PROJECT PLAN: KANSAS BROADBAND ACCESS

ABSTRACT

Recent research by the Institute for Policy & Social Research (IPSR) at the University of Kansas has revealed significant discrepancies between Federal Communications Commission (FCC) broadband access estimates and actual on-the-ground connectivity in Kansas. While the FCC claimed that 85.8% of Kansans have access to broadband speeds of 100+ Mbps download, IPSR's survey data suggested only 43.5% have such access, with the two sources agreeing only 25% of the time. This project aims to provide an additional independent validation of Kansas broadband access using Ookla Speedtest data—real-world speed measurements collected from users' mobile devices connected to home internet services. By analyzing geographic patterns in actual internet speeds and comparing them with FCC maps, Microsoft usage data, and IPSR survey results, this project will strengthen the evidence base for policy decisions regarding broadband infrastructure investment in Kansas and provide a methodology for more accurate assessment of rural broadband gaps.

1 INTRODUCTION

The digital divide in Kansas represents a critical barrier to economic development, educational opportunity, and quality of life, particularly in rural communities. Recent research has exposed fundamental problems with how broadband access is measured and reported, with profound implications for how federal and state resources are allocated to address connectivity gaps.

The IPSR study, *Broadband Access in the State of Kansas* (January 2023), documented systematic overestimation of broadband availability by existing data sources. The FCC's maps report maximum advertised speeds but do not reflect actual service availability or experienced speeds. Microsoft's broadband data, derived from software update patterns, uses a somewhat outdated definition of broadband (25/3 Mbps) and shows that fewer than 50% of Western Kansas households connect at even these speeds. IPSR's own survey of 6,895 Kansans found widespread dissatisfaction with service (46% dissatisfied or very dissatisfied) and revealed that over 1,000,000 Kansans live in regions with less than 100 Mbps download speeds—the modern standard for adequate connectivity.

This project addresses a critical gap in the evidence base by introducing a third independent data source: Ookla Speedtest measurements. Unlike FCC maps (which show theoretical availability) or Microsoft data (which captures only Windows update traffic), Ookla Speedtest data represents actual speed measurements conducted by users testing their connections. These tests are performed via the Speedtest mobile app (both iOS and Android), with phones often connected via WiFi to fixed home internet infrastructure (cable, fiber, DSL, satellite, or fixed wireless). Ookla aggregates these measurements at the geographic tile level (“map tiles”), providing spatially detailed information about real-world internet performance.

The goals for this project are three-fold:

1. **Validate and contextualize existing findings**: Compare Ookla Speedtest measurements with FCC availability maps, Microsoft usage data, and IPSR survey results to determine the degree of agreement or disagreement across data sources and identify which sources provide the most accurate picture of Kansas broadband access.
2. **Map the geography of actual internet speeds**: Create detailed visualizations of Ookla speed test results across Kansas, identifying areas where actual speeds fall short

of modern broadband standards (100/20 Mbps) and examining the urban-rural divide in internet performance.

3. **Provide evidence for policy recommendations**: Strengthen the case for targeted broadband infrastructure investment by demonstrating through multiple independent data sources that large portions of Kansas lack adequate internet connectivity, contradicting FCC estimates that might otherwise suggest the problem is less severe than it actually is.

This project is particularly timely as Kansas policymakers are considering how to allocate unprecedented levels of federal broadband funding. Accurate data is essential to ensure these resources reach the communities most in need. By demonstrating convergent evidence from multiple sources—each with different methodologies and potential biases—this analysis will provide a more robust foundation for decision-making than any single data source could offer alone.

The relevance of this work extends beyond Kansas. The methodological challenge of accurately measuring broadband access affects rural communities nationwide, and the FCC's measurement approach has been criticized across many states. This project's approach of triangulating multiple data sources—administrative data (FCC), usage data (Microsoft), survey data (IPSR), and performance testing data (Ookla)—offers a replicable model for other states seeking to challenge incomplete or inaccurate federal broadband maps and advocate for their underserved communities.

2 Baseline or Initial Analysis

The internet speed data from Ookla and Microsoft are publicly available (see below for links). The first step is to extract Kansas data from the map tile speed-test data from Ookla. Ookla collects both “mobile” and “fixed” data, and “fixed” means that the cell phone performing the speed test was connected to a fixed internet source via WiFi, such as cable or satellite. We wish to focus on the “fixed” data set because this is best for comparing to other sources of internet speed data, in particular the data that Microsoft collects from Windows computers. The map tile data from the Ookla datasets can be used to extract the tiles that correspond to Kansas county shape files. (See below for resources for using R or Python for this type of GIS manipulation.) This will allow the joining of the Ookla speed test data to other county-level Kansas data, such as the data set from Microsoft as well as data from the American Community Survey (ACS). The ACS data includes estimates of households that have “broadband” access (however, all internet other than dial-up was considered “broadband,” which is a very outdated definition). The ACS data helps to interpret the Microsoft data, which shows the percent of the internet speeds that are truly “broadband” (by the FCC definition of 25/3 Mbps). By joining the two data sets, we could extrapolate the percentage of households with (non-dial-up) internet that are at broadband speeds. The Microsoft data set also includes the FCC estimate of broadband availability, as a percentage of people for each county, so these three data sources can be joined into a single data table for comparison analysis.

To get the most accurate speed estimates from Ookla, the data from the four quarterly data sets should be combined for each year. Initial focus should be on 2020 because this allows comparison to the Microsoft data set (derived from one month in 2020), but more recent years from Ookla should also be included for analyses, because these can be combined with ACS data and also compared with data from the IPSR study.

3 Final Analysis

Once the Ookla data has been extracted and combined with Microsoft, FCC, and ACS data at the county level (as described in Section 2), the analysis should focus on visualizing and comparing internet speeds across these different data sources. Maps should be created showing the distribution

of Ookla speeds across Kansas counties, using both quantile-based color coding (to show the full distribution of speeds) and the five speed ranges used in the IPSR report (to enable direct comparison with their survey findings). These visualizations will highlight geographic patterns in internet access, particularly the urban-rural divide, and will help identify counties where different data sources agree or disagree about broadband availability.

Additional analysis should examine correlations between data sources—for example, comparing Ookla median speeds with Microsoft's broadband usage percentages and FCC availability estimates to quantify the degree of agreement across sources. Temporal comparisons using multiple years of Ookla data may reveal trends in improving (or stagnant) internet speeds. Summary statistics should document the percentage of Kansas counties and population that fall below various speed thresholds (25/3 Mbps, 100/20 Mbps), providing updated estimates that can be compared against FCC claims and IPSR survey findings. The goal is to determine whether Ookla data corroborates the IPSR finding that FCC estimates significantly overstate actual broadband access in Kansas.

4 Final Goals & Evaluation

The deliverables for this project include well-documented code files (R or Python) used to generate the wrangled datasets. It is recommended that the code be organized into at least two files: one focused on the GIS work (extracting Kansas map tiles from Ookla data and joining to county geometries), and another containing all other data wrangling and analysis steps. This separation makes the code more maintainable and reusable.

Visualizations are a key deliverable. Maps should show internet speeds at the county level using appropriate color schemes, with separate maps for different years to track changes over time. Comparison maps that overlay multiple data sources (e.g., side-by-side views of Ookla speeds vs. FCC availability) would be particularly valuable for communicating discrepancies between sources.

A written report (5-10 pages) summarizing the findings should accompany the code and visualizations. The report should describe the methodology, present key findings about broadband access in Kansas, highlight areas of agreement or disagreement between data sources, and discuss policy implications. The report should be accessible to policymakers and advocates who may not have technical backgrounds.

5 Related Work

1. *Broadband in Kansas: The Challenges of Digital Access and Affordability (*University of Kansas Institute for Policy & Social Research, March 2023) by Donna Ginther *et al.*

   https://kuscholarworks.ku.edu/entities/publication/993b65ce-f011-414e-957d-fe4b6ab02564

   This report details the results of survey data, focus groups, and interviews that focused on broadband internet access in Kansas after the pandemic years, and compares the results with other available data sources. It also makes detailed policy recommendations with justifications.

2. *Microsoft Airband Initiative*

   https://www.microsoft.com/en-us/corporate-responsibility/airband-initiative

   Microsoft's "Digital Access In the US" initiative is a part of this larger world-wide initiative for closing the global digital divide. Their work of collecting and processing internet speed data in US Counties was part of this initiative. The data is available (from 2019 and 2020 only) from the project's GitHub page: https://github.com/microsoft/USBroadbandUsagePercentages?tab=readme-ov-file

3. *Report from the Federal Communications Commission (FCC) on broadband deployment: Fourteenth Report:* https://docs.fcc.gov/public/attachments/FCC-21-18A1.pdf

This report from January 2021 details the status of both fixed and mobile broadband access in the United States. Broadband in the report is defined as 25 Mbps download, 3 Mbps upload speed. The report also details the efforts taken to close the digital divide. The data files are not needed from the FCC because the needed FCC data is incorporated into the datasets from Microsoft.

4. *The Ookla Open Data Project* (GitHub page):

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

   This web page is a rich source of references for obtaining and working with Ookla's Speedtest data.

5. *Analyzing US Census Data: Methods, Maps, and Models in R,* CRC Press 2023, by Kyle Walker.

   This is a very helpful book that shows how to use the R programming language and resources from the US Census to prepare, map and present map data.

6 Data & Technical Requirements

This project requires extensive data wrangling skills for combining data sets from different sources. It also requires skill with GIS. The Ookla GitHub page listed in this section provides links to brief tutorials for using either Python or R with GIS methods. The tutorials demonstrate working with either Parquet files or Shapefiles. Shapefile (SHP) is a widely-adopted format for sharing geospatial data, and it is supported by nearly every GIS-capable software and library, so this is recommended format when working in R or Python. Some useful R packages include tigris (connects to Census API, used to get county shapes) and sf (spatial functions).

The Speedtest data from Ookla is available via links on their GitHub project page, "Speedtest by Ookla Global Fixed and Mobile Network Performance Map Tiles." https://github.com/teamookla/ookla-open-data. There are quarterly datasets for each year starting from 2019 until the current year, and more data is added shortly after the conclusion of the quarter. The data is collected using both Android and iOS phones using the Ookla app, and it is separated into "fixed" and "mobile" data sets. (The cell phone provides GPS location data, and it could be connected to a fixed source via WiFi). Both Apache Parquet and Shapefiles are available.

The Microsoft Airband Initiative has downloadable CSV files for October 2020 on the GitHub page for the Broadband Usage project: https://github.com/microsoft/USBroadbandUsagePercentages?tab=readme-ov-file. The data file includes "broadband availability per FCC," which is the percent of people per county with access to fixed terrestrial broadband speeds, according to the Fourteenth Broadband Deployment Report linked to above, as well as "broadband usage", which is the percent of people that actually use internet at broadband speeds (FCC definition) as measured by Microsoft.

Resources for obtaining and interpreting American Community Survey Data: The Census Bureau releases two ACS datasets to the public: the 1-year ACS, which covers areas of population 65,000 and greater, and the 5-year ACS, which is a moving average of data over a 5-year period that covers geographies down to the Census block group.

https://www.census.gov/acs/www/methodology/sample_size_and_data_quality/

Downloading data from the Census is generally not a simple matter. You must pick the parameters, and then a data table is generated for you. A helpful resource for using this resource:

https://data.census.gov/table?tid=ACSDP1Y20https://www.census.gov/data/what-is-data-census-gov/guidance-for-data-users/how-to-materials-for-using-data-census-gov.html21.DP02