**Broadband Coverage in the US: When Sources Disagree**

DSPG team: Sarah McDonald, Raghav Sawhney, Kateryna Savchyn

SDAD team: Teja Pristavec, Josh Goldstein, Stephanie Shipp

**We Don’t Know How Many Americans are Without Broadband**

The Federal Communications Commission (FCC) provided [$1.49 billion](https://docs.fcc.gov/public/attachments/DOC-353840A1.pdf)in 2018 to fund broadband development in rural areas. The same year, the United States Department of Agriculture (USDA) created the [ReConnect Program](https://www.usda.gov/reconnect/program-overview), providing $600 million in funding for broadband improvement in rural areas.

Despite these programs and the agencies’ efforts to expand broadband coverage, internet at broadband speeds—defined as 25mbps download and 3mbps upload—is still not available to many Americans. However, estimates differ on how many individuals are without access to broadband, and thus limited in their ability to participate in today’s increasingly online world.

Typically, broadband availability estimates come from the FCC. The FCC provides an authoritative data source on broadband coverage and subscriptions that informs funding policies. However, the dataset has known limitations. The FCC reports a block group as having broadband available even if only one subscriber has it, if the provider is not currently providing coverage but could feasibly start doing so in an area within a standard service interval, and it collects advertised rather than actual internet speeds. These factors contribute to potential coverage overestimation. Whereas the FCC claims that approximately 25 million Americans do not have broadband access, the Microsoft Airband Initiative finds that around 163 million Americans do not use internet at broadband speeds.

Broadband access is associated with beneficial outcomes like improved firm productivity, increased employment opportunities, higher political participation, and better education outcomes. Given that broadband facilitates participation in an increasingly online economy, a better understanding of the state of US broadband coverage is needed. Accurately identifying areas with low broadband access has implications for broadband policies and funding, and is particularly important for rural communities struggling to be competitive.

To investigate where US broadband coverage may not be good, our Data Science for Public Good team set out to investigate three broadband data sources: the FCC, the American Community Survey (ACS), and Microsoft Airband data. Our aims were to understand the extent of coverage according to each dataset, to examine discrepancies between FCC and its alternatives, and to address both aims with particular attention to rural areas.

**Three Datasets Measuring Broadband Coverage**

We used three public data sources in comparing US broadband coverage, each available at different geographic levels and reported by a different party. The FCC collects broadband availability and subscription data from providers, and makes data across different years at block group, census tract, and county levels. For better comparisons with other sources, we 2015 data. Our second source, the US Census Bureau’s ACS, is an annual, nationally representative US population survey. It provides estimates on population sociodemographic characteristics and select topics, including internet access. Contrary to FCC, it relies on self-reports. Broadband estimates became available for the first time at census block level in its 2013-2017 (5-year) data, which we use in our analyses. Finally, the Microsoft Airband Initiative data is neither provider- nor consumer-reported; instead, it comes from Microsoft’s one-time initiative to collect data on broadband coverage using customer access. Microsoft analyzed its server logs when electronic devices downloaded Microsoft Windows and Office updates, accessed Microsoft’s Bing search, or used Xbox gaming consoles. The company aggregated their 2018 data at the county level.

**Capturing Coverage and Discrepancies**

Our challenge in comparing broadband coverage across three data sources—each using different geographic levels and methods of obtaining information—was constructing indicators that capture coverage in similar ways.

We calculated FCC broadband coverage as the proportion of the population with access to at least one provider offering at least 25mbps download speed. We used FCC provider-reported maximum advertised downstream speeds, coupled with ACS population estimates. Similarly, we calculated ACS broadband coverage as the proportion of the population self-reporting access to a broadband connection (excluding satellite and cellular). Our Microsoft broadband coverageindicator captures the proportion of population that used Microsoft services at 25mbps broadband download speed, according to Microsoft server logs. We calculated each metric for every available geographic level.

We constructed reported coverage discrepancy measures to examine whether and how broadband access information differs across data sources. Our reference point was FCC coverage as the source used in official statistics and policymaking. We constructed FCC-ACS and FCC-Microsoft broadband coverage discrepancy measures indicating the percent discrepancy in reported coverage between the dataset pairs. We calculated each metric at each geographic level available.

To better understand disadvantage in broadband access for rural versus urban communities, we also classified geographies according to USDA [Rural-Urban Continuum Codes (RUCC)](https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/). RUCC codes classify metropolitan counties by the population size of their metro area, and nonmetropolitan counties by degree of urbanization and adjacency to a metro area. We grouped rural and urban codes to construct an indicator of whether a geography was a metro or nonmetro area.

**Millions Without Broadband, But How Many?**

Figure 1 shows the overall county distribution of broadband coverage according to FCC provider reports, ACS consumer reports, and Microsoft server logs.



Figure 1: County distribution of broadband access by data source.

As Figure 1 shows, the FCC suggests an overall higher broadband coverage within counties compared to ACS and Microsoft data. The FCC receives provider-reported data on advertised coverage from providers, which may introduce a positive coverage bias if advertised estimates do not match actual estimates, and it has [other known data quality issues](https://www.theverge.com/2019/5/30/18644726/fcc-broadband-report-high-speed-rural-statistics-reactions) potentially leading to overestimates. Accordingly, the FCC claims that broadband is not available to 25 million Americans, whereas the Microsoft Airband Initiative—the data source which suggests the lowest broadband coverage—finds that around 163 million Americans do not use internet at 25 mbps. Given that Microsoft’s data is [based on server logs of Bing, Microsoft Update, and XBox use](https://www.dailyyonder.com/rural-download-speeds-worse-reported-microsoft-study-says/2019/02/05/30193/), this sample likely excludes non-Windows users, and may underreport broadband coverage. ACS shows coverage between the FCC and Microsoft extremes, but similarly to Microsoft, it suggests that US broadband coverage is lower than FCC reports.

**Numbers Aside, Urban Areas Doing Better than Rural Ones**

Even though the datasets differ in their estimates of how US broadband coverage, they all show an urban-rural disparity. Figure 2 visualizes broadband coverage reports by [urbanicity](https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/) and shows that on average and across all three datasets, metro areas have higher broadband coverage compared to non-metro regions.

 Figure 2: County distribution of broadband access by data source and urban status.

**Exploring Dataset Discrepancies Through an Interactive Dashboard**

We developed a [dashboard to illustrate our discrepancy findings](http://bband.policy-analytics.net). Users can select a state of interest, choose a geography, and subset visualizations into urban and rural areas. Each visualization overlays major cities in each state. At each geographic level, users can view all available data for the selected area by hovering over a county, tract, or block group. Figure 3 is an example from the dashboard, showing the FCC-ACS broadband coverage discrepancies for Alabama. Our maps show that the range of discrepancies tends to be larger in rural versus urban areas.



Figure 3: Alabama FCC-ACS Broadband Coverage Discrepancy Map at Block Group Level.

**Discussion and Next Steps**

Discrepancies in reported broadband coverage exit, but that regardless of data source, urban areas appear to have higher coverage than rural ones. These report discrepancies are larger in rural than urban areas. Our next step is to develop heuristics for FCC data reliability and provide tools for identifying where FCC data should be supplemented or improved, using the three data sources to obtain model-based broadband access predictions based on urbanicity and sociodemographic information. Having coverage data that is accurate and high quality is needed to ensure that decision-makers direct [broadband development resources](https://broadbandusa.ntia.doc.gov/new-fund-search) to areas that need it most.