David Ilouz
Prof. Robert Bird
Programming and Data Management
12 May 2025

Access to Broadband as a Function of Economic Prosperity

Is access to high-speed broadband internet a function of wealth? In this report, I have conducted an initial review of nationwide broadband coverage and economic conditions at the county level for our prospective client, the National Telecommunications and Information Administration (NTIA). Briefly, this federal agency is focused on expanding broadband access and "advising the President on information, telecommunications, and related technology policy."[1] Below, I have outlined in detail the datasets used, the challenges faced in collecting, cleaning, and analyzing the data, and the rationale behind several important data collection and analysis decisions. I close with an initial overview of the results and questions for potential further exploration.

Data Sources and Structure

This analysis draws on two government-maintained datasets: Income and poverty data from the U.S. Census Bureau, and Broadband coverage data from the Federal Communications Commission:

Income and poverty data was sourced from the Census Bureau's *American Community Survey (ACS) 5-Year Estimates* for 2018–2022,[2] accessed via the **tidycensus** R package:

- **Topic covered:** County-level economic indicators
- **Observations:** The unit of observation is county, and the dataset included 3,222 counties in total (though there were initially 6,444 total observations, with two rows per county, each with a unique data point).
- **Key variables:** I selected median household income (**median_income**) and percentage of the population below the poverty line (**poverty_rate**) as two indicators of economic prosperity.

Broadband coverage data was sourced from the FCC's National Broadband Map.[3] This dataset reports access to fixed broadband (which refers to high-speed internet delivered through wired

[1] National Telecommunications and Information Administration. 2025. Accessed May 8, 2025. https://www.ntia.gov.
[2] U.S. Census Bureau. 2025. "American Community Survey (ACS) 5-Year Data (2018–2022)." Census Bureau API. Accessed May 4, 2025. https://www.census.gov/data/developers/data-sets/acs-5year.html.
[3] Federal Communications Commission. 2024. "Fixed Broadband Summary by Geography Type." National Broadband Map. Accessed May 4, 2025. https://broadbandmap.fcc.gov/data-download/nationwide-data.

infrastructure such as cable, fiber, or DSL) as of December 2023 for various speeds, technology types, and geographical entities.

- **Topic covered:** Fixed broadband availability by county, including baseline and high-speed thresholds.
- **Observations:** The raw data set contains over 62,000 rows, but has been filtered to only include counties and rows where **technology == "Any Technology"** and **biz_res == "R"** indicating residential aggregate broadband coverage by US county.
- **Key Variables:** I selected two metrics from this dataset for analysis. The first is the % of residential units by county with access to 25 Mbps download/3 Mbps upload broadband speed (**speed_25_3**). This is the FCC's former standard, which I use as a baseline for broadband internet access. The second is the % of units with access to 100/20 Mbps (**speed_100_20**), which is the current FCC standard, and which I use as the threshold for access to high-speed broadband internet.

I merged these two datasets into a **combined table**, described below:

- **Observations:** This data set contains 3134 observations, after merging and standardizing county names. The unit of observation is county.
- **Variables:** There are five variables I use in the final merged dataset:
    - **median_income** – A key indicator of county-level economic prosperity, reflecting the midpoint of household income distribution.
    - **poverty_rate** – The share of residents living below the federal poverty line, signaling economic hardship and structural disadvantage.
    - **wealth_score** – This is a composite score based on z-scores for median income and inverse poverty rate (% above poverty line) to summarize overall economic prosperity for each county.
    - **speed_25_3** – Measures the proportion of households with access to basic broadband (25 Mbps download/3 Mbps upload), the FCC's former minimum standard and a proxy for access to baseline broadband internet.
    - **speed_100_20** – Measures the proportion of households with access to basic broadband (100 Mbps download/20 Mbps upload), the FCC's current minimum standard and a proxy for access to high-speed broadband internet.

Challenges and Methodological Choices

When importing, merging, and cleaning the data, I ran into a few challenges and had to make several methodological choices to maintain sound results:

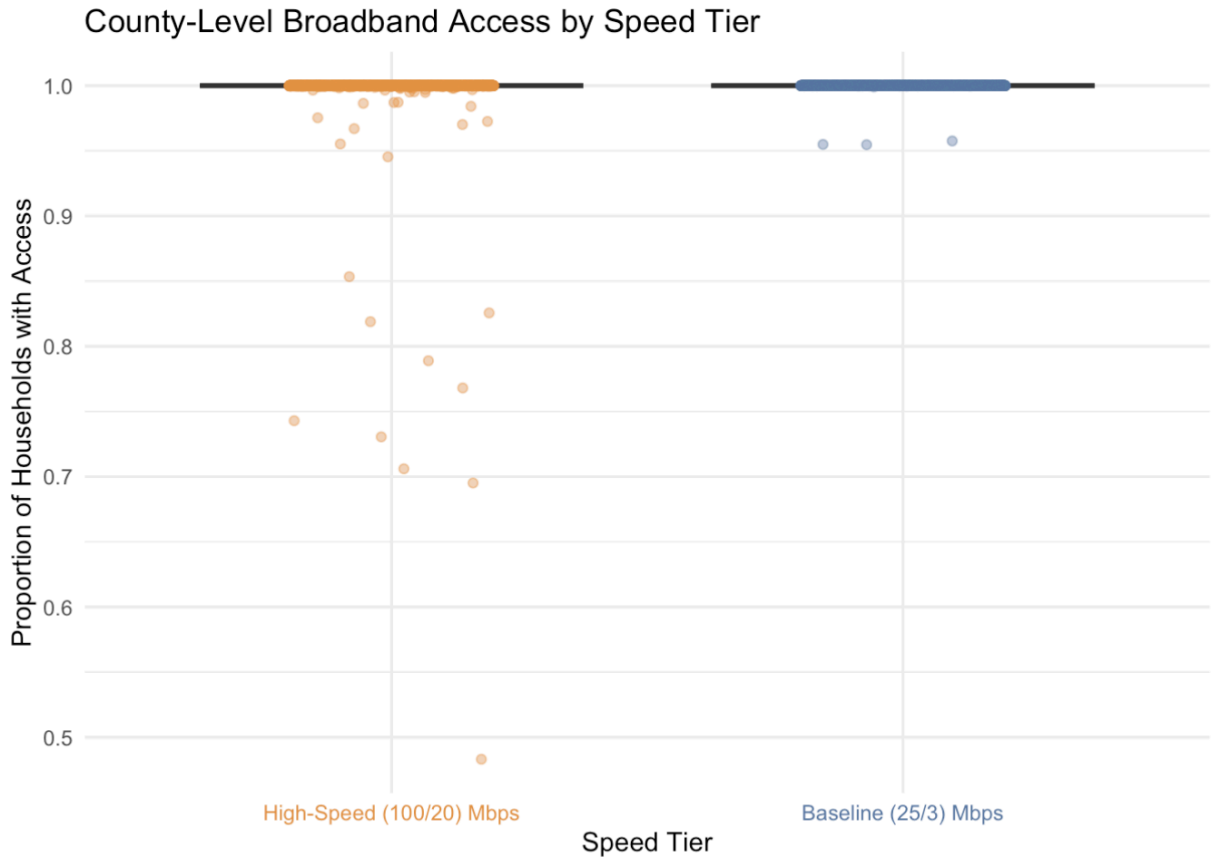
- **Timestamp:** For the broadband data, I used the most recent finalized release: December 2023, providing the clearest and most up-to-date snapshot of access nationwide. For the income and poverty data, I used the 2022 American Community Survey (ACS). Although the timeframes do not perfectly align, they are close enough to allow meaningful analysis without the risk of outdated trends. ACS 2022 data reflects income in the 12 months prior to the survey, which overlaps with much of 2022 and early 2023. This pairing was judged to be a reasonable compromise between recency and data availability.
- **Broadband Data:** First, I chose to use fixed broadband access rather than mobile, as fixed internet (e.g., cable, fiber) is a more relevant indicator of infrastructure access, particularly for households relying on the internet for work, school, or healthcare. Mobile coverage may be more ubiquitous, but is less tied to structural investment.
- **Economic Data:** In preparing the dataset, I normalized the **poverty_rate**, which was initially stored as a whole number percentage (e.g., "13" for 13%). I converted the variable to a proportion (0.13) to ensure comparability with broadband data. I also created a composite **wealth_score**, combining z-scored values of median income and inverse poverty rate. This score offers a broader, normalized picture of county-level economic advantage.
- **Inconsistent Data:** Lastly, I encountered a notable inconsistency in the data for Connecticut. In the broadband dataset, Connecticut observations are organized by county, but in the ACS data, the state is divided into "Planning Regions," which are not coterminous with counties. This mismatch made merging impossible without significant assumptions. I considered assigning a statewide average to all Connecticut counties, but this could have disproportionately skewed results due to Connecticut's relatively high income and broadband coverage. Instead, I chose to exclude Connecticut entirely to preserve the accuracy of the merged dataset.

Visualizations and Analysis

To explore the relationship between economic prosperity and broadband availability, I focused on five core variables: **median_income**, **poverty_rate**, **wealth_score**, **speed_25_3**, **and speed_100_20**. Together, these variables allow us to assess not only the overall distribution of broadband access, but also its potential association with local economic conditions.

Distribution of Broadband Access by Speed Tier

The first visualization presents the distribution of county-level broadband coverage for two different speed tiers: the former FCC standard of 25 Mbps download/3 Mbps upload (**speed_25_3**), and the newer standard of 100/20 Mbps (**speed_100_20**). I use these as proxies for access to baseline and high-speed broadband internet, respectively.

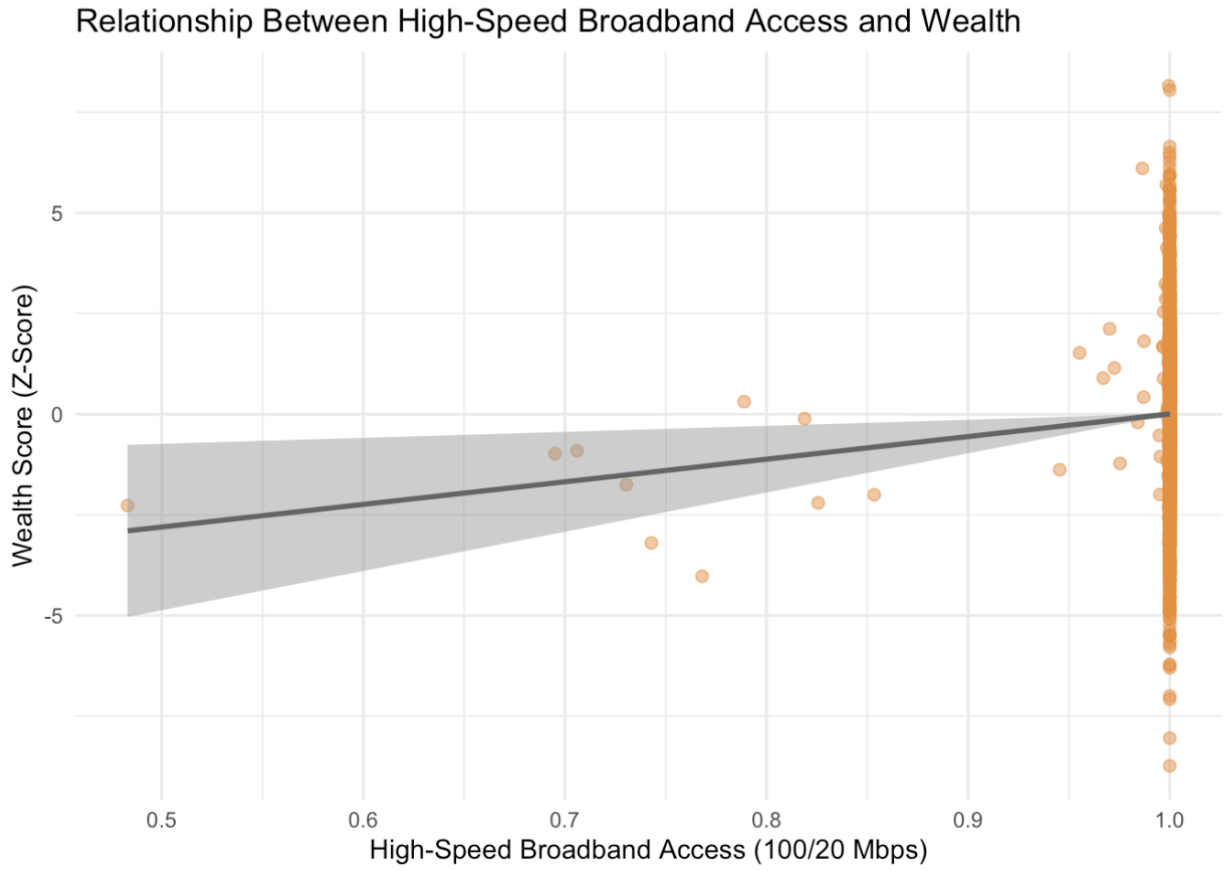


The boxplots show that baseline broadband access is nearly universal, with the vast majority of counties reporting near 100% coverage. The data for high-speed broadband access is similar, but

there is slightly more variation. While the vast majority of counties are at or near full coverage, a more substantial number fall below 90%, and some even below 70%. This suggests that although basic broadband infrastructure is broadly established, gaps in high-speed access exist and may reflect deeper structural or geographic inequities.

Wealth Score vs. High-Speed Broadband Access

The second visualization explores the relationship between economic advantage (as captured by a standardized **wealth_score**) and access to high-speed broadband (**speed_100_20**). Each point represents a county, and the scatter plot includes a linear trendline with a 90% confidence interval (shaded region).



While the trend is slightly upward, indicating that higher-wealth counties are somewhat more likely to have near-universal high-speed access, the relationship is weak and noisy. The spread in wealth at high coverage levels – and the presence of lower-wealth counties with decent access – suggests that wealth alone does not fully explain broadband disparities.

Correlation Results

To further quantify the relationship between wealth and broadband access, I calculated Pearson correlation coefficients between **speed_100_20** and two individual economic indicators:

- **High-speed access and median income:** $r = 0.047$
- **High-speed access and poverty rate:** $r = -0.041$

As expected, the correlation with poverty is negative – counties with higher poverty tend to have less access – but the strength of both relationships is extremely weak. This aligns with the visual results: broadband access may be influenced by wealth, but the effect is small and inconsistent at the national level.

Together, these visualizations and statistics suggest that while economic advantage might play a role in shaping broadband access, other factors (such as geography, infrastructure investment, or provider competition) likely play a more dominant role, and should be investigated in future analysis.

Summary and future exploration

This preliminary analysis suggests that access to high-speed broadband is only weakly correlated with county-level economic indicators like median income and poverty rate. While some wealthier counties do have slightly better coverage, many lower-wealth areas are also well-connected, and some affluent counties still face access gaps. This indicates that wealth alone is an insufficient predictor of broadband infrastructure, pointing to the influence of other structural or geographic factors. Future exploration could take several directions: analyzing urban vs. rural disparities, examining regional patterns (e.g., South vs. Northeast), or building a time series view of how access has changed over multiple years. Incorporating additional variables such as population density, federal or state investment levels, or racial and demographic characteristics could also enrich the analysis and provide more actionable insights for policymakers.

Bibliography

Federal Communications Commission. 2024. "Fixed Broadband Summary by Geography Type." National Broadband Map. Accessed May 4, 2025. https://broadbandmap.fcc.gov/data-download/nationwide-data.

National Telecommunications and Information Administration. 2025. Accessed May 8, 2025. https://www.ntia.gov.

U.S. Census Bureau. 2025. "American Community Survey (ACS) 5-Year Data (2018–2022)." Census Bureau API. Accessed May 4, 2025. https://www.census.gov/data/developers/data-sets/acs-5year.html.