BROADBAND DATA VALIDATION: COMPARING US BROADBAND COVERAGE

DSPG Team: Kat Savchyn (Georgetown University), Sarah McDonald (Marymount University), Raghav Sawhney (Virginia Tech)

SDAD Team: Teja Pristavec, Josh Goldstein, Stephanie Shipp

SPONSORS: John Pender, Peter Hahn, Zachary Keeler, USDA Economic Research Service

Introduction

Does broadband development affect rural prosperity, as measured by changes in property values and quality of life indicators in rural communities?

- Conducted a **comprehensive literature review of value modeling** consisting of factors other than broadband that shape property values. We found relevant environmental, neighborhood, structural, and locational attributes that influence property prices.
- Explored methods of **validating Federal Communications Commission (FCC) broadband data, using multiple data sources.**

Data

FCC FORM 477

- Self-reported data from broadband providers, potentially introducing error if advertised coverage estimates are reported rather than actual estimates.
- If a census tract had just one subscriber, the entire census tract is assumed to have broadband.
- Claims that broadband is not available to 25 million Americans.

American Community Survey

- Table B28002: Presence and Types of Internet Subscriptions in Household for the 2013-2017 data specific to broadband.
- Self-reported data from households, but does not include group quarters.

BroadbandNow

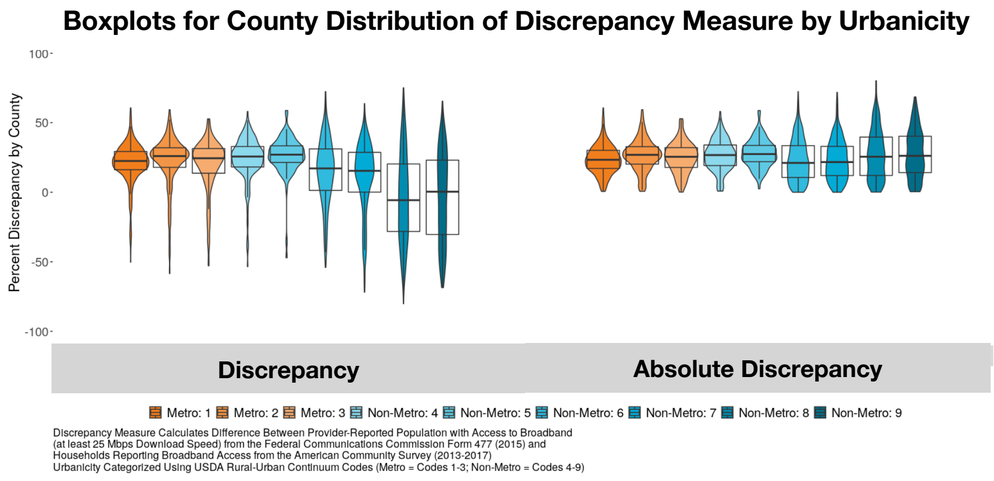
- Online search engine which lists information for broadband availability on the zip code and city level.
- Provides information on:
  - number of providers,
  - maximum and average upload and download speeds
  - change in broadband speeds over time.

Microsoft Airband

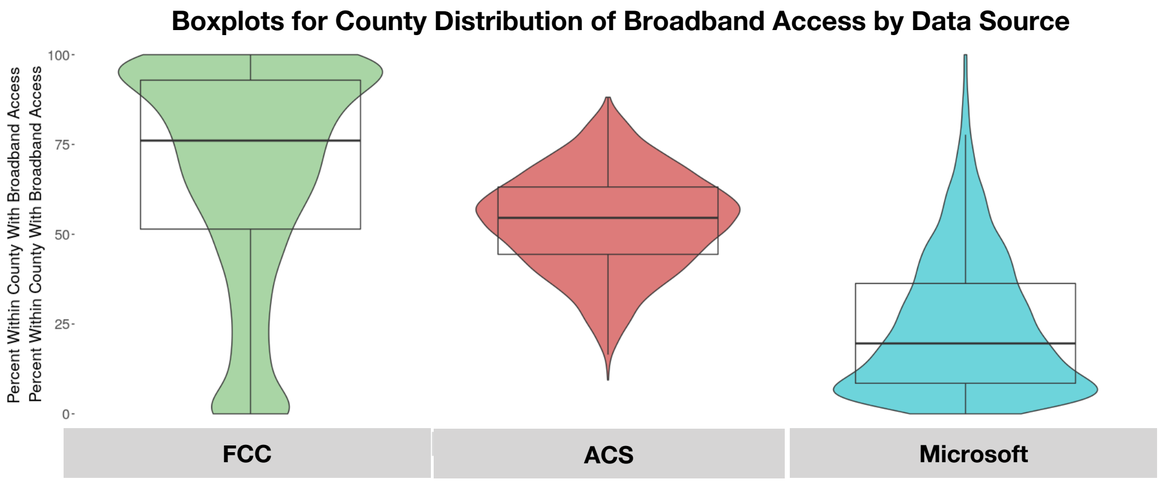
- Finds that 163 million Americans do not use the internet at 25 mbps, or broadband speed.
- Analyzing its servers logs when electronic devices downloaded Microsoft Windows and/or Office updates, as well as when devices used Microsoft's Bing search engine and Xbox gaming consoles.

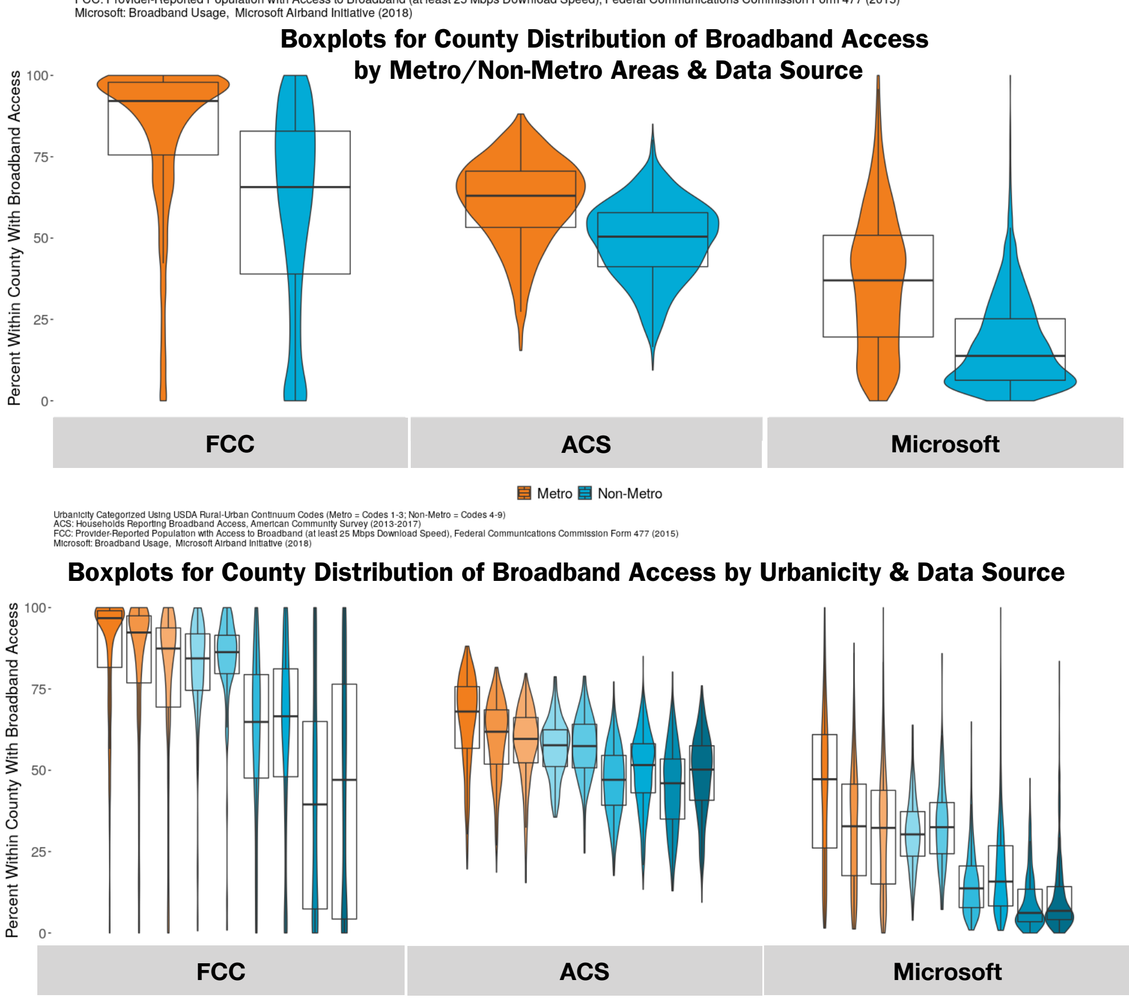
Validation Methodology

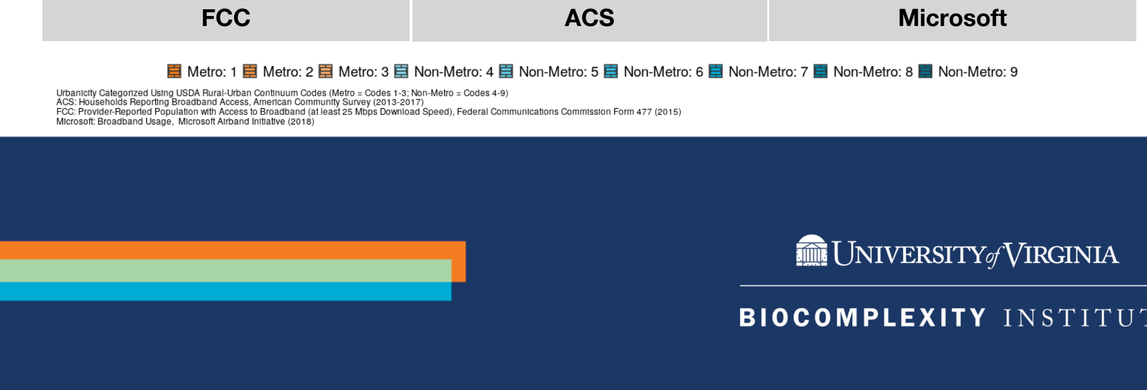
- Acquired and aggregated data at county, city, census tract, and block group levels to observe differences in coverage.
- Developed FCC coverage metric using a threshold of 25 mbps download speed and supplemented with American Community Survey data to determine % of population with broadband access. The absolute difference in percent between this metric and other data sources identifies level of discrepancy:
  1) ACS access to broadband (excluding cellular and satellite)
  2) FCC internet service connections of at least 10 mbps [block group, census tract]
  3) Microsoft usage percentage [county]
  4) BroadbandNow coverage [city]
- Examined data by USDA-specified Rural-Urban Continuum Codes (RUCC). RUCC codes distinguish metropolitan counties by the population size of their metro area, and nonmetropolitan counties by degree of urbanization and adjacency to a metro area.



Findings

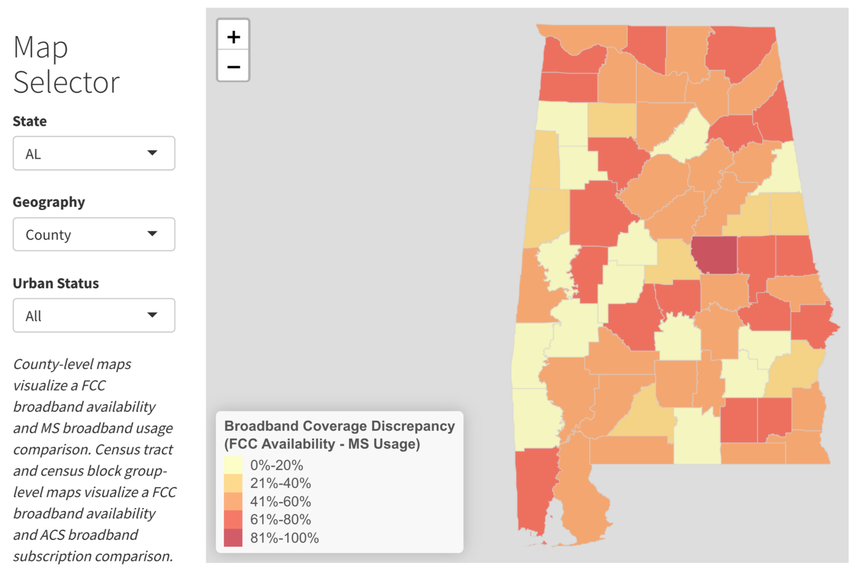






Interactive Dashboard

We developed an interactive dashboard that visualizes discrepancies in broadband coverage reports at county, census tract, and census block levels. The dashboard allows users to select a state and level of geography, and to filter by urban or rural area.



The dashboard is available at http://bband.policy-analytics.net.

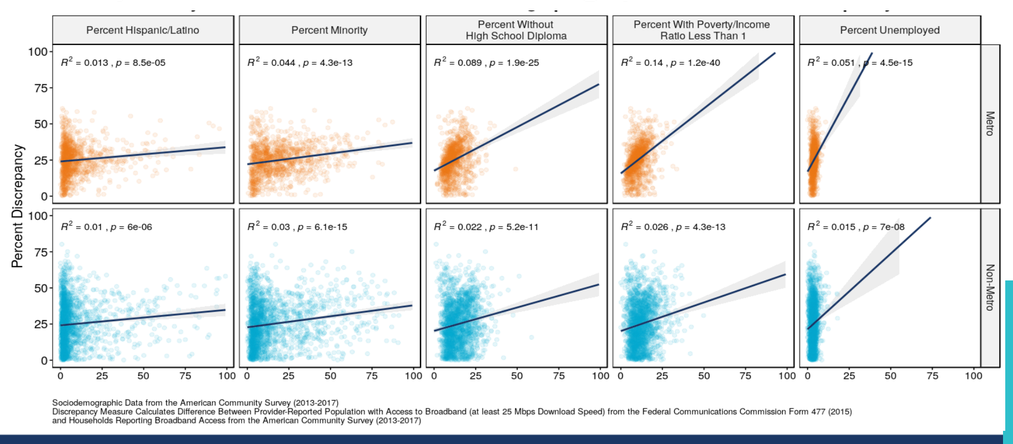
This map shows differences in % broadband availability (FCC) and % broadband usage (Microsoft) for Alabama at county level.

Discussion & Further Research

Findings: FCC, ACS, and Microsoft Airband data give very inconsistent estimates of broadband accessibility. However, they all estimate lower coverage in rural areas compared to urban areas.

Next steps: Explore the relationship between percent discrepancy and sociodemographic factors for counties by urbanicity. We provide our preliminary exploration of discrepancy and sociodemographic factors correlations below. Sociodemographic factors are more likely to influence discrepancies in metro than non-metro areas.

We use FCC, ACS, and Microsoft data to obtain model-based broadband access predictions based on sociodemographic information (e.g. urbanicity, unemployment).



UNIVERSITY of VIRGINIA
BIOCOMPLEXITY INSTITUTE
Data Science for the Public Good

http://bband.policy-analytics.net/