

Big Data, Little Spaces, High Speed: Using Stata to analyze the determinants of broadband access in the United States

Thursday, July 26, 2012

Stata Conference

David Beede

Brittany Bond

Economics and Statistics Administration

U.S. Department of Commerce

DRAFT

NOTE: The statements and opinions expressed in this presentation are solely our own, presented in our personal capacity, and are not intended to represent the official views or policies of the Economics and Statistics Administration, the U.S. Department of Commerce or the Executive Branch.

Table of Contents

- Introduction.
- Data.
- Adoption of Broadband by Race/Ethnicity in US.
- Availability of Broadband by Race/Ethnicity in US.
- Focus on Fayetteville, NC Metropolitan Statistical Area using Maps
- Focus on Fayetteville, NC Metropolitan Statistical Area using Really Preliminary Simplistic Regressions

Introduction

- This project is part of an ongoing partnership between my agency, Economics and Statistics Administration (ESA) and another Commerce Department agency, the National Telecommunications and Information Administration (NTIA), which funds this project.
- NTIA and ESA have been and are continuing to estimate household broadband technology adoption rates and their relationships with household economic and demographic characteristics.
- Today I am introducing more recently inaugurated joint NTIA-ESA work estimating broadband technology access its relationships with economic and demographic characteristics of households and workplaces.

Summary

- Newly available Commerce Department data provide detailed information about availability of broadband service throughout the US.
- These data can be merged with Census Bureau demographic and economic data on households and workplaces.
- The combined data reveal interesting new insights on broadband service availability.



National Broadband Map v.3.0
How connected is my community?



Longitudinal Employer-Household Dynamics



**AMERICAN
COMMUNITY
SURVEY**
U.S. CENSUS BUREAU

Merge

Max. Available Advertised
Wireline and Wireless Upload
and Download Speeds by Census
Block in June 2011

Number of Jobs (Total and by
Industry) by Census Block of
Workplace Location in 2010

Population by Age, Sex, Race,
Ethnicity by Census Block of
Residence in 2010

Income, Educational Attainment,
Occupation, Commuting Time by
Census Block Group, 2006-2010
Average



- www.broadbandmap.gov
- Updated semi-annually

For each Census Block, multiple records with:

- Service Provider (Direct and Holding Company Name and FCC ID Number)
- Broadband Delivery Technology
- Advertised Upload/Download Speeds



Challenges

- Multiple speeds/Census Block
- Wireless: Census blocks on margins of coverage areas had <100% coverage
- Wireline: Service data reported by providers for only a sample of points in large (> 2 square mile) Census blocks

Longitudinal Employer-Household Dynamics

Economic Census;
Quarterly Census of Employment
and Wages (QCEW);
Covered Employment and Wages
(CEW);

Survey of Income and Program
Participation (SIPP);
March Current Population
Survey (CPS)

Unemployment Insurance (UI)
system;
State records on jobs and firms;
Person Characteristics File (PCF);
Composite Person Record (CPR);

Censuses

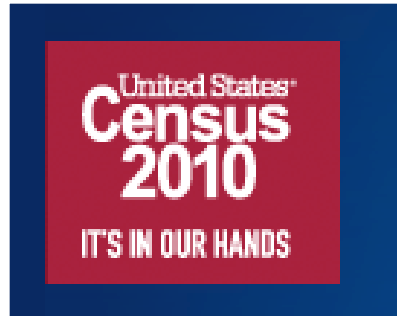
Surveys

**Administrative
Records**

**Longitudinal Employer-Household Dynamics Program
(LEHD)**

Local Employment Dynamics

<http://lehd.did.census.gov/led/onthemap/LODES6/>



- We extracted data on basic demographics - age, sex, race, ethnicity - from the 2010 Decennial Census Summary File 1 data (at Census Block level).



- American Community Survey replaces former “long form” Decennial Census data.
- Includes data such as income, educational attainment, occupation, commuting time.
- Used 2006-2010 ACS data which data are available at the Census Block Group level

Appendix E: Detailed Tables for the 2006-2010 ACS 5-Year Estimates

Table Number	Table Title	Table Restrictions for All Geographies Except Block Groups	Table Available for Block Groups on Summary File?	Summary File Sequence Number	Summary File Starting and Ending Positions
B00001	Unweighted Sample Count Of Population	See Note 1	Yes	9	7-7
B00002	Unweighted Sample Housing Units	See Note 1	Yes	9	8-8
B01001	Sex By Age		Yes	10	7-55
B01001A	Sex By Age (White Alone)			10	56-86
B01001B	Sex By Age (Black Or African American Alone)			10	87-117
B01001C	Sex By Age (American Indian And Alaska Native Alone)			10	118-148
B01001D	Sex By Age (Asian Alone)			10	149-179
B01001E	Sex By Age (Native Hawaiian And Other Pacific Islander Alone)			11	7-37
B01001F	Sex By Age (Some Other Race Alone)			11	38-68
B01001G	Sex By Age (Two Or More Races)			11	69-99
B01001H	Sex By Age (White Alone Not Hispanic Or Latino)			11	100-130
B01001I	Sex By Age (Hispanic Or Latino)			11	131-161
B01002	Median Age By Sex			11	162-164
B01002A	Median Age By Sex (White Alone)		Yes	11	165-167
B01002B	Median Age By Sex (Black Or African American Alone)		Yes	11	168-170
B01002C	Median Age By Sex (American Indian And Alaska Native Alone)		Yes	11	171-173
B01002D	Median Age By Sex (Asian Alone)		Yes	11	174-176
B01002E	Median Age By Sex (Native Hawaiian And Other Pacific Islander Alone)		Yes	11	177-179
B01002F	Median Age By Sex (Some Other Race Alone)		Yes	11	180-182
B01002G	Median Age By Sex (Two Or More Races)		Yes	11	183-185
B01002H	Median Age By Sex (White Alone Not Hispanic Or Latino)		Yes	11	186-188
B01002I	Median Age By Sex (Hispanic Or Latino)		Yes	11	189-191
B01003	Total Population		Yes	11	192-192
B02001	Race		Yes	12	7-16
C02003	Race		Yes	12	103-121
B02005	American Indian And Alaska Native Alone For Selected Tribal Groupings			12	17-62

There are nearly 30 pages worth of tables!

An Aside on Using Stata to Access American Community Survey Summary Files

- Census provides 12,508 SAS programs (state x table x estimate vs. MOE) to access the data!
-
- But it only took one page of Stata code that takes a table number + brief label prefix, gets all necessary information from Census's "look-up table," and generates datasets for entire US with variables labeled.

Percent of Households Adopting Broadband Internet Technologies for Use at Home

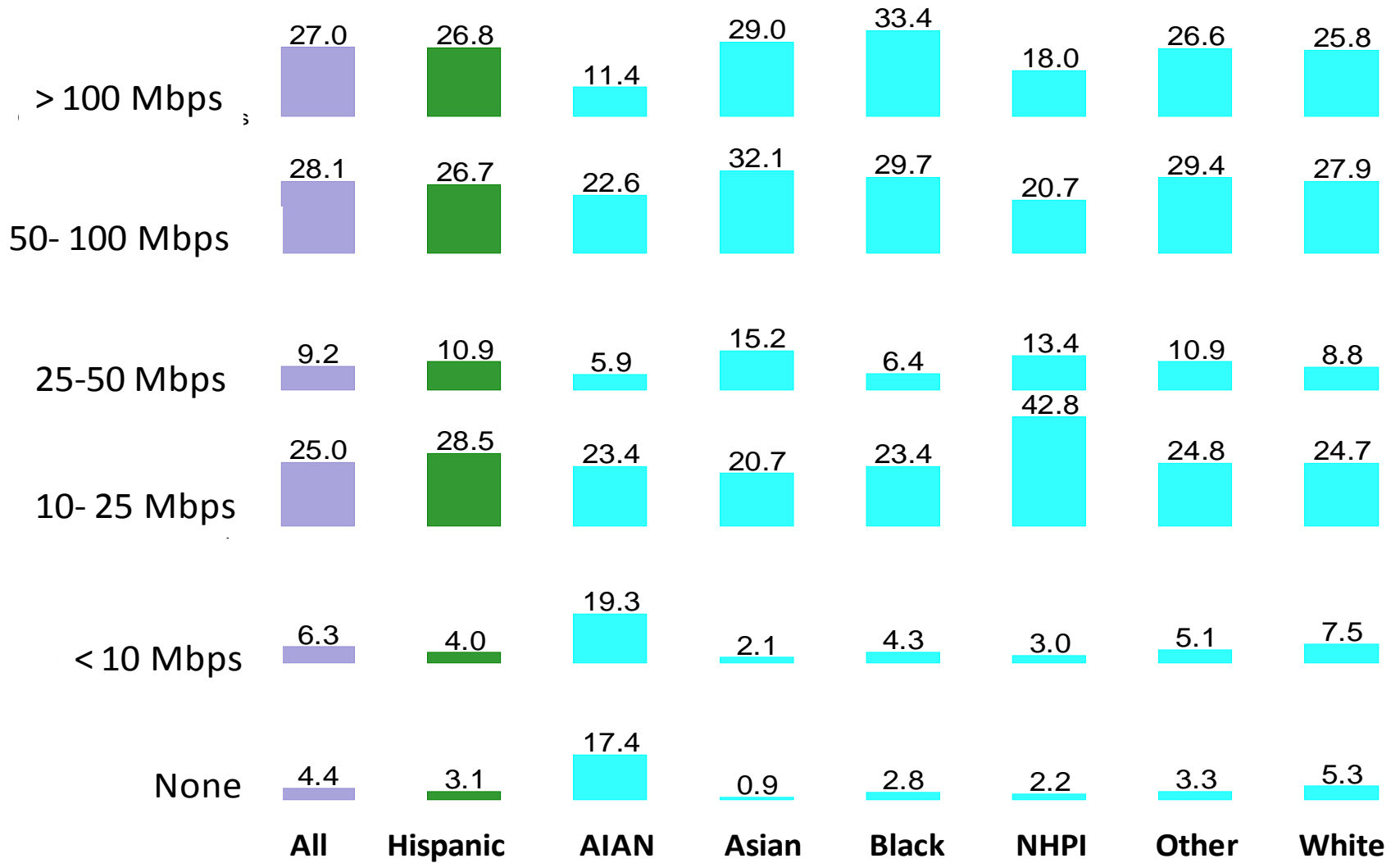
	All	Hispanic	American Indian and Alaska Native Non-Hispanic	Asian Non- Hispanic	Black Non- Hispanic	White Non- Hispanic
Oct 2010	68	57	52	81	56	72
Oct 2009	64	48	48	77	49	68

Raw and Regression-Adjusted Gaps between Non- Hispanic Whites and Other Groups

	Hispanic		Asian Non-Hispanic		Black Non-Hispanic	
	Raw	Adjusted	Raw	Adjusted	Raw	Adjusted
Oct 2010	15	11	-9	0	16	11
Oct 2009	20	14	-9	0	19	10

Sources: U.S. Department of Commerce, 2010 and 2011.

Wireline Download Speeds



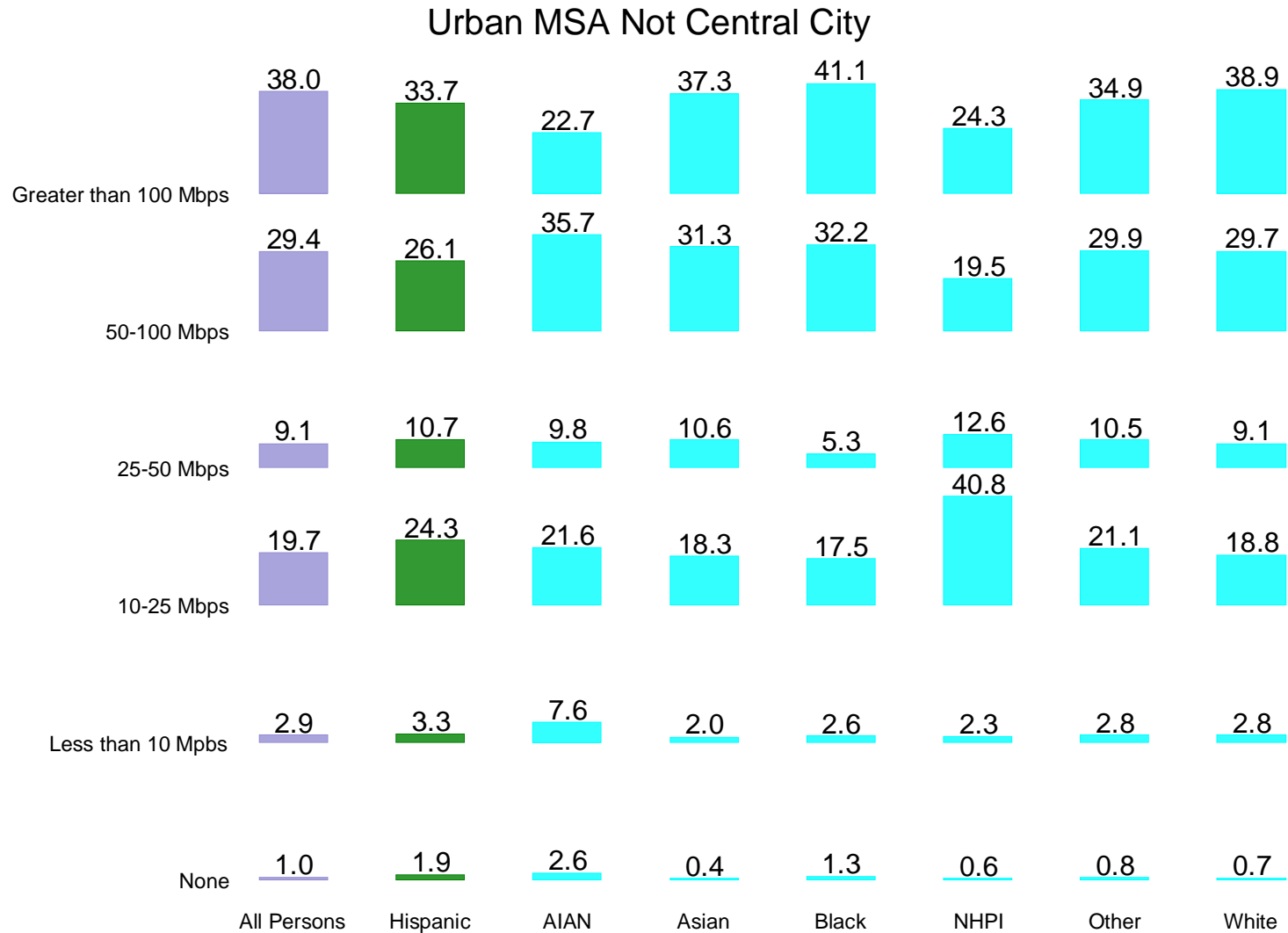
Wireless Download Speeds



Summary Statistics by Race/Ethnicity

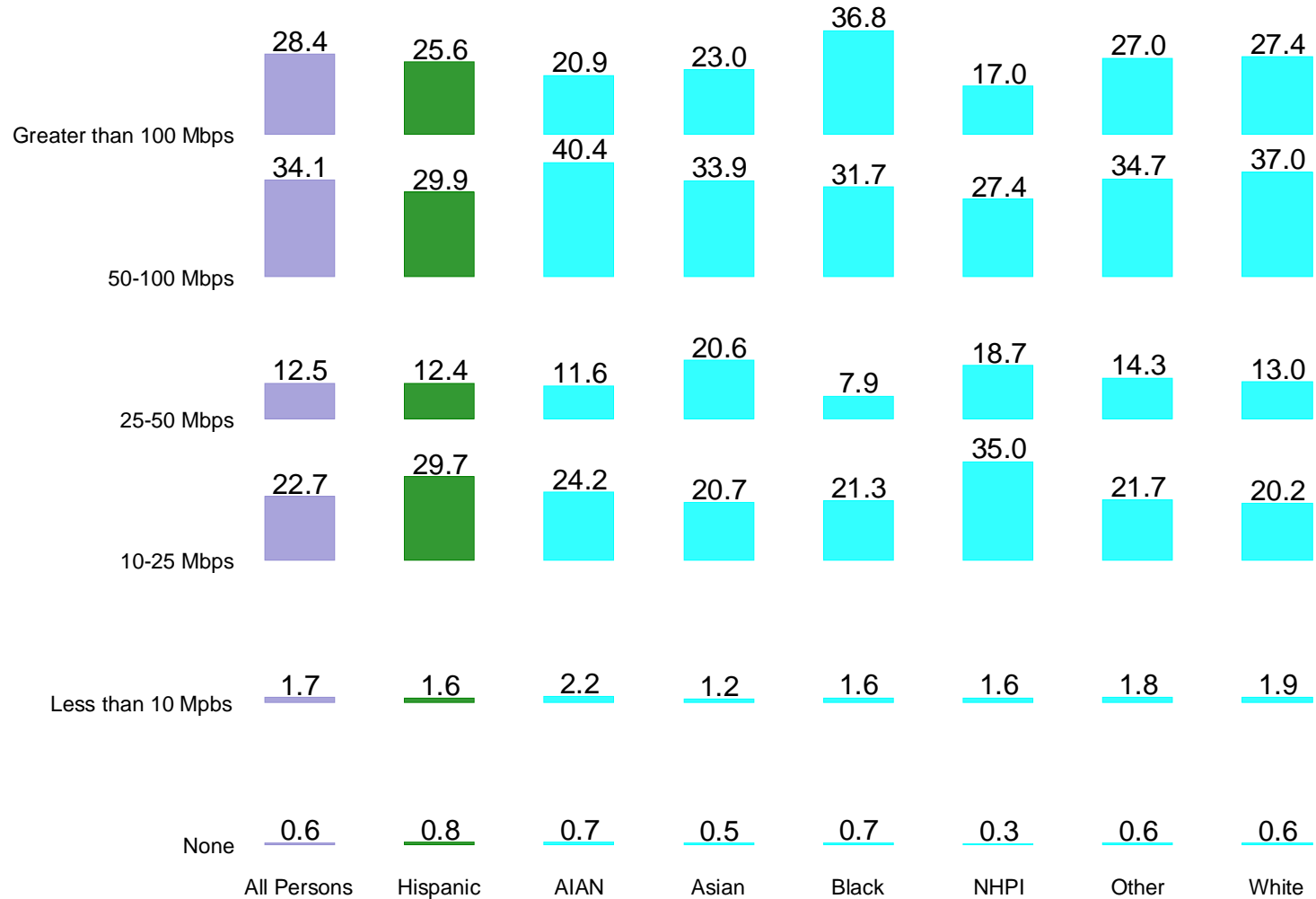
	All	Hispanic	AIAN	Asian	Black	NHPI	Other	White
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
<i>By Urban/Rural and MSA status</i>								
Rural non-MSA ("non-metro rural")	9.5	2.7	30.2	0.9	5.0	4.3	6.2	12.6
Rural non-Central City MSA ("exurbs")	9.6	3.7	12.8	1.7	4.1	3.1	6.4	12.9
Urban non-MSA ("small towns")	6.9	4.8	12.9	2.3	6.1	9.3	6.6	7.9
Urban non-Central City MSA ("suburbs")	41.4	43.4	21.7	46.4	35.0	44.6	42.6	42.2
Central City MSA ("central cities")	32.6	46.4	22.4	48.8	49.9	38.9	38.2	24.5
<i>By Commercial and Residential Status</i>								
Commercial & Residential	57.7	61.4	39.4	75.3	52.6	66.7	62.2	56.5
Residential only	42.3	38.6	60.6	24.7	47.4	33.3	37.8	43.5
<i>By Population Density Quartile (Conditional on Population Greater Than Zero)</i>								
Less than 153/sq. mi.	25.0	9.4	71.2	7.1	13.4	11.2	18.0	32.6
153-1,835/sq. mi.	25.0	15.8	21.5	18.9	22.6	17.3	22.8	28.4
1,835-5,319/sq. mi.	25.0	26.3	21.2	26.4	26.7	26.6	27.7	24.2
Greater than 5,319/sq. mi.	25.0	48.5	16.2	47.6	37.3	44.9	31.5	14.8

Wireline Download Speeds



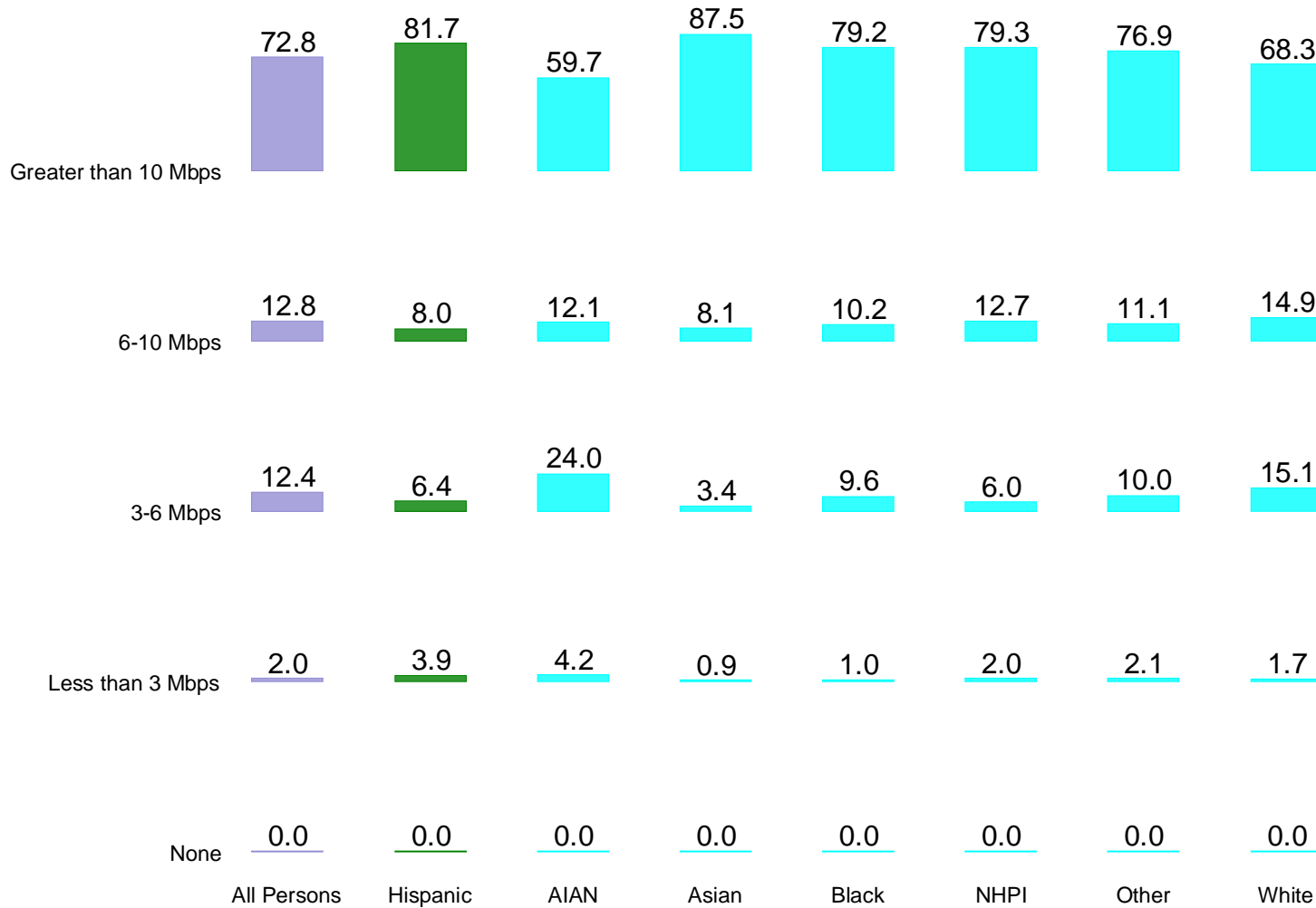
Wireline Download Speeds

MSA Central City

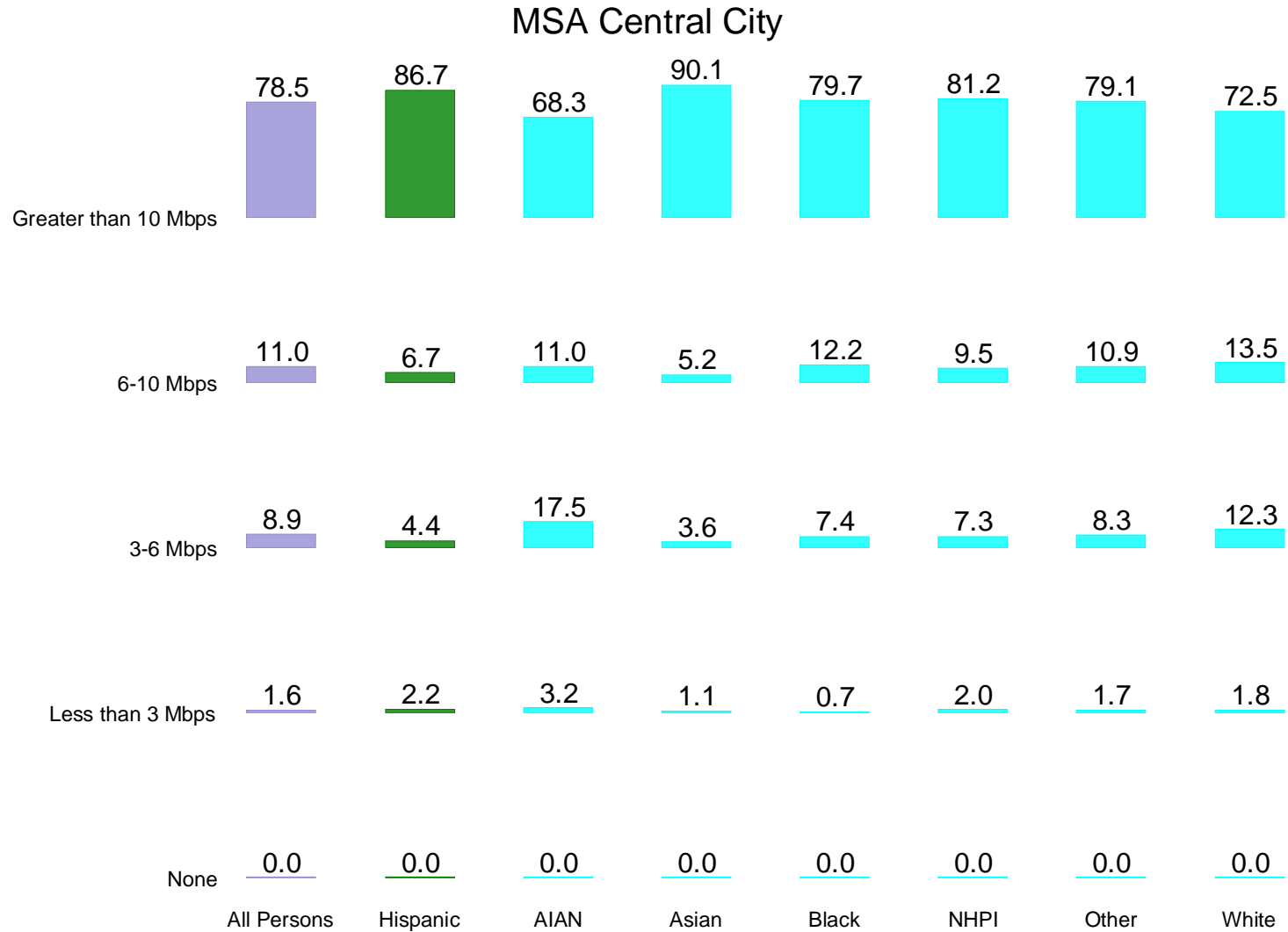


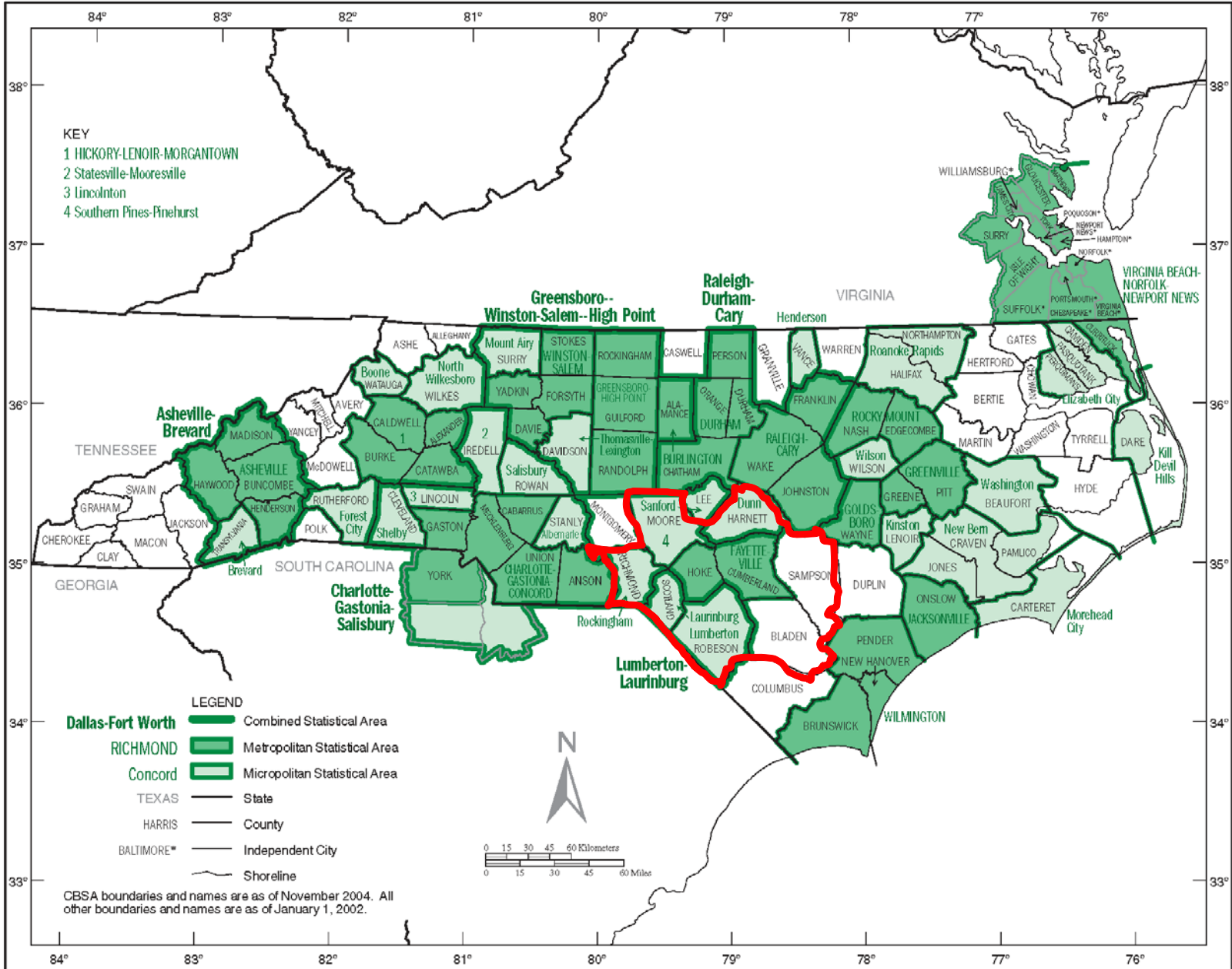
Wireless Download Speeds

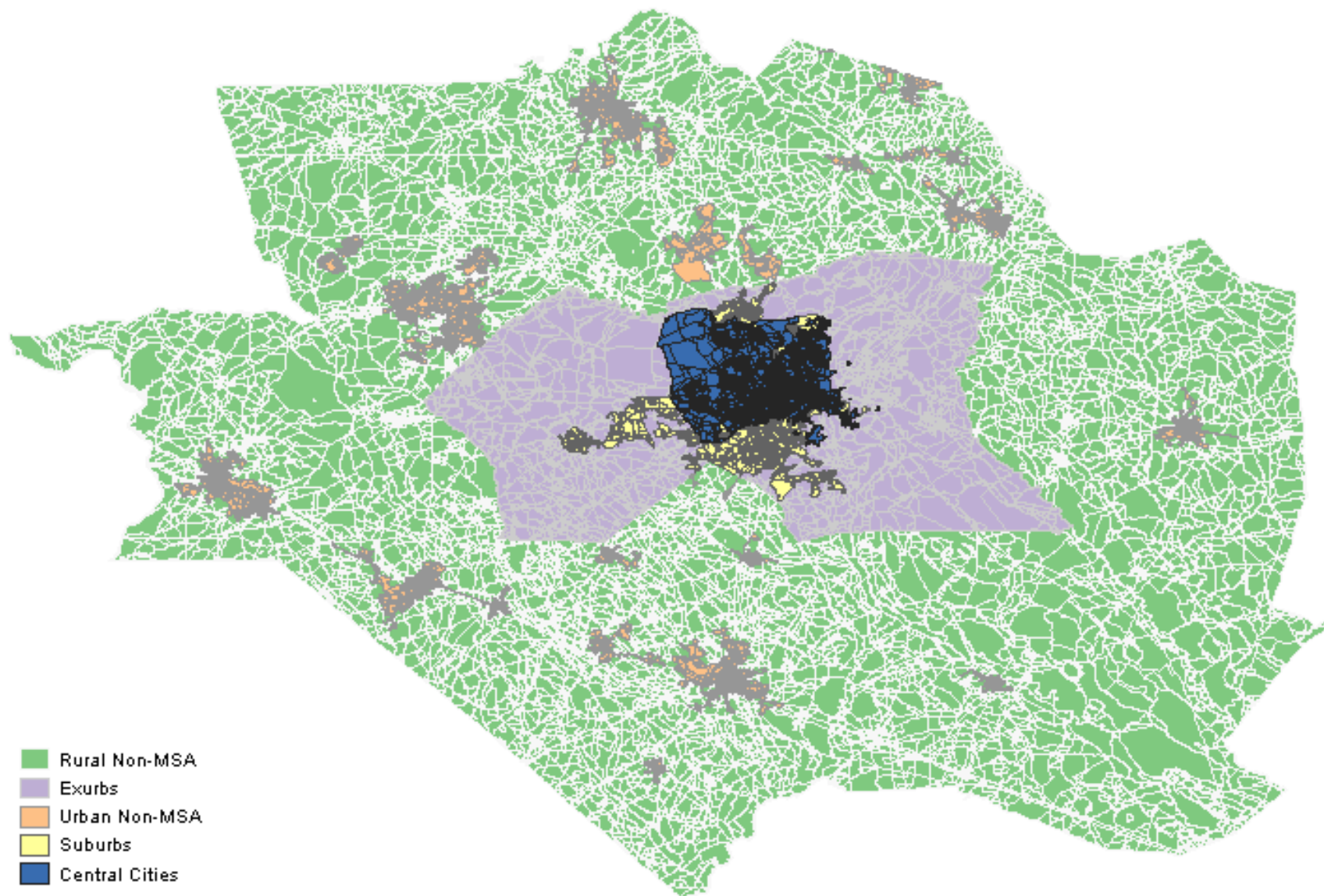
Urban MSA Not Central City



Wireless Download Speeds



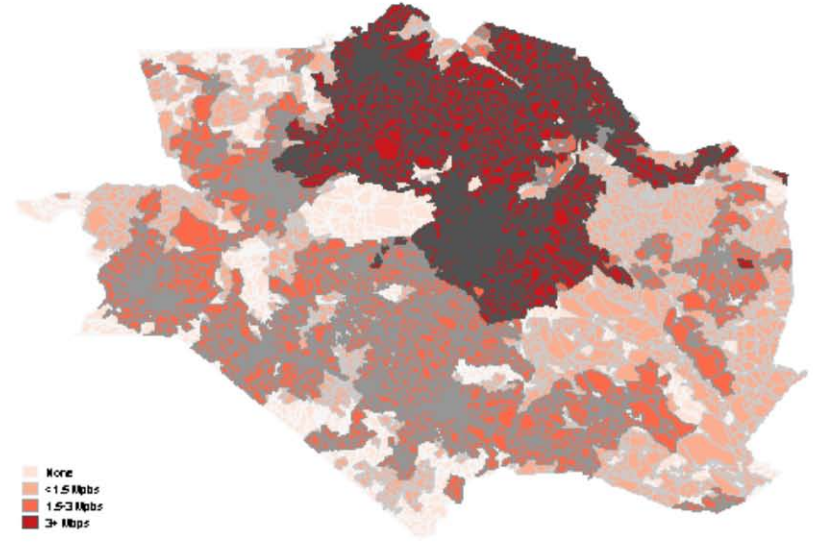
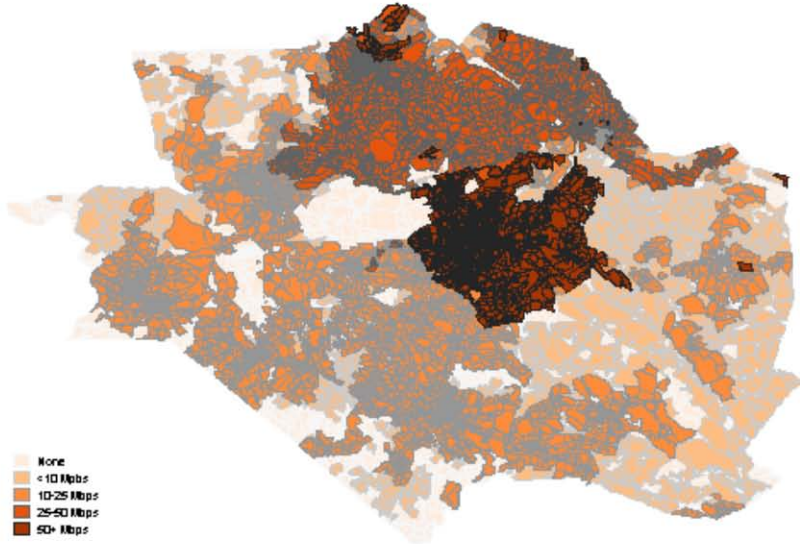




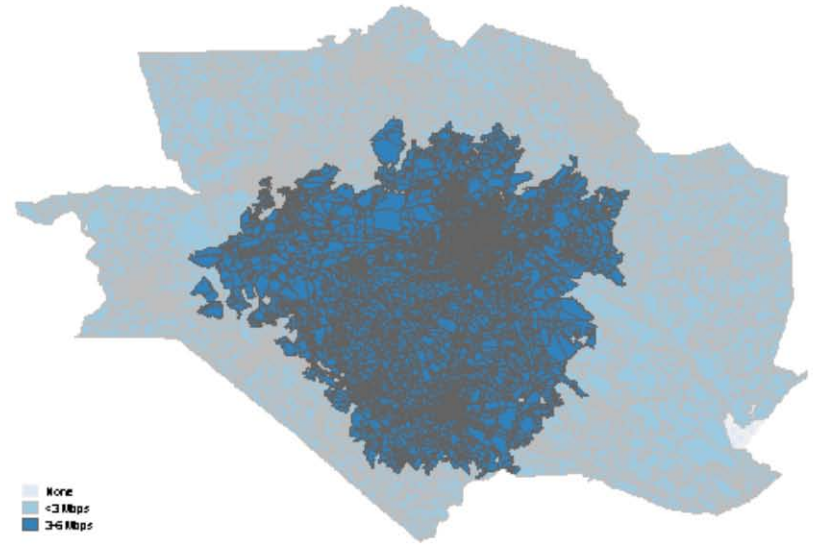
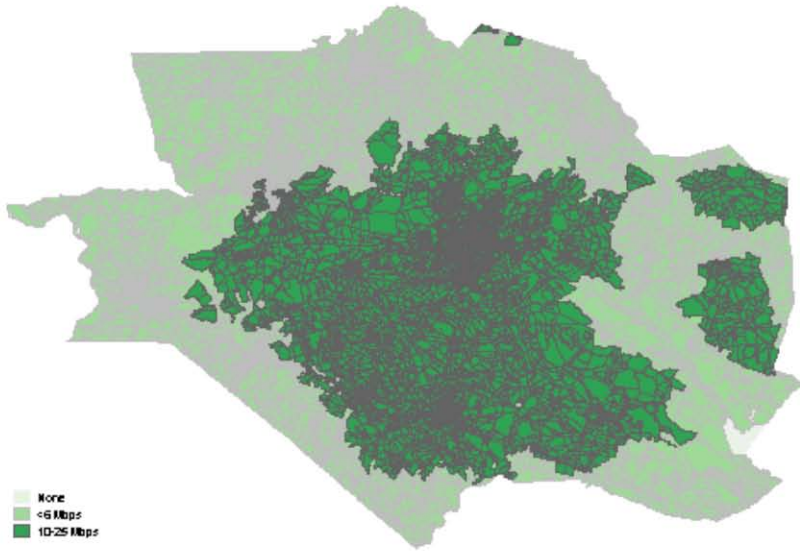
Download Speeds

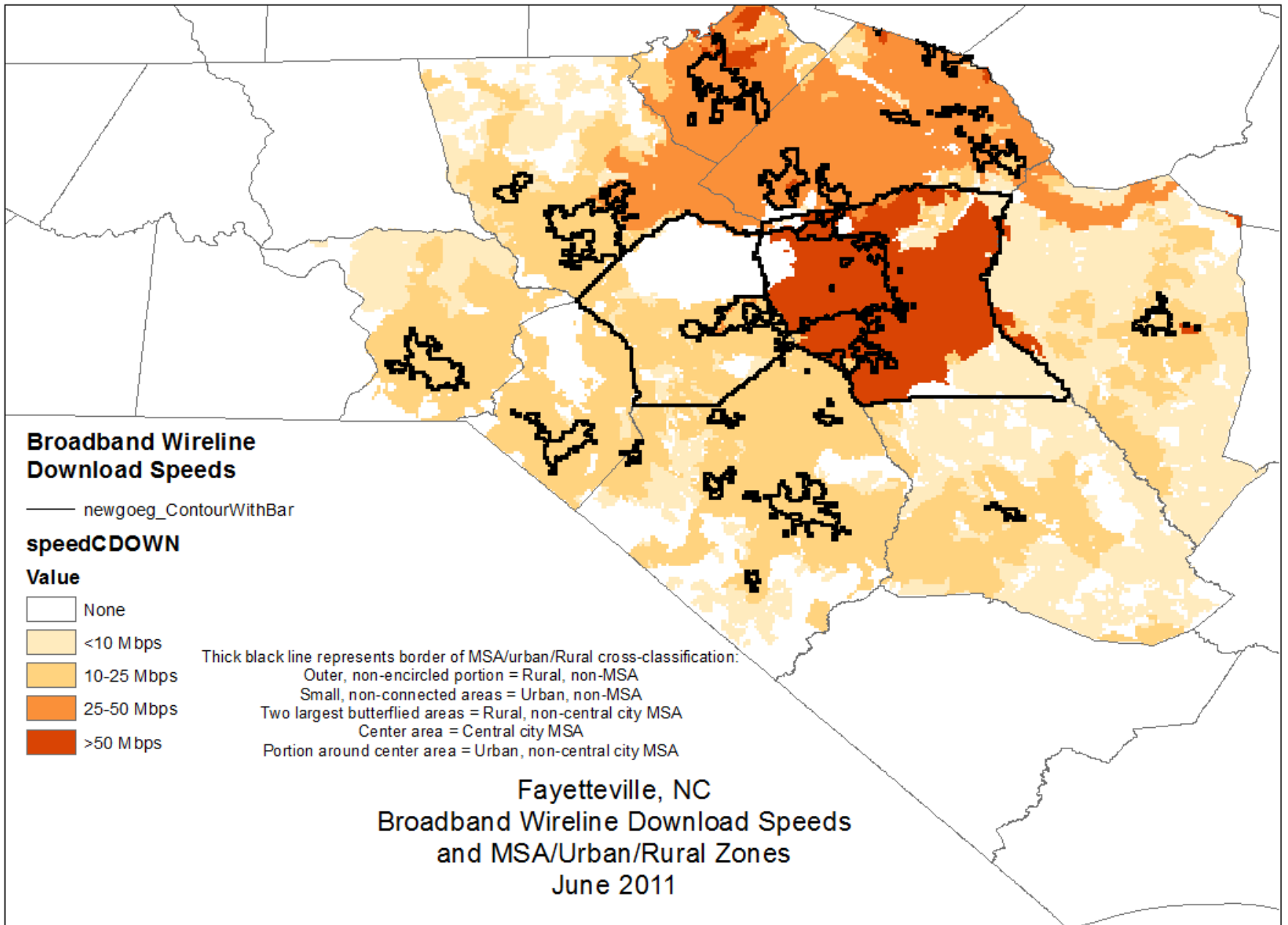
Upload Speeds

Wireline Service



Wireless Service





Extremely Preliminary and Exploratory Regressions

- Focus on wireless download speeds (only two speed ranges – high and low) .
- Estimate linear probability model.
- But “model” is too strong a word at this point – these regressions clearly confound supply (especially population density) and demand factors.
- Still, they may help reveal some interesting features to help inform future work.

**Dependent Variable = 1 if have wireless download speed 10-25 Mbps;
= 0 if < 6 Mbps**

Black non-Hispanic	0.134*	0.002	-0.002
Asian non-Hispanic	0.763*	-0.004	-0.011
Other non-Hispanic	0.433*	0.047*	0.042*
Hispanic	0.068*	0.057*	0.048*
Log population density			0.001
Log job density			0.002*
Female			0.001
Age 18-44			0.003
Age 45-64			-0.005
Age 65+			-0.021*
High School			-0.131*
Some College			0.120*
College			-0.104*
Graduate School			-0.342*
Income \$15-35K			-0.059*
Income \$35-60K			0.008
Income \$60-100K			-0.024
Income \$100K +			0.150*
Constant	0.441*	0.500*	0.534*
Fixed (Census Tract) Effects?	No	Yes	Yes
R-squared	0.038	0.779	0.780
N	30,660	30,660	30,660

Future Work

- Merge NBM data on broadband availability with Current Population Survey data on broadband adoption (supply and demand!).
- Use terrain as way to identify supply effects.
- Use other methods to deal with spatial correlation.
- Deal with margins of error in ACS variables.