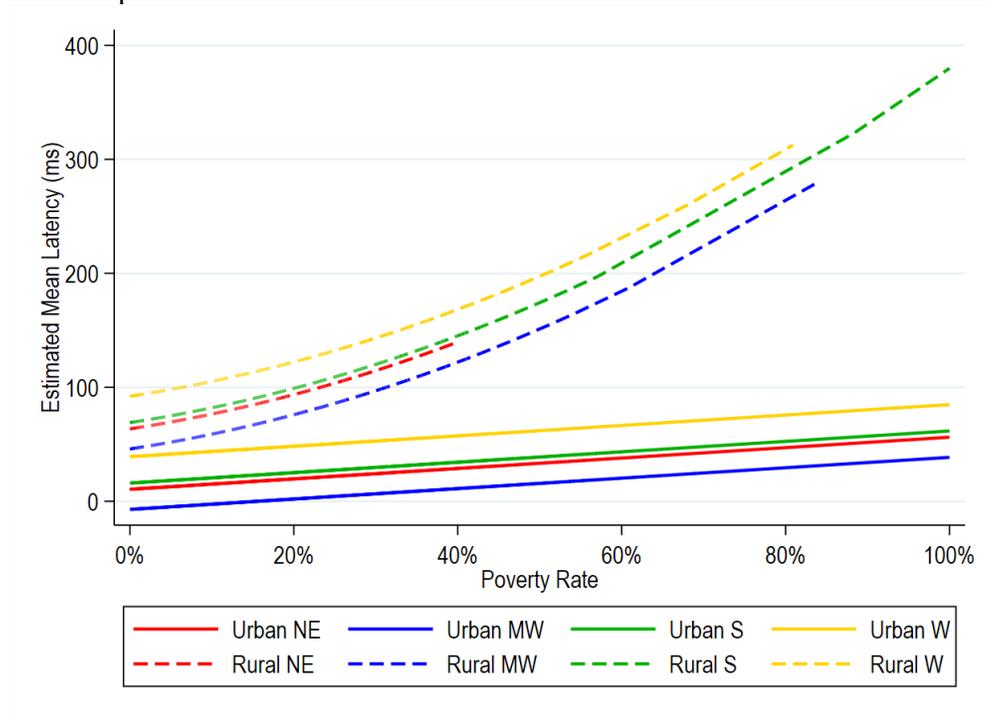
```

Appendix Table A-5

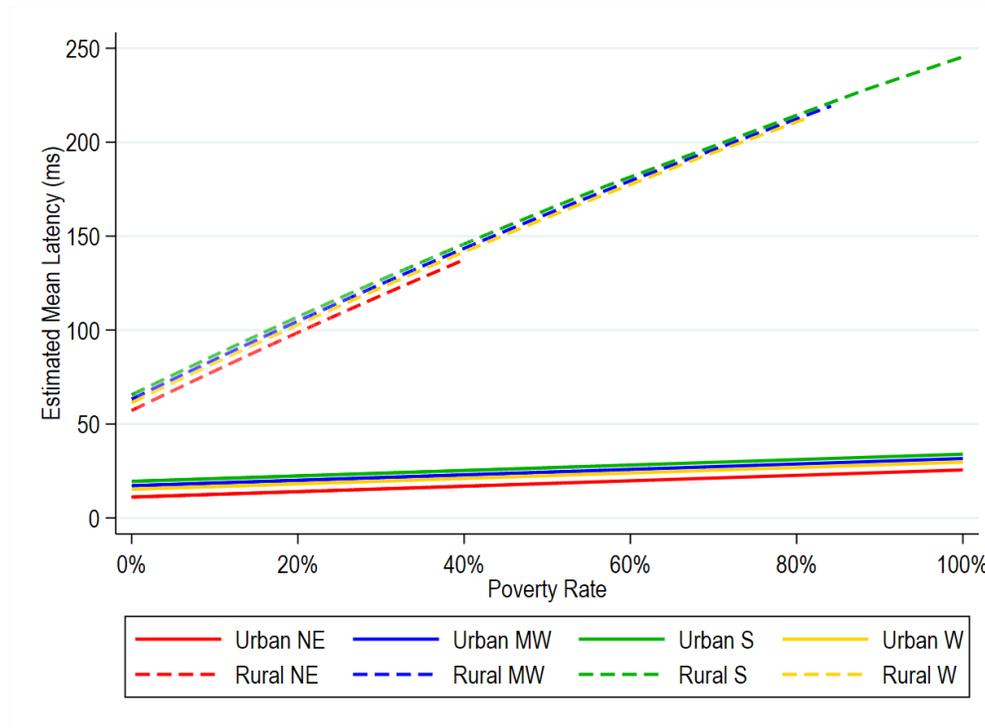
	<b><u>Region</u></b>				
	Northeast	Midwest	South	West	Total
<b><u>Cluster Typology</u></b>					
Regions of Unstable Connection	443	2,464	4,709	1,756	9,372
Regions of Stable Connection	8,364	4,846	8,476	7,406	29,092
Regions of Disparity	120	967	1,790	830	3,707
Regions of Opportunity	98	110	135	166	509
<i>Not Significant</i>	<i>3,834</i>	<i>8,479</i>	<i>10,874</i>	<i>5,226</i>	<i>28,413</i>
Total	12,859	16,866	25,984	15,384	71,093

*Race-Analytic Weighted Predicted Marginal Effects on Latency (Poverty Rate)*

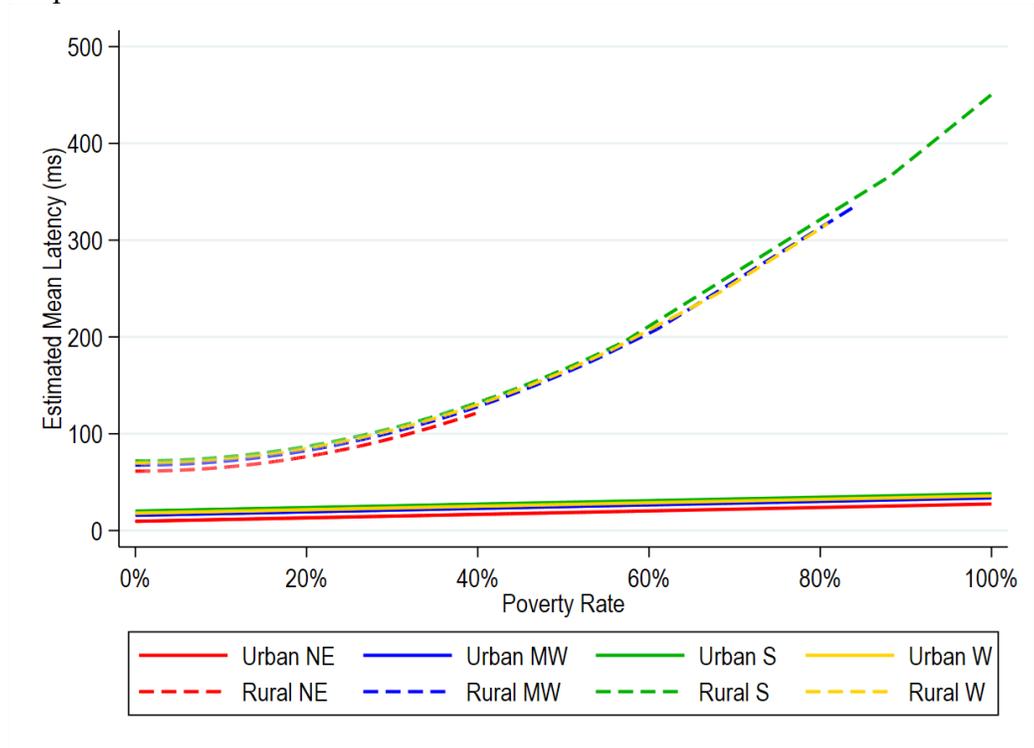
Non-Hispanic American Indian



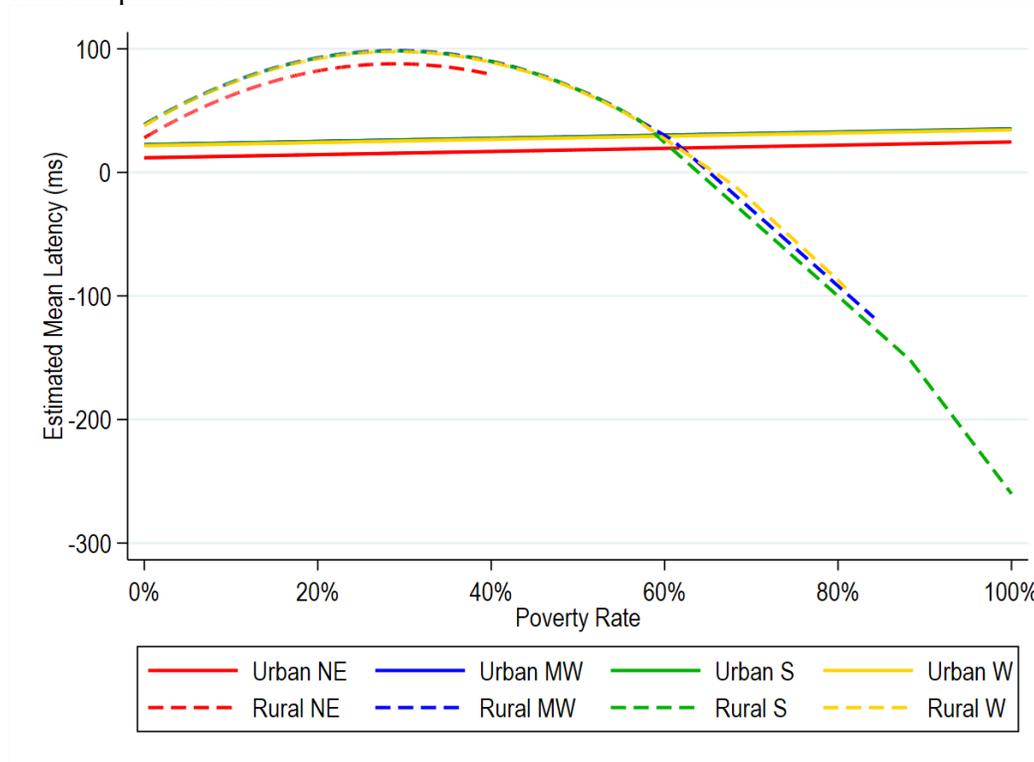
Non-Hispanic Black



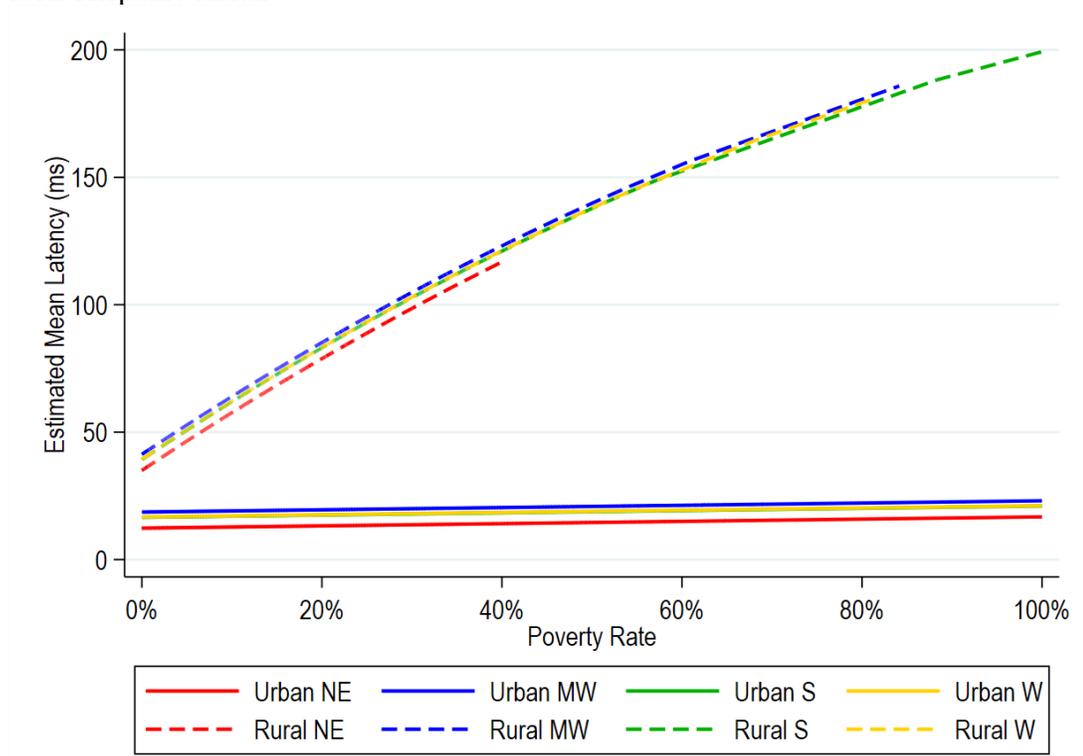
### Hispanic / Latino



### Non-Hispanic White

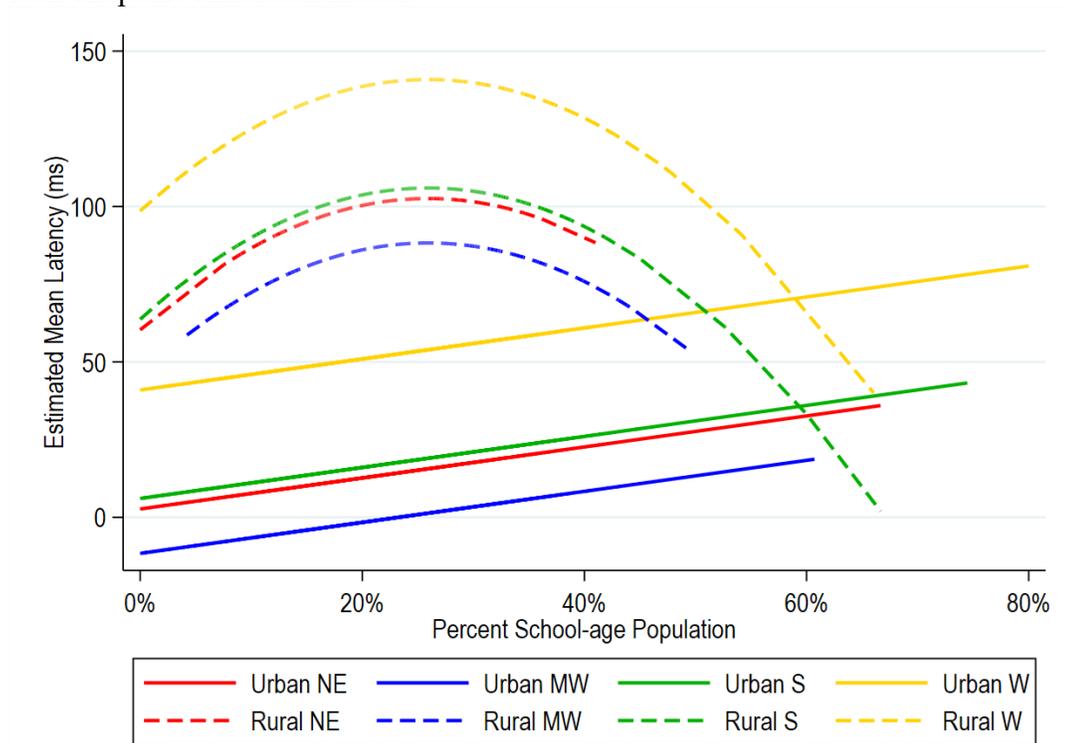


### Non-Hispanic Asian

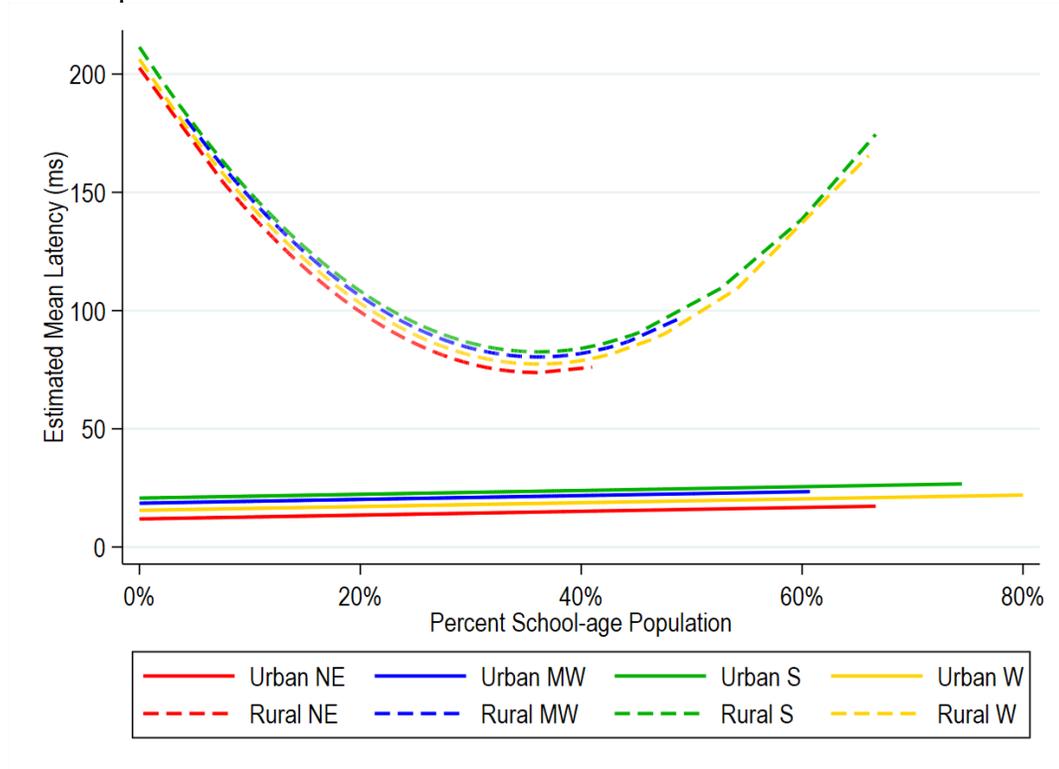


*Race-Analytic Weighted Predicted Marginal Effects on Latency (% School-Age Population)*

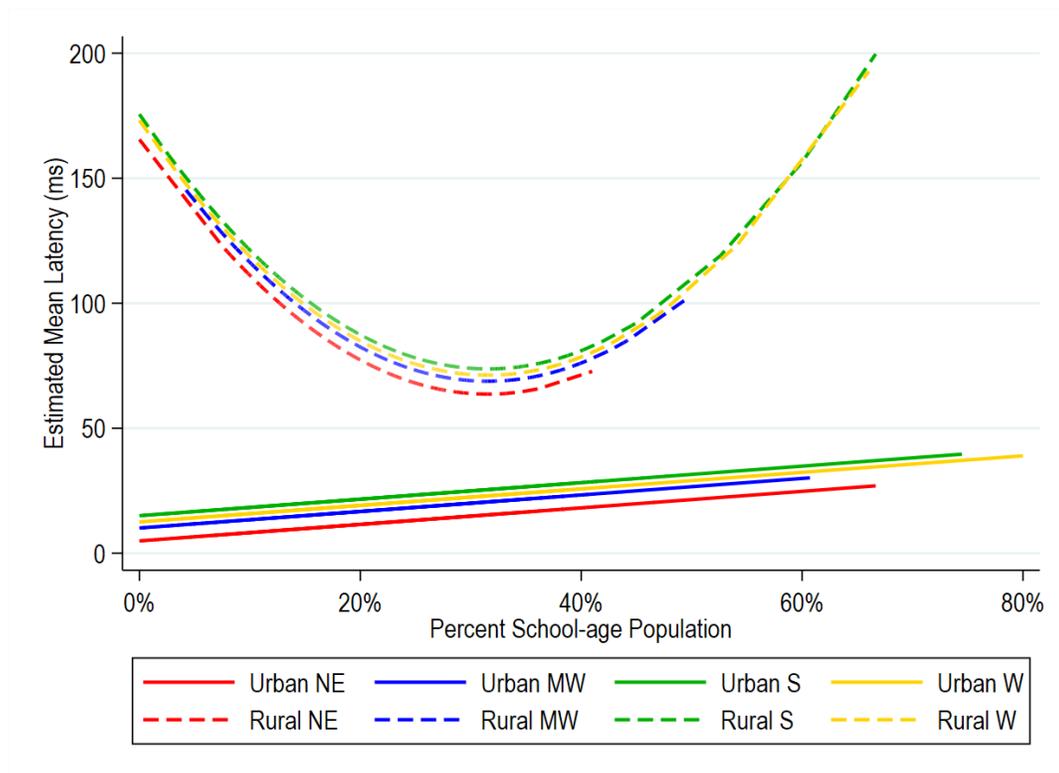
### Non-Hispanic American Indian



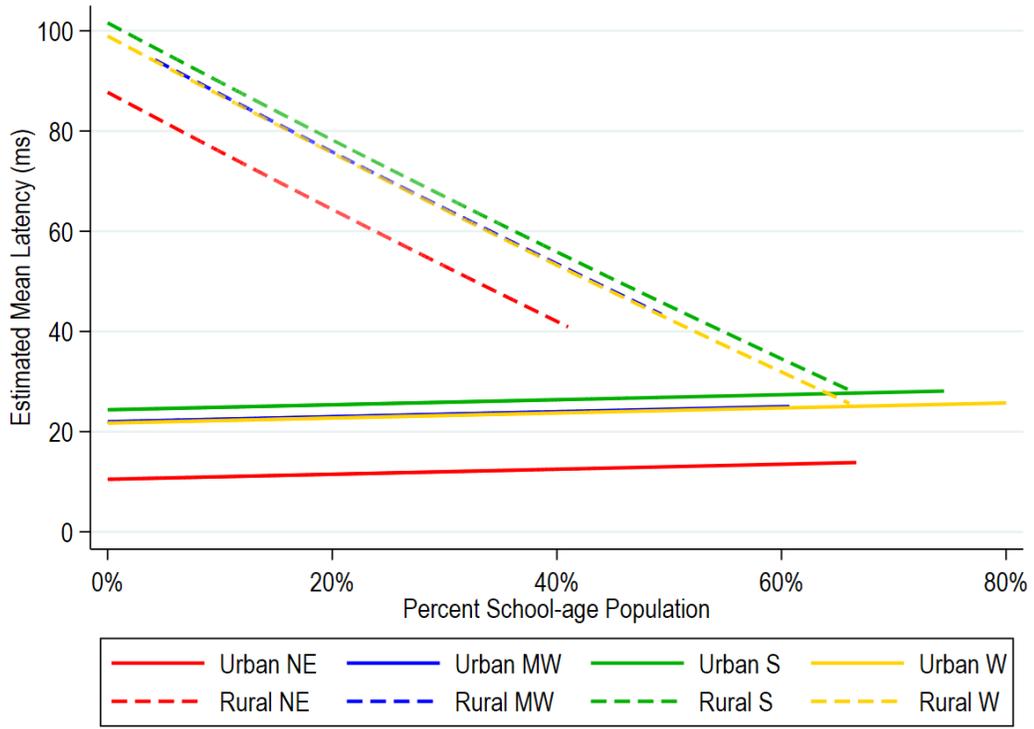
### Non-Hispanic Black



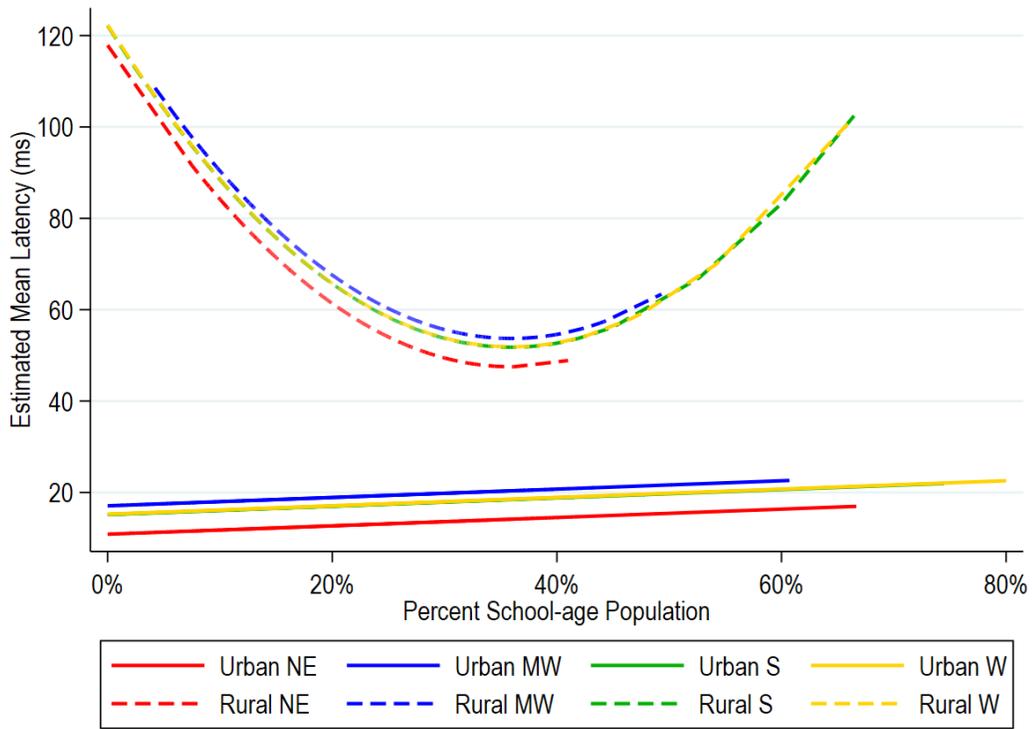
### Hispanic / Latino



### Non-Hispanic White



### Non-Hispanic Asian



## BIBLIOGRAPHY

- Allington, Richard L., and Anne McGill-Franzen. 2003. "The Impact of Summer Setback on the Reading Achievement Gap." *Phi Delta Kappan* 85(1):68–75.
- Alter, Theodore, Jeffrey Bridger, Sheila Sager, Kai Schafft, and William Shuffstall. 2007. "Getting Connected."
- Anselin, Luc, Ibnu Syabri, and Youngihn Kho. 2010. "GeoDa: An Introduction to Spatial Data Analysis." Pp. 73–89 in *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, edited by M. M. Fischer and A. Getis. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Anselin, Luc. 1995. "Local Indicators of Spatial association—LISA." *Geographical Analysis* 27(2):93–115.
- Atske, S., & Perrin, A. (2021). *Home broadband adoption, computer ownership vary by race, ethnicity in the US Pew Research Center*.
- Auchincloss, Amy H., Ana V. Diez Roux, Daniel G. Brown, Trivellore E. Raghunathan, and Christine A. Erdmann. 2007. "Filling the Gaps: Spatial Interpolation of Residential Survey Data in the Estimation of Neighborhood Characteristics." *Epidemiology* 18(4):469–78.
- Bacher-Hicks, Andrew, Joshua Goodman, and Christine Mulhern. 2021. "Inequality in Household Adaptation to Schooling Shocks: Covid-Induced Online Learning Engagement in Real Time." *Journal of Public Economics* 193:104345.
- Bader, Michael D. M., and Jennifer A. Ailshire. 2014. "CREATING MEASURES OF THEORETICALLY RELEVANT NEIGHBORHOOD ATTRIBUTES AT MULTIPLE SPATIAL SCALES." *Sociological Methodology* 44(1):322–68.
- Bass, Bernard M., and Bruce J. Avolio. 1994. *Improving Organizational Effectiveness Through Transformational Leadership*. SAGE.

- Befort, Christie A., Niaman Nazir, and Michael G. Perri. 2012. "Prevalence of Obesity among Adults from Rural and Urban Areas of the United States: Findings from NHANES (2005-2008)." *The Journal of Rural Health: Official Journal of the American Rural Health Association and the National Rural Health Care Association* 28(4):392–97.
- Belinfante, Alexander. 2001. "Telephone Subscribership in the United States." *Federal Communications Commission, Industry Analysis Division*.
- Bischoff, Kendra. 2008. "School District Fragmentation and Racial Residential Segregation: How Do Boundaries Matter?" *Urban Affairs Review* 44(2):182–217.
- Black, Edward J., Catherine R. Sloan, Daniel O'Connor, John T. Nakahata, Brita D. Strandberg, and Kelley A. Shields. 2009. "Federal Communications Commission."
- Blank, G., & Dutton, W. H. (2014). Next Generation Internet Users. In *Society and the Internet* (pp. 36–52). <https://doi.org/10.1093/acprof:oso/9780199661992.003.0003>
- Blank, G., & Grosej, D. (2015). Digital Divide| Examining Internet Use Through a Weberian Lens. *International Journal of Communication Systems*, 9(0), 21. <https://ijoc.org/index.php/ijoc/article/view/3114>
- Blank, Grant, and Darja Grosej. 2015. "Digital Divide| Examining Internet Use Through a Weberian Lens." *International Journal of Communication Systems* 9(0):21.
- Bourdieu, P., & Passeron, J.-C. (1977). *Reproduction in education, culture and society*. London: Sage.
- Box, George E. P. 1976. "Science and Statistics." *Journal of the American Statistical Association* 71(356):791–99.
- Cairncross, Frances. 2002. "The Death of Distance." *RSA Journal* 149(5502):40–42.
- Census Bureau, U. S. (2004). *American FactFinder*. U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau. <https://play.google.com/store/books/details?id=CXXgh95hEhsC>

- Chaney, Robert A., and Liliana Rojas-Guyler. 2016. "Spatial Analysis Methods for Health Promotion and Education." *Health Promotion Practice* 17(3):408–15.
- Chaudhuri, Anindya, Kenneth S. Flamm, and John Horrigan. 2005. "An Analysis of the Determinants of Internet Access." *Telecommunications Policy* 29(9):731–55.
- Chen, Shanting, and Allen B. Mallory. 2021. "The Effect of Racial Discrimination on Mental and Physical Health: A Propensity Score Weighting Approach." *Social Science & Medicine* 285:114308.
- Chiariotti, Federico, Stepan Kucera, Andrea Zanella, and Holger Claussen. 2019. "Analysis and Design of a Latency Control Protocol for Multi-Path Data Delivery With Pre-Defined QoS Guarantees." *IEEE/ACM Transactions on Networking* 27(3):1165–78.
- Cliff, Andrew David, J. K. Ord, P. Haggett, Emeritus Professor in the Department of Geography P Haggett, and G. R. Versey. 1981. *Spatial Diffusion: An Historical Geography of Epidemics in an Island Community*. CUP Archive.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2008). Scaling the digital divide: Home computer technology and student achievement. *Education Policy Colloquia Series, Harvard University, Cambridge, MA. Retrieved August, 19, 2009.*  
<https://edutechdebate.org/wp-content/uploads/2010/07/computers-north-carolina.pdf>
- Clougherty, Jane E., Iyad Kheirbek, Holger M. Eisl, Zev Ross, Grant Pezeshki, John E. Gorczynski, Sarah Johnson, Steven Markowitz, Daniel Kass, and Thomas Matte. 2013. "Intra-Urban Spatial Variability in Wintertime Street-Level Concentrations of Multiple Combustion-Related Air Pollutants: The New York City Community Air Survey (NYCCAS)." *Journal of Exposure Science & Environmental Epidemiology* 23(3):232–40.
- Communities, Vibrant, John B. Horrigan, and Ellen Satterwhite. 2012. "STATE BROADBAND INDEX."
- Conley, P. A., & Hamlin, M. L. (2009). Justice-learning: Exploring the efficacy with low-

- income, first-generation college students. *Michigan Journal of Community Service Learning*, 16(1), 47–58. <https://eric.ed.gov/?id=EJ888073>
- Cooper, Harris, Barbara Nye, Kelly Charlton, James Lindsay, and Scott Greathouse. 1996. “The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review.” *Review of Educational Research* 66(3):227–68.
- Cressie, Noel. 1988. “Spatial Prediction and Ordinary Kriging.” *Mathematical Geology* 20(4):405–21.
- Crowder, Kyle, and Scott J. South. 2011. “SPATIAL AND TEMPORAL DIMENSIONS OF NEIGHBORHOOD EFFECTS ON HIGH SCHOOL GRADUATION.” *Social Science Research* 40(1):87–106.
- Cutler, R., Rui, Y., Gupta, A., Cadiz, J. J., Tashev, I., He, L.-W., Colburn, A., Zhang, Z., Liu, Z., & Silverberg, S. (2002). Distributed meetings: a meeting capture and broadcasting system. *Proceedings of the Tenth ACM International Conference on Multimedia*, 503–512. <https://doi.org/10.1145/641007.641112>
- Cutler, Ross, Yong Rui, Anoop Gupta, J. J. Cadiz, Ivan Tashev, Li-Wei He, Alex Colburn, Zhengyou Zhang, Zicheng Liu, and Steve Silverberg. 2002. “Distributed Meetings: A Meeting Capture and Broadcasting System.” Pp. 503–12 in *Proceedings of the tenth ACM international conference on Multimedia, MULTIMEDIA '02*. New York, NY, USA: Association for Computing Machinery.
- Daniel Xu, H. 2020. “Health Challenges for Rural Families: Issues, Policies, and Solutions.” Pp. 1–22 in *Handbook of Research on Leadership and Advocacy for Children and Families in Rural Poverty*. IGI Global.
- Darity, W. A., Sharpe, R. V., & Swinton, O. H. (2009). *The state of blacks in higher education*. <https://mpr.ub.uni-muenchen.de/id/eprint/34411>
- DeBell, M., & Chapman, C. (2006). Computer and Internet use by students in 2003. Statistical analysis report. NCES 2006-065. *National Center for Education Statistics*.

<https://eric.ed.gov/?id=ED493283>

Diamond, John B. 2006. "Still Separate and Unequal: Examining Race, Opportunity, and School Achievement in 'Integrated' Suburbs." *The Journal of Negro Education* 75(3):495–505.

DiMaggio, P., Hargittai, E., Neuman, W. R., & Robinson, J. P. (2001). Social Implications of the Internet. *Annual Review of Sociology*, 27(1), 307–336.

<https://doi.org/10.1146/annurev.soc.27.1.307>

DiMaggio, Paul, Eszter Hargittai, W. Russell Neuman, and John P. Robinson. 2001.

"Social Implications of the Internet." *Annual Review of Sociology* 27(1):307–36.

Doelling, David R., Norman G. Loeb, Dennis F. Keyes, Michele L. Nordeen, Daniel

Morstad, Cathy Nguyen, Bruce A. Wielicki, David F. Young, and Moguo Sun. 2013.

"Geostationary Enhanced Temporal Interpolation for CERES Flux Products." *Journal of Atmospheric and Oceanic Technology* 30(6):1072–90.

Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2020). *COVID-19 and student learning in the United States: The hurt could last a lifetime*. McKinsey & Company.

Douthit, N., S. Kiv, T. Dwolatzky, and S. Biswas. 2015. "Exposing Some Important

Barriers to Health Care Access in the Rural USA." *Public Health* 129(6):611–20.

Downes, Tom, and Shane Greenstein. 2002. "Universal Access and Local Internet Markets in the US." *Research Policy* 31(7):1035–52.

Drabenstott, M. (2001). New Policies for a New Rural America. *International Regional Science Review*, 24(1), 3–15. <https://doi.org/10.1177/016001701761012962>

Drake, Coleman, Yuehan Zhang, Krisda H. Chaiyachati, and Daniel Polsky. 2019. "The

Limitations of Poor Broadband Internet Access for Telemedicine Use in Rural

America: An Observational Study." *Annals of Internal Medicine* 171(5):382–84.

Faber, J. W., & Sharkey, P. (2015). Neighborhood Effects. In *International Encyclopedia*

*of the Social & Behavioral Sciences* (pp. 443–449). <https://doi.org/10.1016/b978-0-08->

097086-8.32189-4

- Fiduccia, P. C. (2022). *Life Through Internet Lag* (J. W. Sipple (ed.)) [PhD]. Cornell University.
- Fish, R. S., Kraut, R. E., Root, R. W., & Rice, R. E. (1992). Evaluating video as a technology for informal communication. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 37–48. <https://doi.org/10.1145/142750.142755>
- Fish, Robert S., Robert E. Kraut, Robert W. Root, and Ronald E. Rice. 1992. “Evaluating Video as a Technology for Informal Communication.” Pp. 37–48 in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '92*. New York, NY, USA: Association for Computing Machinery.
- Fishman, Aryeh B. 2021. “FEDERAL COMMUNICATIONS COMMISSION.”
- Fotheringham, A. Stewart, Martin E. Charlton, and Chris Brunsdon. 2001. “Spatial Variations in School Performance: A Local Analysis Using Geographically Weighted Regression.” *Geographical and Environmental Modelling* 5(1):43–66.
- Fotheringham, A. Stewart. 2009. “‘The Problem of Spatial Autocorrelation’ and Local Spatial Statistics.” *Geographical Analysis* 41(4):398–403.
- Francis, D. V., & de Oliveira, A. C. M. (2019). Do school counselors exhibit bias in recommending students for advanced coursework? *The B.E. Journal of Economic Analysis & Policy*. <https://www.degruyter.com/document/doi/10.1515/bejeap-2018-0189/html>
- Francis, Dania V., and Christian E. Weller. 2022. “Economic Inequality, the Digital Divide, and Remote Learning During COVID-19.” *The Review of Black Political Economy* 49(1):41–60.
- García, Emma, and Elaine Weiss. 2020. “COVID-19 and Student Performance, Equity, and U.s. Education Policy: Lessons from Pre-Pandemic Research to Inform Relief, Recovery, and Rebuilding.” *Economic Policy Institute*.

- Geelhoed, E., Parker, A., Williams, D. J., & Groen, M. (2009). Effects of latency on telepresence. *HP Labs Technical Report: HPL-2009-120*.  
<http://shiftright.com/mirrors/www.hpl.hp.com/techreports/2009/HPL-2009-120.pdf>
- Goolsbee, A., & Guryan, J. (2006). The impact of Internet subsidies in public schools. *The Review of Economics and Statistics*. <https://direct.mit.edu/rest/article-abstract/88/2/336/57588>
- Goudeau, S., Sanrey, C., Stanczak, A., Manstead, A., & Darnon, C. (2021). Why lockdown and distance learning during the COVID-19 pandemic are likely to increase the social class achievement gap. In *Nature Human Behaviour* (Vol. 5, Issue 10, pp. 1273–1281). <https://doi.org/10.1038/s41562-021-01212-7>
- Gould, E., Perez, D., & Wilson, V. (n.d.). *Latinx workers—particularly women—face devastating job losses in the COVID-19 recession*. Retrieved May 13, 2022, from <https://policycommons.net/artifacts/1409545/latinx-workers-particularly-women-face-devastating-job-losses-in-the-covid-19-recession/2023810/>
- Grubestic, Tony H. 2006. “A Spatial Taxonomy of Broadband Regions in the United States.” *Information Economics and Policy* 18(4):423–48.
- Grubestic, Tony H., and Alan T. Murray. 2002. “Constructing the Divide: Spatial Disparities in Broadband Access.” *Papers in Regional Science: The Journal of the Regional Science Association International* 81(2):197–221.
- Gulson, Kalervo N., and Colin Symes. 2007. *Spatial Theories of Education: Policy and Geography Matters*. Routledge.
- Gulson, Kalervo N., and Colin Symes. 2007. *Spatial Theories of Education: Policy and Geography Matters*. Routledge.
- Haack, Catherine, and Pierre Lefebvre. 2020. “Pandemic School Closures May Increase Inequality in Test Scores.” *Canadian Public Policy. Analyse de Politiques* 46(S1):S82–87.

- Hanks, A., Solomon, D., & Weller, C. E. (2018). Systematic inequality: How America's structural racism helped create the black-white wealth gap. *Center for American Progress, 21*.
- Hargittai, E. (2004). Internet Access and Use in Context. *New Media & Society, 6*(1), 137–143. <https://doi.org/10.1177/1461444804042310>
- Hargittai, Eszter. 2004. "Internet Access and Use in Context." *New Media & Society 6*(1):137–43.
- Harris, C., Straker, L., & Pollock, C. (2017). A socioeconomic related "digital divide" exists in how, not if, young people use computers. *PloS One, 12*(3), e0175011. <https://doi.org/10.1371/journal.pone.0175011>
- Hassani, S. N. (2006). Locating digital divides at home, work, and everywhere else. *Poetics, 34*(4), 250–272. <https://doi.org/10.1016/j.poetic.2006.05.007>
- Helsper, E. J. (2012). A Corresponding Fields Model for the Links Between Social and Digital Exclusion. *Communication Theory: CT: A Journal of the International Communication Association, 22*(4), 403–426. <https://doi.org/10.1111/j.1468-2885.2012.01416.x>
- Hemphill, F. C., & Vanneman, A. (2011). Achievement Gaps: How Hispanic and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress. Statistical Analysis Report. NCES 2011-459. *National Center for Education Statistics*. <https://eric.ed.gov/?id=ED520960>
- Hengl, Tomislav, Gerard B. M. Heuvelink, and Alfred Stein. 2004. "A Generic Framework for Spatial Prediction of Soil Variables Based on Regression-Kriging." *Geoderma 120*(1):75–93.
- Hobbs, T. D., & Hawkins, L. (2020). The results are in for remote learning: It didn't work. *The Wall Street Journal*. [https://www.panoramic.com/wp-content/uploads/2020/06/The-Results-Are-In-for-Remote-Learning\\_-It-](https://www.panoramic.com/wp-content/uploads/2020/06/The-Results-Are-In-for-Remote-Learning_-It-)

Didn%E2%80%99t-Work-WSJ.pdf

- Hodson, Gordon, John F. Dovidio, and Samuel L. Gaertner. 2002. "Processes in Racial Discrimination: Differential Weighting of Conflicting Information." *Personality & Social Psychology Bulletin* 28(4):460–71.
- Hoek, Gerard, Rob Beelen, Kees de Hoogh, Danielle Vienneau, John Gulliver, Paul Fischer, and David Briggs. 2008. "A Review of Land-Use Regression Models to Assess Spatial Variation of Outdoor Air Pollution." *Atmospheric Environment* 42(33):7561–78.
- Hogrebe, Mark C., and William F. Tate. 2012. "Geospatial Perspective: Toward a Visual Political Literacy Project in Education, Health, and Human Services." *Review of Research in Education* 36(1):67–94.
- Hollman, A. K., Obermier, T. R., & Burger, P. R. (2021). Rural Measures: A Quantitative Study of The Rural Digital Divide. *Journal of Government Information: An International Review of Policy, Issues and Resources*, 11, 176–201.  
<https://doi.org/10.5325/jinfopoli.11.2021.0176>
- Hollman, Angela K., Timothy R. Obermier, and Paul R. Burger. 2021. "Rural Measures: A Quantitative Study of The Rural Digital Divide." *Journal of Government Information: An International Review of Policy, Issues and Resources* 11:176–201.
- Hovardas, T. (2016). Primary school teachers and outdoor education: Varying levels of teacher leadership in informal networks of peers. *The Journal of Environmental Education*, 47(3), 237–254. <https://doi.org/10.1080/00958964.2015.1113155>
- Howard, P. N., Busch, L., & Sheets, P. (2010). Comparing digital divides: Internet access and social inequality in Canada and the United States. *Canadian Journal of Communication*, 35(1). <https://www.academia.edu/download/30666170/2192-5865-1-PB.pdf>
- Jerrett, Michael, Sara Gale, and Caitlin Kontgis. 2010. "Spatial Modeling in Environmental

and Public Health Research.” *International Journal of Environmental Research and Public Health* 7(4):1302–29.

Junuzovic, S., Inkpen, K., Hegde, R., Zhang, Z., Tang, J., & Brooks, C. (2011). What did i miss? in-meeting review using multimodal accelerated instant replay (air) conferencing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 513–522). Association for Computing Machinery.

<https://doi.org/10.1145/1978942.1979014>

Kane, T. J., Riegg, S. K., & Staiger, D. O. (2006). School Quality, Neighborhoods, and Housing Prices. *American Law and Economics Review*, 8(2), 183–212.

<https://doi.org/10.1093/aler/ahl007>

Katz, Raul L., Javier Avila, and Giacomo Meille. 2011. “Economic Impact of Wireless Broadband in Rural America.” *Stanfordville, NY: Telecom Advisory Services, at Http://rca-Usa. Org/wp-content/uploads/2011/02/Economic-Study-02. 24 11.*

Khatiwada, L. K., & Pigg, K. E. (2010). Internet Service Provision in the U.S. Counties: Is Spatial Pattern a Function of Demand? *The American Behavioral Scientist*, 53(9), 1326–1343. <https://doi.org/10.1177/0002764210361686>

Khatiwada, Lila K., and Kenneth E. Pigg. 2010. “Internet Service Provision in the U.S. Counties: Is Spatial Pattern a Function of Demand?” *The American Behavioral Scientist* 53(9):1326–43.

Khilnani, A., Schulz, J., & Robinson, L. (2020). The COVID-19 pandemic: new concerns and connections between eHealth and digital inequalities. *Journal of Information, Communication and Ethics in Society*, 18(3), 393–403. <https://doi.org/10.1108/JICES-04-2020-0052>

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *The American Psychologist*, 39(10), 1123.

<https://psycnet.apa.org/journals/amp/39/10/1123/>

- Kiesler, Sara, Jane Siegel, and Timothy W. McGuire. 1984. "Social Psychological Aspects of Computer-Mediated Communication." *The American Psychologist* 39(10):1123.
- Kolak, Marynia, Jay Bhatt, Yoon Hong Park, Norma A. Padrón, and Ayrin Molefe. 2020. "Quantification of Neighborhood-Level Social Determinants of Health in the Continental United States." *JAMA Network Open* 3(1):e1919928.
- Kruger, Lennard G. 2014. "Reauthorization of the Satellite Television Extension and Localism Act (STELA)." Retrieved September 16, 2021 ([https://ipmall.law.unh.edu/sites/default/files/hosted\\_resources/crs/R43490\\_140805.pdf](https://ipmall.law.unh.edu/sites/default/files/hosted_resources/crs/R43490_140805.pdf)).
- Kryczka, A., Arefin, A., & Nahrstedt, K. (2013). AvCloak: A Tool for Black Box Latency Measurements in Video Conferencing Applications. *2013 IEEE International Symposium on Multimedia*, 271–278. <https://doi.org/10.1109/ISM.2013.52>
- Lareau, A. (1987). Social Class Differences in Family-School Relationships: The Importance of Cultural Capital. *Sociology of Education*, 60(2), 73–85. <https://doi.org/10.2307/2112583>
- Lareau, A. (2011). *Unequal Childhoods*. University of California Press. <https://doi.org/10.1525/9780520949904>
- LaRose, Robert, Jennifer L. Gregg, Sharon Strover, Joseph Straubhaar, and Serena Carpenter. 2007. "Closing the Rural Broadband Gap: Promoting Adoption of the Internet in Rural America." *Telecommunications Policy* 31(6):359–73.
- Leatherman, John C. 2000. "Internet-Based Commerce: Implications for Rural Communities." *Reviews of Economic Development Literature and Practice* 5.
- Legendre, Pierre. 1993. "Spatial Autocorrelation: Trouble or New Paradigm?" *Ecology* 74(6):1659–73.
- Lehr, W. H., Osorio, C., Gillett, S. E., & Sirbu, M. A. (2006). *Measuring broadband's economic impact*. <https://dspace.mit.edu/handle/1721.1/102779?show=full>

- Lehr, William Herndon, Carlos Osorio, Sharon E. Gillett, and Marvin A. Sirbu. 2006. "Measuring Broadband's Economic Impact."
- Leiner, Barry M., Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard Kleinrock, Daniel C. Lynch, Jon Postel, Larry G. Roberts, and Stephen Wolff. 2009. "A Brief History of the Internet." *SIGCOMM Comput. Commun. Rev.* 39(5):22–31.
- Li, Yan, and Maria Ranieri. 2013. "Educational and Social Correlates of the Digital Divide for Rural and Urban Children: A Study on Primary School Students in a Provincial City of China." *Computers & Education* 60(1):197–209.
- Lin, D. W., and M. L. Liou. 1988. "A Tutorial on Digital Subscriber Line Transceiver for ISDN." Pp. 839–46 vol.1 in 1988., *IEEE International Symposium on Circuits and Systems*. [ieeexplore.ieee.org](http://ieeexplore.ieee.org).
- Loeb, T. B., A. J. Atkins-Jackson, and A. F. Brown. 2021. "No Internet, No Vaccine: How Lack of Internet Access Has Limited Vaccine Availability for Racial and Ethnic Minorities." *The Conversation*.
- Malecki, C. K., & Demaray, M. K. (2003). What Type of Support Do They Need? Investigating Student Adjustment as Related to Emotional, Informational, Appraisal, and Instrumental Support. *School Psychology Quarterly: The Official Journal of the Division of School Psychology, American Psychological Association*, 18(3), 231–252. <https://doi.org/10.1521/scpq.18.3.231.22576>
- Malone, Laurence J. 2001. "Commonalities: The R.E.A. and High-Speed Rural Internet Access." *arXiv [cs.CY]*.
- Maxwell, Elliot. 2005. "A New Future for Telecommunications Policy : Learning from Past Mistakes."
- McConnaughey, J., Wendy Lader, and R. Chin. 1998. "``Falling through the Net II: New Data on the Digital Divide", National Telecommunications and Information Administration." *Department of Commerce, US Government, Washington, DC*.

- McIntosh, K., Moss, E., Nunn, R., & Shambaugh, J. (2020). Examining the Black-white wealth gap. *Washington DC: Brooking Institutes*. <https://memphis.uli.org/wp-content/uploads/sites/49/2020/07/Examining-the-Black-white-wealth-gap.pdf>
- McLoughlin, Glenn J. 2005. "The National Telecommunications and Information Administration (NTIA): Budget, Programs, and Issues." LIBRARY OF CONGRESS WASHINGTON DC CONGRESSIONAL RESEARCH SERVICE.
- Miller, David W. 1990. "Social History Update: Spatial Analysis and Social History." *Journal of Social History* 24(1):213–20.
- Monge, Peter R., Lynda White Rothman, Eric M. Eisenberg, Katherine I. Miller, and Kenneth K. Kirste. 1985. "The Dynamics of Organizational Proximity." *Management Science* 31(9):1129–41.
- Montes, Sabrina. 2003. "Information Technologies in the US Economy." *Digital Economy*.
- Mooney, Stephen J., Michael Dm Bader, Gina S. Lovasi, Kathryn M. Neckerman, Andrew G. Rundle, and Julien O. Teitler. 2020. "Using Universal Kriging to Improve Neighborhood Physical Disorder Measurement." *Sociological Methods & Research* 49(4):1163–85.
- Moran, P. A. P. 1948. "The Interpretation of Statistical Maps." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 10(2):243–51.
- Moss, Mitchell L., and Anthony M. Townsend. 1998. "Spatial Analysis of the Internet in US Cities and States." in *conference on "Technological Futures", Durham, England*.
- Murat, M., & Bonacini, L. (2020). *Coronavirus pandemic, remote learning and education inequalities* (No. 679). GLO Discussion Paper. <https://www.econstor.eu/handle/10419/224765>
- Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: race and the disciplining of young students. *Psychological Science*, 26(5), 617–624. <https://doi.org/10.1177/0956797615570365>

- Owens, A. (2016). Inequality in children's contexts: Income segregation of households with and without children. *American Sociological Review*, 81(3), 549–574.  
<https://journals.sagepub.com/doi/abs/10.1177/0003122416642430>
- Owens, Ann, Sean F. Reardon, and Christopher Jencks. 2016. "Income Segregation Between Schools and School Districts." *American Educational Research Journal* 53(4):1159–97.
- Parmentier, Ingrid, Ryan J. Harrigan, Wolfgang Buermann, Edward T. A. Mitchard, Sassan Saatchi, Yadvinder Malhi, Frans Bongers, William D. Hawthorne, Miguel E. Leal, Simon L. Lewis, Louis Nusbaumer, Douglas Sheil, Marc S. M. Sosef, Kofi Affum-Baffoe, Adama Bakayoko, George B. Chuyong, Cyrille Chatelain, James A. Comiskey, Gilles Dauby, Jean-Louis Doucet, Sophie Fauset, Laurent Gautier, Jean-François Gillet, David Kenfack, François N. Kouamé, Edouard K. Kouassi, Lazare A. Kouka, Marc P. E. Parren, Kelvin S. H. Peh, Jan M. Reitsma, Bruno Senterre, Bonaventure Sonké, Terry C. H. Sunderland, Mike D. Swaine, Mbatchou G. P. Tchouto, Duncan Thomas, Johan L. C. H. Van Valkenburg, and Olivier J. Hardy. 2011. "Predicting Alpha Diversity of African Rain Forests: Models Based on Climate and Satellite-Derived Data Do Not Perform Better than a Purely Spatial Model." *Journal of Biogeography* 38(6):1164–76.
- Pasculli, A., A. Pugliese, R. W. Romeo, T. Sanò, Adolfo Santini, and Nicola Moraci. 2008. "The Uncertainty in the Local Seismic Response Analysis." *AIP Conference Proceedings*.
- Peters, O. (1999). Learning and teaching in distance education: Analyses and interpretations from an international perspective. *Education + Training*, 41(8), 384–386. <https://doi.org/10.1108/et.1999.41.8.384.3>
- Philip, L., Cottrill, C., Farrington, J., Williams, F., & Ashmore, F. (2017). The digital divide: Patterns, policy and scenarios for connecting the "final few" in rural

- communities across Great Britain. *Journal of Rural Studies*, 54, 386–398.  
<https://doi.org/10.1016/j.jrurstud.2016.12.002>
- Proctor, B. D., Semega, J. L., & Kollar, M. A. (2016). Income and poverty in the United States: 2015. *US Census Bureau, Current Population Reports*, 14.  
<https://census.gov/content/dam/Census/library/publications/2016/demo/p60-256.pdf>
- Rachfal, Colby Leigh, and Angele A. Gilroy. 2019. “Broadband Internet Access and the Digital Divide: Federal Assistance Programs.” *Congressional Research Service Report RL30719*. Available Online: <https://fas.org/sgp/crs/misc/RL30719>. Pdf.
- Rao, N., A. Maleki, F. Chen, W. Chen, C. Zhang, N. Kaur, and A. Haque. 2019. “Analysis of the Effect of QoS on Video Conferencing QoE.” Pp. 1267–72 in *2019 15th International Wireless Communications Mobile Computing Conference (IWCMC)*.  
[ieeexplore.ieee.org](http://ieeexplore.ieee.org).
- Reardon, S. F., & Bischoff, K. (2011). Income inequality and income segregation. *AJS; American Journal of Sociology*, 116(4), 1092–1153. <https://doi.org/10.1086/657114>
- Reardon, Sean F., Ericka Weathers, Erin Fahle, Heewon Jang, and Demetra Kalogrides. 2019. “Is Separate Still Unequal? New Evidence on School Segregation and Racial Academic Achievement Gaps.”
- Reglitz, Merten. 2020. “The Human Right to Free Internet Access.” *Journal of Applied Philosophy* 37(2):314–31.
- Ross, Zev, Michael Jerrett, Kazuhiko Ito, Barbara Tempalski, and George D. Thurston. 2007. “A Land Use Regression for Predicting Fine Particulate Matter Concentrations in the New York City Region.” *Atmospheric Environment* 41(11):2255–69.
- Sanders, F. H. 1998. “Broadband Spectrum Surveys in Denver, CO, San Diego, CA, and Los Angeles, CA: Methodology, Analysis, and Comparative Results.” Pp. 988–93 vol.2 in *1998 IEEE EMC Symposium. International Symposium on Electromagnetic Compatibility. Symposium Record (Cat. No.98CH36253)*. Vol. 2.

- Selwyn, Neil. 2015. "Data Entry: Towards the Critical Study of Digital Data and Education." *Learning, Media and Technology* 40(1):64–82.
- Shapley, K., Sheehan, D., Maloney, C., Caranikas-Walker, F., Huntsberger, B., & Sturges, K. (2007). Evaluation of the Texas Technology Immersion Pilot: Findings from the second year. *Texas Center for Educational Research*.  
<https://eric.ed.gov/?id=ED536295>
- Shedd, Carla. 2015. *Unequal City: Race, Schools, and Perceptions of Injustice*. Russell Sage Foundation.
- Short, J., Williams, E., & Christie, B. (1976). *The Social Psychology of Telecommunications*. Wiley.  
<https://play.google.com/store/books/details?id=Ze63AAAAIAAJ>
- Short, John, Ederyn Williams, and Bruce Christie. 1976. *The Social Psychology of Telecommunications*. Wiley.
- Skerratt, Sarah. 2010. "Hot Spots and Not Spots: Addressing Infrastructure and Service Provision through Combined Approaches in Rural Scotland." *Sustainability: Science Practice and Policy* 2(6):1719–41.
- Smith, Amy Symens, and Edward Trevelyan. 2019. *The Older Population in Rural America: 2012-2016*. US Department of Commerce, Economics and Statistics Administration, US ....
- Sparks, C. (2013). What is the "Digital Divide" and why is it Important? *Javnost - The Public*, 20(2), 27–46. <https://doi.org/10.1080/13183222.2013.11009113>
- Srinuan, C., & Bohlin, E. (2011). *Understanding the digital divide: A literature survey and ways forward*. <https://www.econstor.eu/handle/10419/52191>
- Stein, A., and L. C. A. Corsten. 1991. "Universal Kriging and Cokriging as a Regression Procedure." *Biometrics* 47(2):575–87.
- Stelitano, L., Doan, S., Woo, A., Diliberti, M., Kaufman, J. H., & Henry, D. (2020). The

- Digital Divide and COVID-19: Teachers' Perceptions of Inequities in Students' Internet Access and Participation in Remote Learning. Data Note: Insights from the American Educator Panels. Research Report. RR-A134-3. *Rand Corporation*.  
<https://eric.ed.gov/?id=ED609427>
- Stern, M. J., & Wellman, B. (2010). Rural and Urban Differences in the Internet Society—Real and Relatively Important. *The American Behavioral Scientist*, 53(9), 1251–1256.  
<https://doi.org/10.1177/0002764210361688>
- Stern, Michael J., and Barry Wellman. 2010. “Rural and Urban Differences in the Internet Society—Real and Relatively Important.” *The American Behavioral Scientist* 53(9):1251–56.
- Stratmann, Thomas, and Matthew Baker. 2020. “Examining Certificate-of-Need Laws in the Context of the Rural Health Crisis.”
- Telford, Taylor. 2019. “Income Inequality in America Is the Highest It’s Been since Census Bureau Started Tracking It, Data Shows.” *Washington Post* 26.
- Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46(sup1), 234–240. <https://doi.org/10.2307/143141>
- Tully, Stephen. 2014. “A Human Right to Access the Internet? Problems and Prospects.” *Human Rights Law Review* 14(2):175–95.
- US Census Bureau. 2020. “Changes to Counties and County Equivalent Entities: 1970-Present.” *United States Census Bureau*. Retrieved (<https://www.census.gov/programs-surveys/geography/technical-documentation/county-changes.html>).
- Vogels, Emily, Andrew Perrin, Lee Rainie, and Monica Anderson. 2020. “53% of Americans Say the Internet Has Been Essential during the COVID-19 Outbreak: Americans with Lower Incomes Are Particularly Likely to Have Concerns Related to the Digital Divide and the Digital Homework Gap.” *Pew Research Center*.
- Wahba, G. 1990. “Back Matter.” *Spline Models for Observational Data* 153–69.

- Wahba, Grace. 1975. "Smoothing Noisy Data with Spline Functions." *Numerische Mathematik* 24(5):383–93.
- Wang, Hansi Lo. 2018. "Native Americans on Tribal Land Are 'the Least Connected' to High-Speed Internet." *National Public Radio*.
- Warren, Martyn. 2007. "The Digital Vicious Cycle: Links between Social Disadvantage and Digital Exclusion in Rural Areas." *Telecommunications Policy* 31(6):374–88.
- Webster, Frank. 1995. "Information and the Idea of an Information Society." \_\_\_\_\_. *Theories of the Information Society*. London: Routledge 6–51.
- Whitacre, Brian E., and Bradford F. Mills. 2007. "Infrastructure and the Rural—urban Divide in High-Speed Residential Internet Access." *International Regional Science Review* 30(3):249–73.
- Whitacre, Brian E., and Bradford F. Mills. 2010. "A Need for Speed? Rural Internet Connectivity and the No Access/dial-Up/high-Speed Decision." *Applied Economics* 42(15):1889–1905.
- White, P., & Selwyn, N. (2013). MOVING ON-LINE? AN ANALYSIS OF PATTERNS OF ADULT INTERNET USE IN THE UK, 2002–2010. *Information, Communication and Society*, 16(1), 1–27. <https://doi.org/10.1080/1369118X.2011.611816>
- White, Patrick, and Neil Selwyn. 2013. "MOVING ON-LINE? AN ANALYSIS OF PATTERNS OF ADULT INTERNET USE IN THE UK, 2002–2010." *Information, Communication and Society* 16(1):1–27.
- Williams, S. N., Armitage, C. J., Tampe, T., & Dienes, K. (2020). Public perceptions and experiences of social distancing and social isolation during the COVID-19 pandemic: a UK-based focus group study. *BMJ Open*, 10(7), e039334. <https://doi.org/10.1136/bmjopen-2020-039334>
- Wong, David W. S. 2004. "The Modifiable Areal Unit Problem (MAUP)." Pp. 571–75 in *WorldMinds: Geographical Perspectives on 100 Problems: Commemorating the*

*100th Anniversary of the Association of American Geographers 1904–2004*, edited by D. G. Janelle, B. Warf, and K. Hansen. Dordrecht: Springer Netherlands.

Yamashita, Naomi, Andy Echenique, Toru Ishida, and Ari Hautasaari. 2013. “Lost in Transmittance: How Transmission Lag Enhances and Deteriorates Multilingual Collaboration.” Pp. 923–34 in *Proceedings of the 2013 conference on Computer supported cooperative work*. New York, NY, USA: Association for Computing Machinery.

Zahnd, Whitney E., Nathaniel Bell, and Annie E. Larson. 2021. “Geographic, Racial/ethnic, and Socioeconomic Inequities in Broadband Access.” *The Journal of Rural Health: Official Journal of the American Rural Health Association and the National Rural Health Care Association*. doi: 10.1111/jrh.12635.