Mapping connections between broadband expansion and equitable economic prosperity

CUSP Capstone final report

Dan Levine, Max Magid, Zuda Yan, Kaiwen Zhang

Abstract

Access to broadband internet is vital for economic development, however no comprehensive research has connected measures of broadband availability and performance and economic success. This study detects correlation between past broadband expansion and subsequent economic growth across locales with certain base levels of economic performance. We identify the demographic and economic factors which correspond to lower broadband scores and also find the measures common in places which saw a boost in prosperity following broadband expansion. We found that there is a delayed correlation between an increase in broadband access and an increase in economic development. We combine broadband, economic, and equity considerations to locate the places that may achieve the most equitable and absolute gains from broadband expansion and present a prioritized list of counties. This project presents a new, nationwide, Census Tract-level composite dataset of broadband performance and an interactive map for exploration and use in further research.

Introduction

Problem definition

The Internet has become an integral part of the 21st-century economy. However, in many parts of the United States residents still struggle with lack of access to high-quality broadband.

The Federal government has dedicated substantial funding to expanding broadband infrastructure and US Ignite, a non-profit organization supporting smart cities projects, is ready to help cities improving connectivity.[1] However, while expanding broadband access will likely have a positive impact on communities all over the country, there has been little research that comprehensively measures where broadband access will best increase economic performance or where improvements will close critical, inequitable gaps in access.

This study constructs the first nationwide, linked data on broadband performance and access along with economic indicators and explores the connections between broadband

1 https://www.us-ignite.org

expansion and economic growth.

Moreover, poor access to broadband may mirror and exacerbate disparities and inequities. This study further revealed where there are the greatest disparities in broadband access and how these align with racial segregation and income gaps.

Literature review

Several previous studies have found connections between broadband access and speed and various measures of economic success. However, no previous work had linked all measures of broadband and studied the effects nationwide.

Nazareno and Jose (2021) looked at the effect of broadband deployment in unserved rural areas in the United States through Connect America Fund (CAF).[2] Their preliminary results suggested an overall positive effect on employment rates, but one that varies substantially by region. This study is limited in geographic scope and also identifies the lack of more comprehensive data. We build off the authors' use of Census American Community Survey (ACS) data for finding economic-broadband relationships.

Han's 2021 study examined how broadband affects self-employment and work-from-home for married women using, ACS and FCC data.[3] It found that both adoption and access to broadband positively impact self-employment and work-from-home, and the adoption of broadband has stronger effects. This novel approach used Census data to specifically focus on married women and suggests a way US Ignite and future researchers may further slice the data to gain insights about particular groups.

Koklo's (2012) important early paper on the link between broadband and economic performance found that broadband access increased total employment but did not affect unemployment.[4] This shows that where there is broadband there is more business growth (startup or expansion) and more immigration, but the impact on employment opportunities for residents already in the area is limited. But the difference in broadband speed may have some impact on the unemployment rate.

Lobo, et al. (2019) examined the effects of broadband speed on county unemployment rates in Tennessee by first merging the older National Broadband Map dataset and the newer FCC data.[5] They found that unemployment rates are about 0.26 percentage points lower in counties with high speeds compared to counties with low speeds. Better quality broadband appears to have a disproportionately greater effect in rural areas. The study is

---

2 Luísa Nazareno and Justina Jose. "The Effects of Broadband Deployment in Rural Areas: Evaluating the Connect America Fund Program." 2021. SSRN. http://dx.doi.org/10.2139/ssrn.3897867

3 Han, Luyi, Broadband, Self-Employment, and Work-from-Home — Evidence from the American Community Survey (October 4, 2021). Available at SSRN: https://ssrn.com/abstract=3936667 or http://dx.doi.org/10.2139/ssrn.3936667

4 Jed Kolko. "Broadband and local growth." 2012. Journal of Urban Economics 71(1) https://doi.org/10.1016/j.jue.2011.07.004

5 Bento Lobo, Rafayet Alam, Brian E. Whitacre. "Broadband Speed and Unemployment Rates: Data and Measurement Issues." 2019. Telecommunications Policy 44(1) https://doi.org/10.1016/j.telpol.2019.101829

limited to one state and only at the comparatively coarse county level because older data on broadband is limited and not well linked to newer data.

Data

To comprehensively understand the relationship of broadband access on economic performance, we sourced and combined all available data on broadband. These sources are summarized in Table 1.

Table 1: Broadband data sources

Name	**Measures**	**Years available**	**Update frequency**	**Geography**	**Notes**
FCC Form 477[6]	Availability	2014 - 2021	Every 6 months	Census Tract	Reported by internet service providers. Coarse and imprecise: indicates broadband access if *any* provider serves at least *one* address.
American Community Survey[7]	Availability	2016 - 2019	annual	Census Tract	Household survey
Microsoft Broadband Usage[8]	Percentage of households with broadband-speed downloads	2020	never	ZIP Code	Noise is applied to ZIP-level data to protect privacy. Location is estimated from IP address.
M-Lab Speed Test[9]	Actual network speed from tests initiated by users	2019 - present	daily	County	Location is estimated from IP address.

6 https://www.fcc.gov/general/broadband-deployment-data-fcc-form-477

7 https://www.census.gov/programs-surveys/acs

8 https://github.com/microsoft/USBroadbandUsagePercentages

9 https://www.measurementlab.net/data/

Name	Measures	Years available	Update frequency	Geography	Notes
Ookla speed test[10]	Actual download and upload speed from tests initiated by users	2019 - present	quarterly	Quadkey zoom 16 (~$600m^2$ raster cells)	Location is estimated from IP address. Incomplete coverage (data only available where tests were initiated)
Broadband Now[11]	Number of providers and lowest advertised monthly cost	2020	never	ZIP code	Dataset also includes M-Lab 12-month average speeds and FCC 477 data
Connect America Fund[12]	Location, speed, year of deployment, and grant information	2000-2020	annual	Census Block	
National Neighborhood Data Archive (NaNDA)[13]	Availability	2014-2018	annual	Census Tract	Compiled data from FCC form 477

10 https://github.com/teamookla/ookla-open-data

11 BroadbandNow Research. https://github.com/BroadbandNow/Open-Data

12 Universal Service Administrative Company. https://opendata.usac.org/High-Cost/High-Cost-Connect-America-Fund-Broadband-Map-CAF-M/r59r-rpip

13 Clarke, Philippa, Gomez-Lopez, Iris, Li, Mao, and Chenoweth, Megan. National Neighborhood Data Archive (NaNDA): Broadband Internet Access by Census Tract, United States, 2014-2018. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2020-02-21. https://doi.org/10.3886/E117866V1

Name	**Measures**	**Years available**	**Update frequency**	**Geography**	**Notes**
Indicators of Broadband Need[14] *aggregate dataset. includes:* Ookla Speed Test M-Lab Speed Test Microsoft Broadband Usage ACS FCC Form 477	Speed and availability, with various measures:	2021	Will be updated every 6 months. (Only one updated so far)	Census Tract	Aggregated, most recent data.

14 United States Department of Commerce, National Telecommunications and Information Administration (NTIA). "Indicators of Broadband Need."

https://broadbandusa.maps.arcgis.com/apps/webappviewer/index.html?id=e2b4907376b548f892672ef6afbc0da5

Data organizing and cleaning

To collect the various measures of broadband availability and economic performance into single comparable scores, we first needed to organize each data source at a common geographic level.

The code we created for cleaning and combining data, as well as all the subsequent analysis methods described below, are available online on a public GitHub repo.[15]

We selected Census Tracts as a fine-grained, compact, and relatively stable unit of geography, and one at which much of our data was natively available. (Specifically, we used 2010-vintage Census Tracts, to match data available from the Census ACS and FCC.)

Spatial interpolation

Data which was available at the ZIP Code level was spatially interpolated to Census Tract using areal weighting. The computed value at each Tract is the value at each overlapping ZIP Code area, weighted by the percentage of the Tract which the ZIP Code area covers. This interpolation used the Python `tobler` package.[16]

Speed test data from Oookla was summarized by year at the Tract level with a test-weighted average: local median speeds were summed for each year, then the computed Tract speed was the average of the speed at all overlapping tiles in each of four quarters, weighted by the number of tests at each location in each quarter.

Limitations

The lack of historic or longitudinal data on broadband availability, speed, or access limited the potential for analysis. Some sources were released as a one-time snap-shot; while those offering earlier vintages extend back no further than 2016.

Without earlier data, it is difficult to know how early adoption of broadband may have helped certain areas excel economically, or, more generally, to understand what causal effect broadband availability may have had on economic performance.

The Microsoft data shows a particular clever way of using actual, measured internet traffic to measure internet speed. Other companies that transmit large volumes of data likely have access to similar information. If such data were recorded and available across a wider timeline, it could greatly improve the scale of possible analysis.

The geographic imprecision of the broadband data we use also challenges these analyses. Ookla, M-Lab, and Microsoft data all use geolocation based on IP addresses. This is not a precise location of the actual end use, but generally a location of nearby ISP routing equipment. M-Lab specifically cautions against inferring greater than county-level

15 https://github.com/dlevine01/broadband-economy

16 https://pysal.org/tobler/index.html

geographic specificity, though the data are reported at the more granular, ZIP code level.[17] Reaggregation to Tract level may add additional error. As each measure is one independent component in a composite broadband score, we deem possible geographic errors to be acceptable.

Methods and findings

Identifying local clusters of high or low broadband use

Local indicators of spatial autocorrelation were computed across the Tract-level broadband use data.[18] Simulated alternative distributions were used to estimate pseudo-*p*-values and an α of 0.99 was used as a cutoff for identifying significant outliers. This was computed using the Python `pysal esda` tool.[19]

Create unified broadband scores

We created two different types of combined broadband scores for two different types of analysis. The first broadband score is a longitudinal score we can use to more easily investigate how broadband access has expanded over time. The second type of score provides a more in-depth snapshot of a particular year, in this case 2019, which we use to explore current inequalities of access to high speed broadband internet services.

We use FCC data from form 477 to find data from 2011 to June 2019 as the backbone of our longitudinal score. It gives a score from 0 to 5 for upload and download speeds. From 2011 to 2014 this score indicates percentages of households with residential fixed high-speed connections at least 3 Mbps download speed and 768 kbps upload speed. A score of 0 indicates 0% of households meet this criteria, and a score of 5 means 80% of households meet this criteria. The other numbers proceed in 20% household increments. (ie. 1 means that 0%-20% have a high-speed connection). In 2015 the definition of broadband-speed changed to residential fixed high-speed connections at least 10 Mbps downstream and 1 Mbps upstream. To adjust for this change, we weight the pre-2015 scores slightly less to account for the more lenient score.

Once we have the weighted FCC scores, we supplement by adding in data from the National Neighborhood Data Archive (NaNDA) developed by researchers at the University of Michigan's Institute for Social Research. They use the same rating scale as the FCC, and derive much of their data from the same FCC source. However, there are some small differences in the assigned value, so we incorporate it in. This dataset spans 2014-2018.

Beginning in 2017, the ACS includes statistics on broadband and internet access. It does not use the 0-5 scale, but instead shows the percentage of households with broadband access.

---

17 https://github.com/m-lab/stats-pipeline/blob/master/docs/geo-precision.md

18 Anselin, L. (1995), Local Indicators of Spatial Association—LISA. Geographical Analysis, 27: 93-115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

19 Rey, Sergio & Anselin, Luc. (2007). PySAL: A Python Library of Spatial Analytical Methods. Review of Regional Studies. 10.52324/001c.8285. https://pysal.org/esda/index.html

This value was converted to a 0 to 5 scale, then added to the composite score with the same weight as the FCC and NaNDA scores.

By putting together these factors we were able to easily measure how the broadband profile of a tract was changing over time.

However, there were some other broadband indicators we had access to that did not easily fit into the 0-5 scoring format, such as the indicators of broadband need assembled by the National Telecommunications and Information Administration and internet speed test data.

Using principal component analysis we were able to compress these different factors into a single score that explained nearly 50% of all variance between locales. This score is harder to understand and trace from one year to the next, but does provide a more accurate snapshot than our longitudinal data. When developing our inequity analysis, this is the composite score we used.

Create economic performance typologies

To find which areas showed signs of economic growth we needed to define prosperity. We had a multitude of measures available in the (ACS) data. We decided to define economic growth using an unsupervised clustering method because this reveals the different sorts of growth patterns from this complex data without relying on presupposed categories.

We started by summarizing a multitude of indicators using principal component analysis (pca). Once we had that score for each census tract over each year between 2012 to 2019, we found the computed the difference in pca score over a three-year period. We then found local indicators of spatial associations (LISA statistics) to control for the geographic component of economic growth.

Using the three-year difference between a variety of economic indicators and the LISA statistics, we ran a *k*-means clustering to find the types of economic growth. We found that for every three-year period we looked at there were five different economic growth types, (although occasionally a single tract would be in its own sixth group as an outlier). Unsurprisingly, most tracts fell into an economically stationary category, which we called group 3. Groups 4 and 5 both experienced growth, although in general group 5 tended to see a greater average increase in pricing of homes and median income, while group 4 tended to match or exceed group 5 for increase in employment, poverty reduction, and increase in high earners. There was also a geographic component to the groupings, with group 5 Tracts mostly found in cities, while group 4 Tracts were more scattered. Figure 1 shows the relative change in several economic categories by cluster group. This plot reflects the changes from 2012-2015, although its general pattern held for each three-year period. Figure 2 shows economic classifications in the state of Indiana (a generally representative geography), showing class 5 areas within metropolitan areas and college towns, class 1 across the most rural areas, and general large swaths of class 3. The economic classifications nationwide can be seen and explored through the [interactive map tool]() we have built.

Figure 1. Standardized change in economic factors by cluster, 2012-2015.

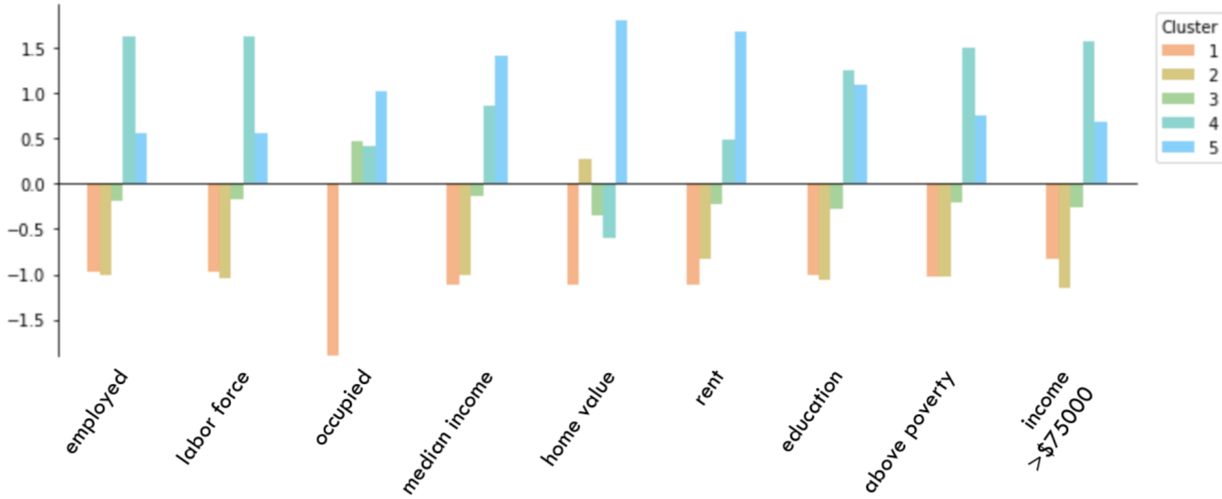
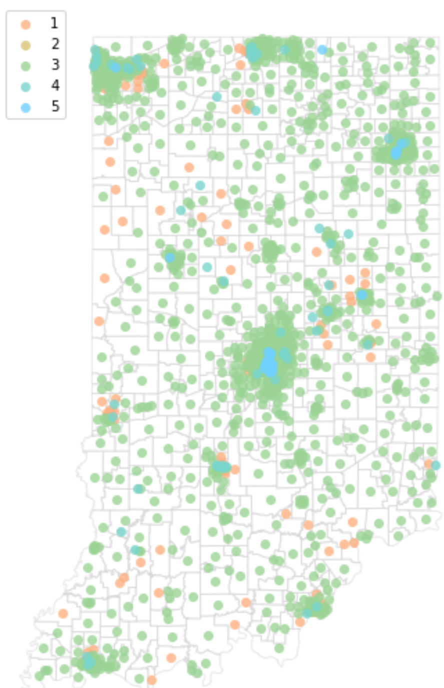


Figure 2: Economic clusters in Indiana

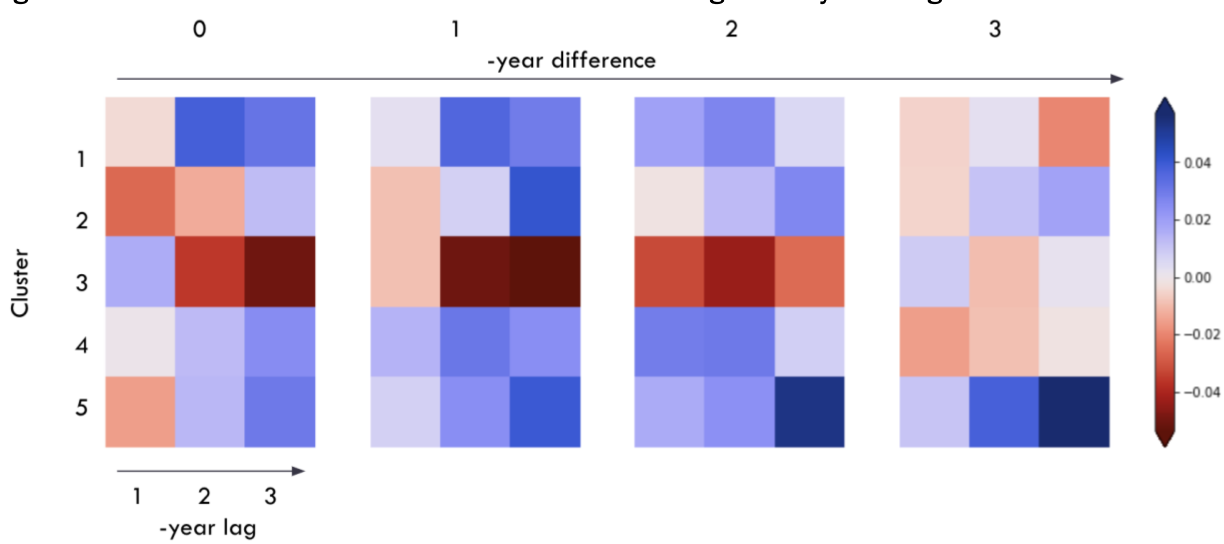


Test lagged effect of increases in broadband performance across economic clusters

Once we had the longitudinal broadband score and the economic clusters, we could compare the two to find any correlations. We did this by grouping the Tracts into their respective counties and then calculating the increases in broadband scores per county along with the count of Tracts in each economic group in each county (normalizing by the total number of Tracts in each county).

Once our county-level broadband score increases and economic growth types were prepared we generated a correlation table. One axis shows the change in broadband score over one year, two years, and three years; the other axis shows the relative count for each economic group. We then projected the economic growth into the future. For instance, if the broadband axis looked at 2014-2015 (one-year lag), 2013-2015 (two-year lag), 2012-2015 (three-year lag) then the economic group axis would look at the groups from 2012-2015 (0 year difference), 2013-2016 (one-year difference), 2014-2017 (two-year difference), and 2015-2018 (3 year difference). The correlation is shown in Figure 3.

Figure 3: Correlations between broadband and economic growth by time lag



In this way we discovered that there is a strong positive correlation between a three year increase in broadband scores and eventual group 5-style growth in the future.

Test equity and disparity of broadband access

To test for disparities of broadband access we regressed broadband score on several demographic variables available through ACS data. This showed whether each indicator had a positive or negative correlation to broadband. We added Lasso regularization to control for multicollinearity in this analysis.

We evaluated which factors were most important as the value of the coefficients divided by the standard deviation of the variable. Through this method we discovered the following factors (in order of importance) indicate *lower* broadband scores:

- high unemployment
- lower rates of people born out of state
- lower levels of owner-occupied places
- lower levels of household income
- higher proportions of Black, Indigenous, and Hispanic people
- lower rents
- higher rates of people in service professions
- higher rates of people without bachelor's degrees
- lower home values
- higher number of renters
- higher rates of people employed in management, business, science, and art
- lower population density

We also further investigated inequities by taking a closer look at density. We discovered that a histogram of the log of tract density revealed a double peak structure, likely corresponding to tracts within and outside of large metropolitan areas. When splitting tracts into four categories based on their densities (corresponding to urban, suburban, small town, and rural), we found that within each category the broadband score does not fluctuate significantly by density, but between groups it does.

We also dug deeper into racial disparities and found that tracts with large proportions of Black, Hispanic, and Indigenous people often had disproportionately low broadband scores, while the reverse was true of tracts with high proportions of White and Asian.

Recommendations

We found a strong correlation between broadband change over the course of three years and future improvement in economic conditions. The lag between a change in broadband score and economic improvement suggests a possible causal link between the two.

To find the demographic makeup of the places experiencing this growth, we identify the average demographic profile of these places before any change occurs. That is, for each set of locations, we looked at demographics in year zero, broadband improvements in years 0-3, and economic performance in years 3-6.

With this look-back, we found that the demographic variables that most closely correlated with increased broadband and greatest economic growth were:

- high levels of education;
- high numbers of non-citizens;
- high home median values; and
- relatively high rent

Policy implications

Our results suggest several demographic factors that correspond to an economic dividend from increased broadband access, as well as several factors corresponding to lagging broadband availability.

We combine these two considerations to suggest target areas where broadband expansions will redress inequities and may lead to greater economic returns. The list of counties in Table 2 is ordered by importance.

The list reveals that a huge issue is access more so than simply infrastructure. Many counties on this list have broadband capabilities, but for one reason or another connectivity remains out of reach for too many people. On the other hand, these counties show the opportunity available to possibly kickstart growth through smart investment in expanding access.

Table 2: Recommended counties for broadband improvements to address inequities and enable economic growth

Name	State	# of Targeted Tracts	pca_score
Los Angeles	CA	283	-0.090194
San Diego	CA	39	-0.051646
Alameda	CA	23	-0.069572
Miami-Dade	FL	18	-0.034586
San Bernardino	CA	18	-0.024785
Kings	NY	18	-0.017785
Cook	IL	15	-0.006625
Bronx	NY	13	-0.019034
Orange	CA	13	-0.010609
Monterey	CA	11	-0.171034
Santa Barbara	CA	10	-0.092233
Ventura	CA	9	-0.049969

Riverside	CA	9	-0.006419
New York	NY	8	-0.013715
Sacramento	CA	8	-0.008751
De Kalb	GA	6	-0.060759
Fresno	CA	5	-0.041411
Travis	TX	5	-0.013838
Maricopa	AZ	5	-0.003158
Palm Beach	FL	4	-0.014022
Hillsborough	FL	3	-0.024491
Washington	DC	3	-0.021655
Denver	CO	3	-0.011255
Santa Clara	CA	3	-0.006067
Multnomah	OR	3	-0.006042
San Joaquin	CA	3	-0.00417
Broward	FL	3	-0.003969
King	WA	3	-0.003312
Marin	CA	2	-0.047677
San Luis Obispo	CA	2	-0.038232
Stanislaus	CA	2	-0.030225
Osceola	FL	2	-0.023765
Arlington	VA	2	-0.023469
Orleans	LA	2	-0.016269
Middlesex	NJ	2	-0.013748

Suffolk	MA	2	-0.006545
Essex	MA	2	-0.006476
Baltimore City	MD	2	-0.006374
San Francisco	CA	2	-0.005746
Prince Georges	MD	2	-0.0055

Conclusions

This study compiles the first nationwide, Tract-level data on broadband performance and links these values with economic growth. Exploration of the data reveals that broadband expansion in certain areas correlates with greater economic success in following years. We identify many demographic factors which correspond to lower scores and also find the measures which are common in places which saw an economic boost following broadband expansion. We combine these considerations to a list of counties which may achieve the most equitable and absolute gains from broadband expansion.

This study compiles the first nationwide, Tract-level data on broadband performance and links these values with economic growth. We identify measures of inequity in broadband access and correlates of economic gains from broadband expansion. We also provide these measures on an interactive map for further exploration and analysis.

Future research may further explore the causal linkages between broadband expansion and various economic gains. While conclusive causal findings are hampered by limited historic data; further filtering of these data could explore, for instance, whether broadband speed, availability, cost are most indicative of future success, and whether economic gains are realized through business growth, relocation, education gains, upskilling, or other economic development mechanisms.

In each of the areas we suggest for high-impact broadband expansion, further study can reveal what interventions may be most beneficial. In urban areas with generally high-quality infrastructure, programs to increase household access or reduce costs may be most effective. Local low-broadband-use outliers may highlight the areas most in need of this type of service.

Supplemental materials

Code for data preparation and analysis

The Python code we created and used to collect, prepare, combine, rescale, and analyze the data for this project are available online on a public GitHub repo.[20] The complete code is posted so that the methods for this project are fully reproducible.

Visualization and exploration tool

To allow for exploration of the broadband and economic indicators we built an interactive map tool, available online at: https://usignite.carto.com/u/usignite-intern/builder/a7627f0d-a64d-44a3-892a-820b14c0dfab/embed. A sample is shown in Figure A. This tool includes several interactive layers for different views of the Tract-level data:

1. Economic cluster classification. On the 5-class scale described above.
2. Broadband score (normalized to the range of 0 worst) to 1 (best) for easier interpretation)
3. Estimated actual broadband usage (ranging from 0, no use, to 1, 100% use)
4. Local outliers and clusters of high- or low-broadband use

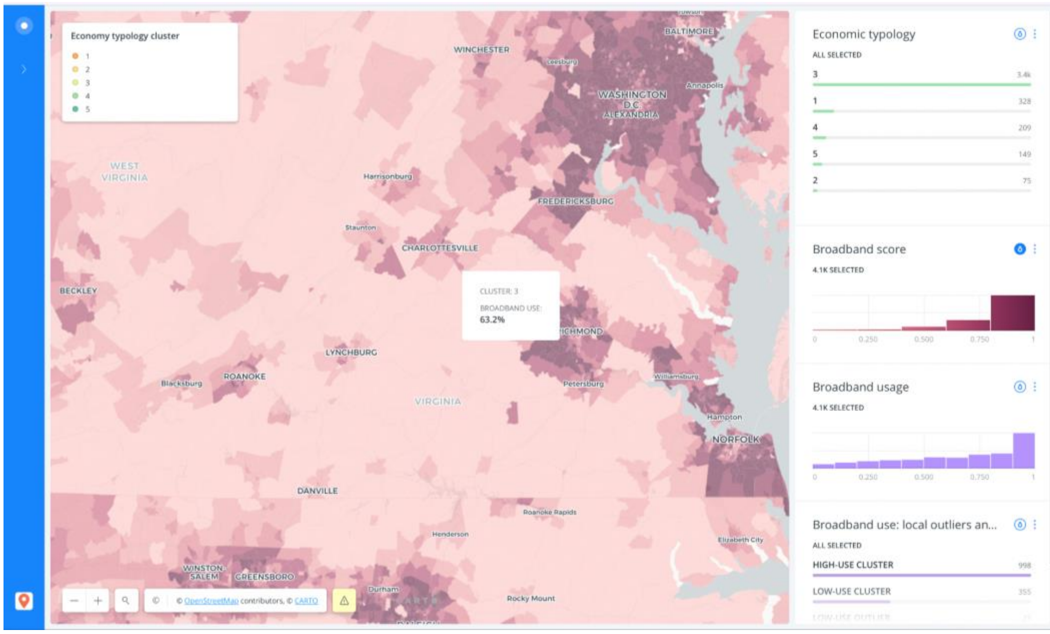
Figure A: Interactive map of economic classification and broadband performance

Each layer is a selectable filter: clicking on one or more economic classes or selecting a range of broadband scores (for instance) will filter the map to show only Tracts matching that classification or those scores.

20 https://github.com/dlevine01/broadband-economy

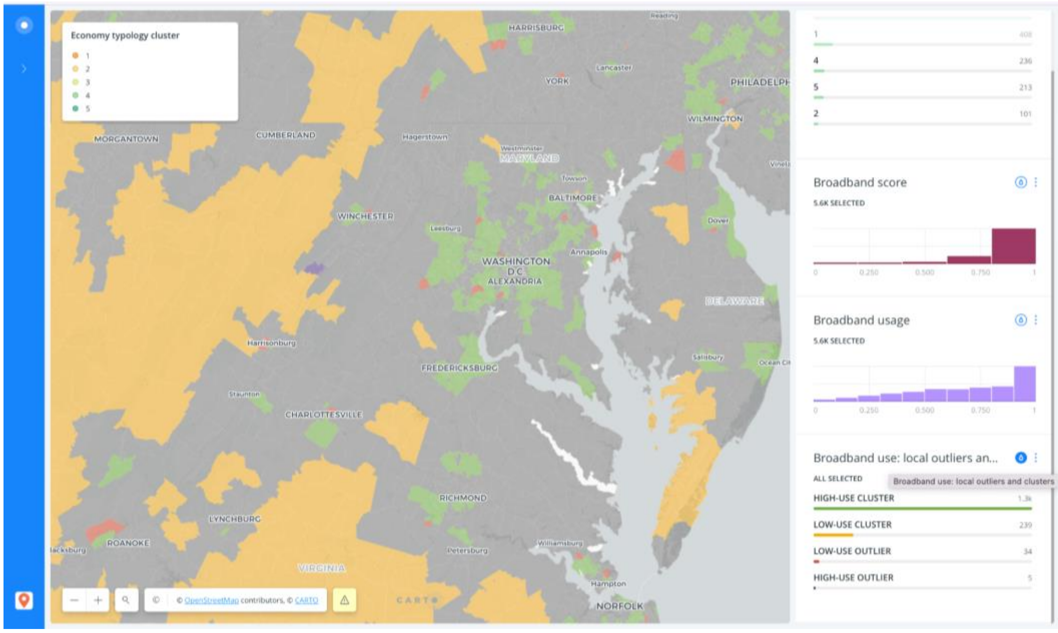
The map is initially shaded based on each Tract's economic class, but it can easily be reshaded by any other factor by clicking the 'Apply auto style' button on any other layer. For instance, applying auto style on the broadband score layer reshaded the map by broadband score (see Figure B).

Figure B: Map showing broadband scores



The low broadband use local outliers indicate Tracts where actual broadband use is significantly lower than would be expected from the use in the surrounding tracts. Filtering to just these areas (by clicking the 'low-use outlier' bar on this layer) highlights Tracts where perhaps greater access to broadband, rather than new infrastructure, could improve broadband performance. These areas are seen in red in the sample map view in Figure C.

Figure C: High- and low-broadband use clusters



This dynamic map is built with the Carto mapping platform. It uses SQL queries of the data stored in the Carto database to construct the map. This means that as different data sources are updated (if, for instance, newly-published economic indicators are extracted, transformed, and loaded), the map will be automatically updated.