Who Gets Broadband When? A Panel Data Analysis of Demographic, Economic, and Technological Factors Explaining U.S. Broadband Deployment

Vamsi Gadiraju, Anthony Panat, Raghav Poddar, Zain Sherriff, Sam Kececi, and Henning Schulzrinne

Columbia University

September 22, 2018

Abstract

We study how US broadband has been deployed from 2014 to 2016 at the block and block-group level, focusing on access technology development and upgrades, expansion in rural and non-rural areas, demographics, and ISP growth. Since broadband definitions have changed over time, we provide statistical analyses using both the 25 Mbps download and 3 Mbps upload speeds as well as 10 Mbps download and 1 Mbps upload speeds. We combine data from Federal Communications Commission (FCC) 477 forms, the 2010 Census, and 2010-2014 American Community Survey (ACS) data to analyze the technology, performance, provider, and service territory characteristics affecting broadband offerings.

Starting with an infrastructure-centric analysis, we analyze the most common technologies in new broadband deployments and then proceed to determine the speed distributions for each. We then focus on the rural/non-rural divide in expansion and investigate whether broadband is becoming more readily available in regions with lower economic growth factors and lower educational rates. As expected, from 2014-2016, urban blocks saw the largest increase in average speeds relative to rural ones.

Aside from this infrastructure-focused analysis, we used a binary logistic regression to analyze the regression coefficients and general statistics between broadband availability and demographic indicators of economic status, population change, and education. Finally, we explore how provider coverage across the nation has changed, finding large increases in the availability of 25/3 Mbps broadband in areas that already had lower-speed broadband.

1 Introduction

92.3% of Americans now have access to fixed broadband under the FCC's current standard, which, since 2015, has changed from 10 Mbps download speed and 1 Mbps upload speed (which we label as "10/1")[1] to 25 Mbps download and 3 Mbps ("25/3" in this paper). However, as the FCC acknowledged in its 2018 Broadband Progress Report, 24 million Americans, overwhelmingly from rural areas, lack fixed broadband at the 25/3+ standard [1]. In addition, a significant percentage of the rural population lacks access to broadband at low speeds; a 2016 FCC report found that 20% of rural Americans lack access even to 4/1 broadband [2]. While FCC and other reports have pointed out key variables, such as population density, topography and household incomes, we attempt a more in-depth exploration, and also point out some less-explored facets of broadband deployment.

Although our paper's focus is not on the economic impact of broadband, it must be taken into account. On a nationwide scale, the broadband and information and communications technologies (ICT) industry contributes $1,209.2 billion in value, or 6.5% of total U.S. GDP [3]. With respect to job creation, a 2016 study on U.S. broadband outlined the establishment of 6.4 percent of all U.S. private employment as the

[1] See Section 2.

result of the broadband/ICT industry [4]. In addition, individual jobs in the digital sector earned around 60% more than the average worker in the U.S. economy [4].

This paper analyzes the factors affecting broadband deployment in the United States, addressing both the questions of who gets broadband and where. As the actual decision processes of Internet Service Providers (ISPs) are not visible to us, we focus on demographic and population statistics, including housing construction and valuation, in order to analyze the likely explanatory effects of socioeconomic variables on broadband deployment. We use both regression and machine-learning techniques to predict expansion of broadband. This approach was motivated by the observation by the FCC that access to higher broadband speeds is far more prevalent in areas with “higher average populations, population densities, per capita incomes, and median household incomes” [1].

To supplement our analysis of socioeconomic factors, we also aim to better understand where the major ISPs deployed new offerings in the 2014 to 2016 time period. By investigating the respective expansions from 2014-2016 and provided speeds of large ISPs, we sought to identify both the type of broadband provided and year-to-year expansion, both in aggregrate and individually for several large ISPs.

To analyze broadband deployment over time, we rely on the public FCC Form 477 data for 2014 through 2016. We are well aware of its limitations, but lack other data sources with similar granularity and multi-year coverage. From the data set, we cannot tell whether an ISP serves one household in a census block at the speed indicated, every household or some fraction in-between. Thus, since partial coverage of blocks is particularly likely in rural areas, the data provides an upper bound on broadband availability. Other data oddities, such as temporary loss of broadband and extremely high speeds for certain techologies, also indicate that the Form 477 data should be analyzed with caution. The actual Form 477 data contains additional data point, such as the number of subscribers to a service tier, that would allow a more nuanced geographic analysis, but this data is not made available to the public. It should also be noted that our analysis ends in December 2016, as newer data has not been published as of August 2018. Unfortunately, data prior to 2014 is not directly comparable due to changes in the Form 477 data gathering process.

The remainder of the paper is organized as follows. Section 2 defines broadband terms and related key concepts which we analyze in depth later. In Section 3, we introduce our findings regarding demographic characteristics and the division between rural and urban census blocks. Section 4 presents our findings describing changes in broadband deployment on a year to year basis. We take a look at corner cases, namely zero population blocks with broadband and areas that temporarily seem to lose, and then re-gain, broadband. We then present our analysis on demographic indicators of broadband deployment in Section 5. In Section 6, we focus on service providers and specifically the broadband expansion of the largest telephone and cable companies in the United States lower-48 market. Finally, we conclude our findings and discuss future work in Section 7.

2 Terminology

We define some key terms below. In general, when discussing broadband, we use the notation “D/U+” to refer to download and upload speed combinations that exceed a download speed of D and an upload speed of U, while “D/U” refers to a specific speed tier. For example, “25/3+” refers to service that offers a download speed of at least 25 Mb/s and an upload speed of at least 3 Mb/s. On occasion, we use the expression “10/1-” to refer to areas that only have 10/1 service or less.

Broadband: The FCC has defined broadband by different minimum download and upload speed tiers, increasing the threshold over time and using different threshold for different purposes, such as broadband availability reporting and eligibility for Universal Service Funds. The FCC defined broadband as 4/1+ in 2010 and updated the definition to 25/3+ in 2015 [5]. The threshold of 10/1 serves as a minimum service requirement in the Connect America Fund (CAF) II [6].

Census block: Census blocks are “[s]tatistical areas bounded by visible features such as roads, streams, and railroad tracks, and by non-visible boundaries such as property lines, city, township, school district, county limits and short line-of-sight extensions of roads. . . . Generally small in area. In a city, a census block looks like a city block bounded on all sides by streets. Census blocks in suburban and rural areas may be large, irregular, and bounded by a variety of features, such as roads, streams, and transmission lines. In remote areas, census blocks may encompass hundreds of square miles. ... The smallest level of geography you can get basic demographic data for, such as total population by age, sex, and race" [7]. Census blocks are delineated once every ten years. In 2010, the United States, excluding territories and possessions, had 11,078,297 census blocks, of which 541,776 are water-only [2]. Census blocks may have zero population, which we discuss in more detail below.

Census block group: "Block Groups (BGs) are statistical divisions of census tracts, are generally defined to contain between 600 and 3,000 people, and are used to present data and control block numbering. A block group consists of clusters of blocks within the same census tract that have the same first digit of their four-digit census block number." Block groups typically cover a contiguous area [8].

Census tract: "Census Tracts are small, relatively permanent statistical subdivisions of a county or equivalent entity that are updated by local participants prior to each decennial census as part of the Census Bureau's Participant Statistical Areas Program. The Census Bureau delineates census tracts in situations where no local participant existed or where state, local, or tribal governments declined to participate. The primary purpose of census tracts is to provide a stable set of geographic units for the presentation of statistical data.

Census tracts generally have a population size between 1,200 and 8,000 people, with an optimum size of 4,000 people. A census tract usually covers a contiguous area; however, the spatial size of census tracts varies widely depending on the density of settlement. Census tract boundaries are delineated with the intention of being maintained over a long time so that statistical comparisons can be made from census to census. Census tracts occasionally are split due to population growth or merged as a result of substantial population decline" [9].

Rural census block or group: We divide census blocks into rural and non-rural by population density, with an urban block having a population density of more than 1,000 people per square mile. This definition was adapted from previous Census definitions and fits the block scope of this study well [10].

American Community Survey (ACS): The ACS is a survey prepared by the U.S. Census Bureau which uses "a series of monthly samples to produce annually updated estimates for the same small areas (census tracts and block groups) formerly surveyed via the decennial census long-form sample" [11]. In addition to offering 1-year ACS estimates, the Census Bureau offers 5-year ACS estimates which are available for areas at the block group level or larger.

(FCC) Form 477: A Form 477 provides information on broadband deployment data as it is the product of the FCC's directive requiring all facilities-based providers "to file data with the FCC twice a year (Form 477) on where they offer Internet access service at speeds exceeding 200 Kbps in at least one direction" [12]. With respect to fixed broadband, Form 477 data is offered at the block level and, specifically, for blocks where providers "can or do offer service to at least one location" [12].

2.1 Broadband Transmission Technology Types

Our analysis of broadband transmission technology types focused on fixed wired technologies, specifically Digital Subscriber Line (DSL), cable (hybrid-fiber coax or HFC), and fiber to the home (FTTH), along with terrestrial fixed wireless. DSL sends data over copper telephone lines and its speed is dependent upon the distance to a service provider facility, typically limited to 25 Mbps. Cable broadband uses the DOCSIS 1.0 through DOCSIS 3.1 standards and can achieve rates up to 1 Gbps. FTTH provides data transmission at speeds up to 1 Gbps today, but speeds of 10 Gbps are achievable. Fixed wireless is found mostly in rural areas, using licensed or unlicensed spectrum to reach subscribers.

Table 1 summarizes the maximum speed for each technology code found in the December 2016 Form 477 data.

[2]https://www.census.gov/geo/maps-data/data/tallies/tractblock.html

Technology Type	Highest Speed after Removing Top 5% of Speeds (Mb/s)	Highest Speed after Removing Top 1% of Speeds (Mb/s)
DSL	30	50
Terrestrial fixed wireless	70	102
Cable DOCSIS 1.0 - 2.0	100	200
Cable DOCSIS 3.0 - 3.1	600	1000
Fiber (FTTH)	1000	1000

Table 1: 2016 maximum download speed by technology type.

The highest download speed for each technology type are found in the data are shown above. We removed outliers by removing both the top 1% and top 5% of speeds for each technology type, since this gave a more realistic representation of the threshold speeds for these technologies, and these upper values are not representative of the general speeds.

Table 2 provides more details on the download speeds by showing the distribution of speeds for each of the technology types.

Technology Type	Mean Speed	25^{th} Percentile Speed	50^{th} Percentile Speed	75^{th} Percentile Speed
DSL	14.8	6.0	12.0	24.0
Terrestrial fixed wireless	28.3	8.0	15.0	30.0
Cable DOCSIS 1.0 - 2.0	45.3	30.0	55.0	90.0
Cable DOCSIS 3.0 - 3.1	535.4	200.0	400.0	500.0
Fiber (FTTH)	658.8	500.0	850.0	1000.0

Table 2: 2016 speed distributions by technology type.

This distribution of speeds shows that DSL in general has the slowest speeds of the main technology types, while fiber consistently provides the fastest speeds. For cable, DOCSIS versions 1.0-2.0 are much slower than versions 3.0-3.1.

3 The Current State of Broadband

The goal of this study is to provide additional analysis into both the nature of existing broadband offerings and detect patterns in recent deployments. As such, we will start off by analyzing the general state of deployment across the country, focusing on replicating some of the common metrics from the FCC Yearly Broadband Reports and providing additional demographic-based ones. Given the growing interest in the rural/urban divide in deployment, we will also focus on the divide between broadband deployment in rural and urban areas, as well as what factors and challenges are most prevalent in rural deployment.

3.1 Methodology

To establish general trends, we initially focused on defining broadband itself. Given recent debate over the definition of broadband, this section will utilize both the 10/1+ and 25/3+ definitions. To this end, we utilized fixed broadband deployment information from the FCC's Form 477 data. In particular, we used the data from the second yearly version of Form 477 data releases from 2014-2016. Within the data set, we focused on analyzing trends within the contiguous USA (excluding US territories) and excluded data on satellite coverage. These omissions were made given the fact that deployment costs and strategies have been noted to be extremely different in non-contiguous region, thus possibly skewing the analysis.

To complement this provider offering data with population data and demographic information, we used the FCC's Commission Staff 2015 population estimates at a block level and the Census Bureau's Planning

Database (PDB) for 2014-2016 at a block group level. At the block group level, we focused on the following attributes from the PDB that were derived from ACS surveys: population distributions (percentages of population in age ranges of 0-5, 5-17, 18-24, 25-44, 45-64, 65+), education distributions (percentage of population that completed high school or higher degrees), median household incomes, median home valuations, recent home construction percentages (percentage of homes built after 2010), and population movement rates (percentage of households where the householder moved in after 2010). While the first four metrics directly correspond to population density, education, income, and age, we used the median home valuation (which we used to calculate the home valuation increase percentage), recent home construction percentage, and moved population percentage to assess economic growth and development in a region. Studies have indicated that these metrics are strongly correlated to both robust economic growth and population growth of a region [13], [14]. We had to account for varying aggregation levels of the data in that the Form 477 data operates at a block level and the PDB data is on the block group level. As such, in our analyses, we aggregated demographic data at the block group level and population counts on the block level given the availability of data sources. Comparisons were made across our computed statistics and select statistics from FCC yearly broadband reports to establish the validity of our analysis methods [1], [2].

3.2 Key Broadband Indicators

Table 3 highlights the broadband availability across America for urban, rural, and tribal areas for 25/3+ broadband deployments. Our analysis was compared to data from the FCC's 2018 Broadband Deployment Report and its latest Internet Access Services Report [1], [15].

Availability	**Analysis Values (%)**	**FCC report data (%) [1]**	**FCC reported adoption rates (%) [15]**
Total United States	93.4	92.3	53.3
Urban Blocks	98.3	97.9	
Rural Blocks	70.6	69.3	
Tribal Regions	66.3	64.6	32.6

Table 3: Percentage of population with broadband availability in 2016 data for 25/3+ deployments

Table 4 encapsulates similar ideas to Table 2 for 10/1+ speed broadband deployments.

Availability	**Analysis Values (%)**	**FCC Report Data (%) [1]**	**FCC Reported Adoption Rates (%)[15]**
Total United States	96.8	96.0	66.2
Urban Blocks	99.1	98.9	
Rural Blocks	84.2	83.9	
Tribal Regions	82.1	81.8	42.2

Table 4: Percentage of population with broadband availability in 2016 data for 10/1+ deployments

Table 5 illustrates access rates at various median household income quartiles for 2016. This table provides us with an easier method to understand the economic profile of regions covered by existing broadband deployments. We compared our values to those aggregated at the county level in the FCC 2018 Broadband Deployment Report [15].

Income Quartile	Analysis Values at block group level	FCC Report Data at County Level [15]
First (0-25%)	52.6	58.2
Second	73.2	69.7
Third	80.5	76.2
Fourth (75-100%)	88.0	84.1

Table 5: Average percentage of population with 25/3+ offering access by median household income quartile in 2016

Table 6 shows general demographic factors pertaining to block groups that did or did not have broadband deployments in 2016. For all key demographic factors listed below, we computed the mean and median values for all block groups that had at least one 25/3+ service offering and for those that had no broadband offerings. It should be noted that for the median household income and the median home value of each block group, we computed the mean and median across all relevant block groups to provide economic insight into the regions with and without broadband. Additionally, the urban population percentage (from the ACS) seems to be a rough estimate at the block group level, with many of the values being 0%, 50%, or 100%.

	25/3+ Mean	25/3+ Median	Mean w/o 25/3+	Median w/o 25/3+
Urban population (%)	69.72	100.00	1.73	0.00
Ages below 5 (%)	6.05	5.54	5.78	5.28
Ages 5-17 (%)	16.45	16.44	16.74	16.67
Ages 18-24 (%)	9.16	7.94	7.82	7.12
Ages 25-44 (%)	25.72	25.03	21.95	21.65
Ages 45-64 (%)	27.50	27.34	29.84	29.52
Ages Above 65 (%)	15.11	13.53	17.86	17.01
No HS degree (%)	13.97	10.55	18.82	16.18
College degree (%)	28.17	22.92	14.48	12.93
Median household income	$ 59,456	$ 52,120	$ 42,493	$ 40,580
Median home value	$ 219,247	$ 160,200	$ 115,856	$ 94,300
Population movement (%)	4.87	2.90	13.97	2.90

Table 6: General demographic factors for block groups with and without broadband in 2016.

3.3 Analysis

These general statistics are particularly relevant for two main reasons: (1) they indicate the trends associated with broadband growth and (2) they validate our method of data analysis. Many of the key metrics here are to be expected: people who are generally wealthier, employed, and educated tend to have the most access to broadband. From a high level, the goal of analyzing these general statistics was to establish a frame of reference for year-to-year changes (next section) and to highlight that our methodology yields results that are extremely similar to values from the FCC, who largely deal with the collection of the data we used.

Based on our analysis, by the end of 2016, 93.4% of Americans lived in a block where there was at least one 25/3+ broadband service offering, marking an improvement from 89.3% access rate at the beginning of 2014. While these numbers may seem high, it is important to note two key factors: 1) by our analysis, approximately 24 million Americans still do not have access to a broadband service offering and 2) availability rates and adoption rates continue to be extremely different (there is nearly a 40% difference between the number of people who have access to broadband and the number of subscribers for 25/3+ speeds across America) [1], [15]. Based on the FCC's Internet Access Report for 2016, it is interesting to note that both the adoption rate and availability rate increased by roughly 3% (to 106 million fixed connections) between December 2015 and December 2016 [15].

Looking at Tables 3 and 4 above, all of our analyzed metrics on broadband availability are within 5% of their corresponding variables from the FCC yearly reports. It should be acknowledged that the discrepancies are slightly larger (still below 5%) for the rural/urban divide metrics. We postulate that this larger deviation may be the result of differing approaches to defining rural/urban blocks. As described in the methodology and terminology section, we included an analysis based on the FCC's 2015 estimates of block-level populations and population density. On the other hand, the FCC yearly reports are based upon the listing of rural/urban blocks from the 2010 Census. Our definition of rural and urban blocks feature the most recent population changes.

With this focus of this study on the deployment characteristics of broadband, it was particularly interesting to consider the nature of the rural/urban divide. Interestingly, almost half of the census blocks that received their first Internet service offerings for 25/3+ broadband from 2014-2016 were rural (based on the Census definition). Given that currently only four percent of urban Americans lack access to broadband with sufficient speeds, it makes sense that many of the new blocks to receive faster Internet access would be in rural or less urbanized regions [1]. More generally, as time passes, broadband deployments are becoming less demographically restrictive (lower education rates, household incomes, etc are covered in newly deployed regions). These findings are elaborated upon in Section 5. Looking at Table 4, we can see that based on the current state of distributions of median household income quartiles, there is less opportunity for growth in higher quartile block groups, meaning that newer deployments may occur in relatively poorer areas.

The wide gap in demographic factors between block groups with and without 25/3+ broadband service offerings can be seen in greater detail in Table 6. Starting with the ACS urban population percentage, there is a clear discrepancy (especially when looking at the median values). This difference re-iterates the fact that most of the regions without broadband are overwhelmingly rural or tribal in nature. In considering education rates, as expected, the percentage of people with high school degrees and the percentage of people with college degrees are both higher in regions that have broadband (differences of about 6% and 10% for the medians, respectively). A similar trend can be seen in median home valuations and median household income, where the differences in the median values are 22% and 41%, respectively. It is interesting to note that the difference in age profiles is less pronounced between the two types of block groups. Additionally, it was surprising to note that the inbound population movement rate percentages (percentage of housing units where householder moved in after 2010) were actually higher in regions without broadband.

4 Year over Year Changes in Broadband Deployment

4.1 Some Areas Seem to Lose Broadband, But Some Only for a Year

We generally expect broadband availability to increase year-over-year — once an ISP starts serving a census block, it would have little incentive to stop doing so. Providers get acquired or sell parts of their service territory, but this would yield only a change in provider, not a decrease in broadband availability or speed. However, as we show below, the Form 477 data shows a non-trivial number of census blocks that lose broadband in the reported data. We also found thousands of blocks where broadband disappeared and then re-appeared across a two-year time span.

Since providers do not explain their data reporting, we can only guess at the reasons for these unexpected changes. Reasons could include reporting errors or an ISP realizes that the actual speeds delivered fall below the 10/1+ or 25/3+ threshold and thus may make an area eligible for USF funding.

In our analysis below, we focus on large holding companies only.

4.1.1 What Constitutes Lost Broadband?

We define lost broadband as a block where a specific holding company provided 25/3+ broadband in 2014, but the same holding company provided only 10/1- broadband in 2015, based on the December 2014, 2015 and 2016 Form 477 data. We focused on the ten holding companies that lost broadband in the largest number of blocks.

To establish whether reporting errors might play a role, we also tallied temporary loss of broadband, i.e., where 25/3+ broadband was available in 2014, only 10/1- in 2015, and then 25/3+ again in 2016. Other than reporting errors, it is also possible that the reporting in 2015 reflected the actual situation, with upgrades in 2016. Unfortunately, there does not appear to be any public information that would allow to ascertain the precise cause.

Holding Company	**Blocks Losing Broadband 2014 to 2015**
Charter Communications	871
Bright House Networks	649
Windstream Corporation	471
Comcast Corporation	465
Cequel Communications	351
Etheric Networks	267
Shenandoah Telecommunications Company	265
Siouxland Wireless	245
Cable One	221
Time Warner Cable	218

Table 7: Number of blocks that lost broadband from 2014 to 2015 and the block was later aquired by a different provider in 2016

New Holding Company	**Blocks Gaining Broadband in 2016**
JAB Wireless	1,968
Charter Communications	720
Neptune Holding US Corp.	583
Comcast Corporation	568
AT&T	505
LTD Broadband	432
Micrologic	362
The Computer Works	285
Liberty Global	102
The Junction Internet	99

Table 8: Number of blocks that gained broadband in 2016 from a different provider than the one in 2014

Holding Company	**Blocks Losing Broadband 2014 to 2015**	**Blocks Losing Broadband in 2016 After Removing Glitch Blocks**
Windstream Corporation	21,956	8,754
Charter Communications	8,910	6,254
Comcast Corporation	4,199	2,400
Mediacom Communications Corp.	3,731	17
Time Warner Cable	3,138	2,520
Verizon Communications	2,176	19
WEHCO Video	1,607	1,151
Golden West Telecom.	1,393	652
Etheric Networks	1,298	948
Cable One	1,197	560

Table 9: Number of blocks losing broadband in 2016 before and after removing temporary loss for each holding company

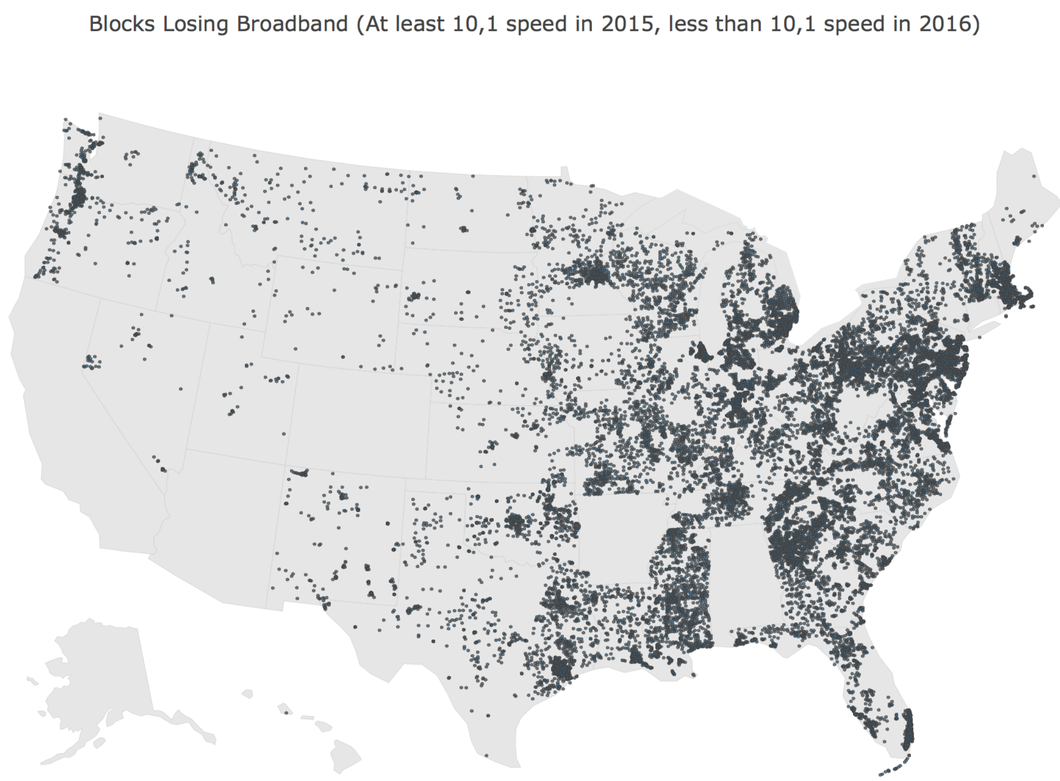

Figure 1: Census Blocks Losing Broadband

4.1.2 Analysis

Our initial analysis showed that there were 67,185 distinct blocks that fit our definition of lost broadband. Of these 67,185 blocks, there were 20,535 distinct blocks that regained broadband in 2016 and 7,195 of those blocks were due to a different provider picking up that area in 2016. From Table 7, we see that Charter and Bright House Networks LLC were two of the holding companies that lost the largest number of blocks to a different holding company in 2016. However, the small number of blocks and even distribution across providers and technology types point to data quality issues. Table 8 shows us that JAB and Charter were able to serve broadband to the most number of blocks in 2016 that had different providers in the loss of broadband from 2014 to 2015. This replacement of providers in blocks that had lost broadband previously potentially shows a buyout of a certain area by these companies.

Next, we took a look at the absolute number of blocks where broadband was lost from 2014 to 2015 and checked for temporary losses of broadband. In Table 9, we see that Verizon and Mediacom lost almost no blocks between 2014 and 2016. As the Federal Register noted in its publication on modernizing the FCC Form 477 data, "[Q]uestions have arisen in various contexts regarding the bases for certain filings by service providers and the extent to which those filings reflect actual user experience. The Commission to date has not systematically examined the precise underlying methodologies that are used by service providers in generating their data nor has it investigated whether actual consumer experience has diverged substantially from the Form 477 filings" [16].

4.2 True Zero Blocks

As we began to investigate broadband deployments, we noticed an interesting fact: there were service offerings in blocks with a recorded Census population of 0. Given the definition of population in the Census, we surmised that this discrepancy was probably the result of either blocks with high seasonal populations or blocks that are close to other blocks with broadband offerings. Blocks with high seasonal populations would have little to no population and recorded in the Census and would still have broadband deployments given that many of the property owners would have the ability and willingness to pay for it. Given the nature of the Form 477 data and the fact that companies list the regions that they could cover with their infrastructure, zero population blocks near other broadband blocks with 25/3+ service offerings may also be listed despite a lack of subscribing customers.

4.2.1 Methodology

This part of the paper, in particular, used the data from the second edition of FCC Form 477 data releases from 2016. To complement the Form 477 data with population data, we used the FCC's Commission Staff 2016 population estimates at a block level. We first detected all blocks with zero populations listed and then proceeded to evaluate whether they had broadband or not. Once these blocks had been identified, block group-level data on demographics were used from the Census Bureau's Planning Database on seasonal home rates, population change, and median home valuation. We filtered out all zero population blocks with broadband that were geographically next to other blocks with 25/3+ offerings and then proceeded to analyze the demographics of the remaining blocks.

4.2.2 Key Figures

Figure 2 shows all the blocks across the country that are listed as having zero population and at least one 25/3+ broadband service offering. It should be noted that the dot sizes in the visualization are considerably larger than the actual size of the block for visualization purposes.

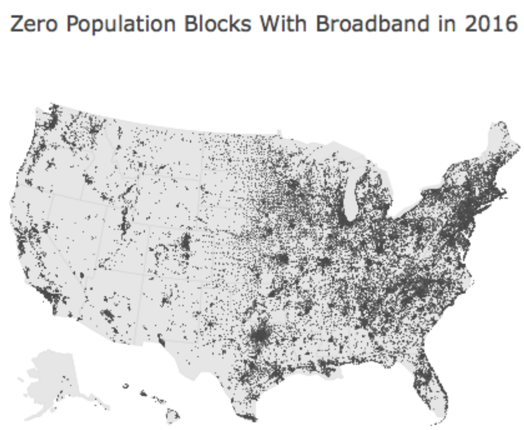

Figure 2: Map of zero population blocks with broadband

4.2.3 Analysis

From our evaluation of the blocks with zero population that were not adjacent to other blocks with broadband offerings, we noticed that there were high seasonal home percentages and low population change at the higher block group level. First, 61.4% of these blocks were in block groups with over 50% seasonal home percentages, indicating a trend of moving seasonal and temporal populations. At face value, this value appears to indicate that these regions may be vacation areas or popular for temporary housing. The wealthier nature of individuals who can afford vacation homes ties in with our earlier argument on the strong correlation between wealth/propensity to spend and broadband coverage. We further evaluated these regions to better understand why they may have broadband, focusing on two main questions: was there a large, rapid population movements that may have caused this zero population and are these generally unpopular regions to live in? To answer the first question, we evaluated the block level data for 2000 and found that fewer than 5% of these blocks had a recorded population previously. To look at the changing popularity of these regions, we evaluated the changes in population at the block group level (ACS and Census data at block group level from Census PDB). We found that below 10% of these zero population blocks were in a block group that faced a population change of over 5%. Generally speaking, this indicated that the broadband deployment was not the result of population movements and furthers our argument that these may be vacation or temporal home regions that have a high seasonal population.

5 Understanding Demographic Indicators

One of the key purposes of this study is to analyze the relationship between key demographic and economic variables and new broadband deployments based on year-to-year changes in the Form 477 data. Although there is a growing amount of literature on broadband deployment, there has not been a comprehensive analysis of the yearly changes that may be useful to understand current broadband deployment and its demographic patterns. Given the rapid changes in corporate interests and strategy and government funding for deployment project, we wanted to analyze how regression coefficients and general statistics relating to demographics were changing over time. We back-tested our regression model across the years from 2014-2016 to provide a better understanding of the importance of our factors in understanding deployment tendencies.

5.1 Methodology

To develop the regression coefficients across our data, we used two main data sources: the FCC Form 477 data on broadband offerings and the Census's Planning Database data on demographic and economic metrics. Additionally, for the correlation section, we used the 25/3 definition for broadband. We adopted the FCC's Commission Staff 2015 population estimates at a block level and the Census Bureau's Planning Database (PDB) for 2014-2016 at a block group level. At the block group level, we focused on the following attributes from the PDB that were derived from ACS surveys: population distributions (percentages of population in age ranges), education distributions (percentage of population that completed high school or college), median household income, median home valuation, recent home construction percentage (percentage of homes built after 2010), and mover rates (percentage of households where the householder moved in after 2010). While the first four metrics directly correspond to population density, education, income, and age, we used the median home valuation (which we used to calculate the home valuation increase percentage), recent home construction percentage, and population movement rates to assess economic growth and development in a region. As in the previous sections, we applied the demographic data at the block group level on the population of a specific block.

To obtain the regression coefficients, we ran a binary logistic regression using the popular statsmodel API. We also sought to attempt classification of sample blocks, so we applied the Nearest Neighbor algorithm (KNN) to determine how predictive demographic and economic similarity to blocks with recent deployments is. The algorithm was trained on all blocks that received broadband deployments for the first time between 2014-2015. We tested our model against all new deployments between 2015-2016. In running the KNN, we also added three additional unique binary values that indicate whether the block was in the same block group as a recent deployment (last two years), whether a major carrier (top 5 carriers) operated in the block group, and whether the block group population increased in the last two years.

5.2 Key Demographic Indicators

Tables 10 and 11 show the mean and median values for key demographic indicators of economic prosperity, education, and urbanity for block groups with new 25/3+ broadband deployment in 2015 and 2016, respectively.

	Mean at Block Group Level	Median at Block Group Level
Urban population (%)	45.5	86.3
Population 0-5 (%)	6.0	5.7
Population 5-17 (%)	16.5	16.7
Population 18-24 (%)	9.1	7.9
Population 25-44 (%)	24.5	23.8
Population 45-54 (%)	27.9	28.3
No HS diploma (%)	14.0	11.88
College degree (%)	24.4	19.6
Median income	53,914	49,002
Median home valuation	175,266	138,800
Population movement (%)	22.6	2.0

Table 10: Key demographic statistics for new deployments of 25/3+ speeds between 2015-2016.

	Mean at Block Group Level	Median at Block Group Level
Urban population (%)	60.8	98.2
Population 0-5 (%)	6.1	5.8
Population 5-17 (%)	16.6	16.8
Population 18-24 (%)	9.4	8.1
Population 25-44 (%)	25.2	24.5
Population 45-54 (%)	27.4	27.8
No HS diploma (%)	13.5	11.1
College degree (%)	27.2	22.0
Median household income	57,066	51,012
Median home valuation	194,378	152,600
Population movement (%)	13.8	2.1

Table 11: Key demographic statistics for new deployments of 25/3+ speeds between 2014-2015.

Tables 12, 13, and 14 show the binary logistic regression coefficients for key demographic indicators in block groups that saw new broadband deployments in 2015, 2016, and between 2014-2016, respectively. The tables also include 90% confidence interval percentages and p-values for each demographic factor. It should be noted that these values were all computed for a single factor coefficient.

	Regression Coefficient	**90% CI Percentage**	**p Value**
Percentage with no HS diploma	-1.1385	58%	0.05
Percentage with College degree	0.9040	63%	0.04
Percentage of recent home construction	0.8246	29%	0.04
Median Household Income	1.2754	21%	0.01
Median Home Value	1.0710	46%	0.02
Population movement rate	0.0057	15%	0.00

Table 12: Logistic regression coefficients for new 25/3+ deployments in 2015.

	Regression Coefficient	90% CI Percentage	p Value
Percentage with no HS diploma	-1.2074	64%	0.04
Percentage with College degree	0.8897	53%	0.05
Percentage of recent home construction	0.7982	27%	0.04
Median Household Income	1.2693	19%	0.04
Median Home Value	1.0550	31%	0.03
Population movement rate	0.0108	26%	0.02

Table 13: Logistic regression coefficients for new 25/3+ deployments in 2016.

	Regression Coefficient	90% CI Percentage	p Value
Percentage with no HS diploma	-1.5482	72%	0.04
Percentage with College degree	1.1349	62%	0.04
Percentage of recent home construction	0.8582	34%	0.02
Median Household Income	1.3244	28%	0.01
Median Home Value	1.0021	44%	0.02
Population movement rate	0.0117	31%	0.05

Table 14: Logistic regression coefficients for new 25/3+ deployments in 2014-16.

Table 14 shows the confusion matrix for our K-Nearest Neighbor algorithm-based predictions for 2016 broadband deployments. The model provided an overall accuracy of 76%.

	Percentages of Potential Deployment Locations for 2016
True Positive (Predicted Deployment, Got Deployment)	10.2%
True Negative (Predicted No Deployment, Got No Deployment)	65.9%
False Positive (Predicted Deployment, Got No Deployment)	18.4%
False Negative (Predicted No Deployment, Got Deployment)	5.6%

Table 15: Confusion matrix for KNN deployment prediction model for 2016.

5.3 Analysis

From our evaluation of the regression coefficients over time, we caught several key trends. Generally, we saw a trend of increasing accessibility across the board, as the coefficients with income, economic growth, education, and population all decreased (albeit only slightly in most cases). This seemingly makes sense given that broadband deployment has become increasingly saturated in richer, more educated parts of the country (as seen in Tables 4, 5, and 6), meaning newer deployments will probably take place in other regions across the country. This trend of demographic accessibility is further reinforced by the mean and median values found in Tables 10 and 11, where the education rate percentages and the economic indicators (median household income and median home valuation) for new deployments are lower than those of existing deployments across the country (as seen in Table 5).

In terms of the relationship between broadband deployment and factors like education, median home valuations, and median household incomes, there are several values to note carefully. While median home valuation and percentage of recent home constructions have a strong correlation with the introduction of Internet services with threshold speeds, the percentage of recent population moved has a weak positive correlation. This low coefficient for the population movement rate makes more sense given the median values computed in tables 5 and 6, as we can see that these percentages are actually higher in the non-broadband regions. On the other hand, median home valuation is probably closely linked with median household income and can be a strong indicator of the householder's propensity to pay for higher speed broadband access. High home valuations are also a strong indication of the economic growth potential of a census block, making these regions more attractive to businesses and the general population [13]. As a result, with more businesses around, there is probably a better reason to deploy higher speed broadband networks. Similarly, the percentage of recent home constructions seems to be a strong indication of a general buyer's market for a region, indicating that it is relatively stable economically. As a result, there may be more incentive for broadband companies to offer higher-speed access here.

These changes in demographic profiles of deployment blocks can help provide greater insight into the parts of the country that are most likely to receive broadband in the near future. The k-Nearest Neighbor algorithm approach yielded approximately a 76% in predicting new deployment regions for the year 2016 (confusion matrix on the model's predictions can be found in Table 15), further indicating that the underlying demographic factors are strongly predictive. Moving forwards, this analysis can be further extended using additional demographic factors to improve our predictive abilities and potentially help in truly understanding year-to-year deployment patterns.

6 Large Provider Analysis

Large Provider Expansion

When analyzing broadband expansion for large broadband providers we first divide the dataset by year and then filter the data to extract broadband Internet by the 25/3+ definition. Furthermore, our criteria for defining large provider was by total blockgroups served in each of these years of which we took the top five. Then we compare this hold-out broadband set with the general Internet set by year and provider. We remove all the block-codes that already had broadband in each preceding year to analyze expansion while mapping them. While analyzing broadband expansion we also compare the expansion of the "large" providers, specifically AT&T, against the expansion of Google Fiber.

Separately, we were also curious whether the planned entry of Google Fiber in a market could explain, by correlation in time, any expansion of 100 Mb/s or faster broadband by the large providers.

Large Provider Expansion Results

Providers expand or contract their customer based in three ways: they gain or lose customers to competing providers in the same area, more households subscribe to broadband and providers expand into areas of the country that have not had broadband before.

The rate of broadband expansion in the US is increasing, with 2.1 million new broadband subscribers added in 2017. Overall telephone companies lost approximately 600,000 broadband subscribers in the year 2017 to Cable Companies. Specifically, the market share of telephone companies dropped from 39% in 2016 to 36% in 2017 with the market share of Cable Companies increased from 61% in 2016 to 64% in 2017 [17]. Analyzing this data at a granular level, we learn that AT&T only added about 100,000 broadband subscribers, while Verizon and CenturyLink each lost 79,000 and 283,000 subscribers in 2017. On the other hand, cable companies Comcast and Charter Communications each added over 1,000,000 subscribers. This trend continued in 2018 and in the first quarter of 2018, cable companies gained about 845,000 new subscribers whereas telephone companies lost about 45,000 subscribers.

Graphs and Tables Explained

The providers serve the areas shaded in gray, at any speed, offering an indication of their "native" service territory. Large telephone providers seem to rarely venture outside their service territory; they upgrade existing facilities to higher speeds instead. The blue areas highlight areas that first acquired 25/3 broadband definition in 2014, the red ones in 2015, and the green ones in 2016. In the tables, we recorded the number of block-codes each of the larger providers serviced. We also calculated the number of households that fell under these census blocks to analyze household reach. The household number is an overestimation as it adds all the households in the specific census block when in reality all the households in that census block might not be serviced by that provider alone. We used the FCC Staff Block Estimates to calculate the number of households for each year [18].

AT&T Expansion

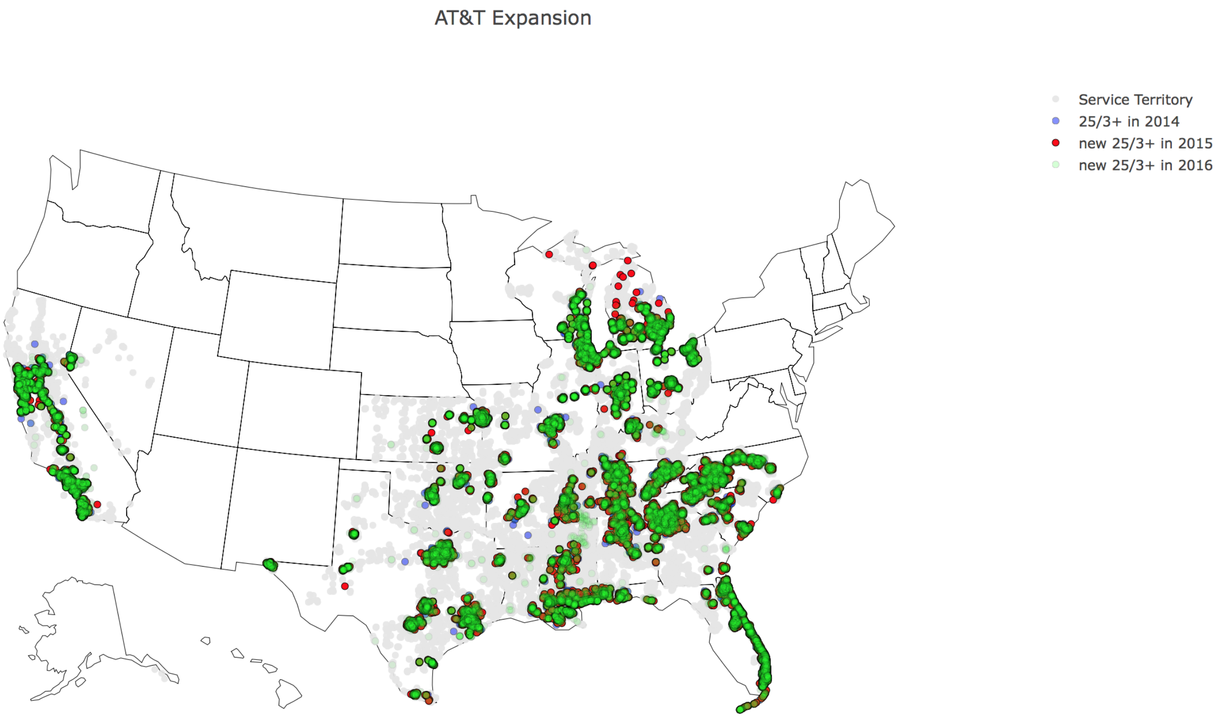

Figure 3: Graph for AT&T Expansion

Above we notice that AT&T expanded across the Southern and Eastern belt in 2014, but gradually expanded its reach to other areas still in the central United States. In 2016, AT&T seems to have taken a more aggressive route and expanded in a scattered manner in the Mid-East region. AT&T also enters into California in 2016. In gray we notice that AT&T has yet to provide broadband to all its subscribers and it seems that AT&T is on the path to do that as the expansion has primarily been in its territory region. Below are the quantitative results of our analysis regarding census blocks serviced and the number of households in these census blocks.

Quantitative Analysis

	Census Blocks	**25/3+ Households**	**Increase in 25/3+ Households Served**	**Subscriptions (all speeds) [17]**	**Total Blockcodes Served**
2014	110,766	1,613,238	-	16,486,000	1,755,318
2015	152,287	2,456,418	52.3%	15,832,000	1,876,969
2016	824,909	25,232,019	927.2 %	15,618,000	1,821,792

Table 16: AT&T Expansion Data

CenturyLink Expansion

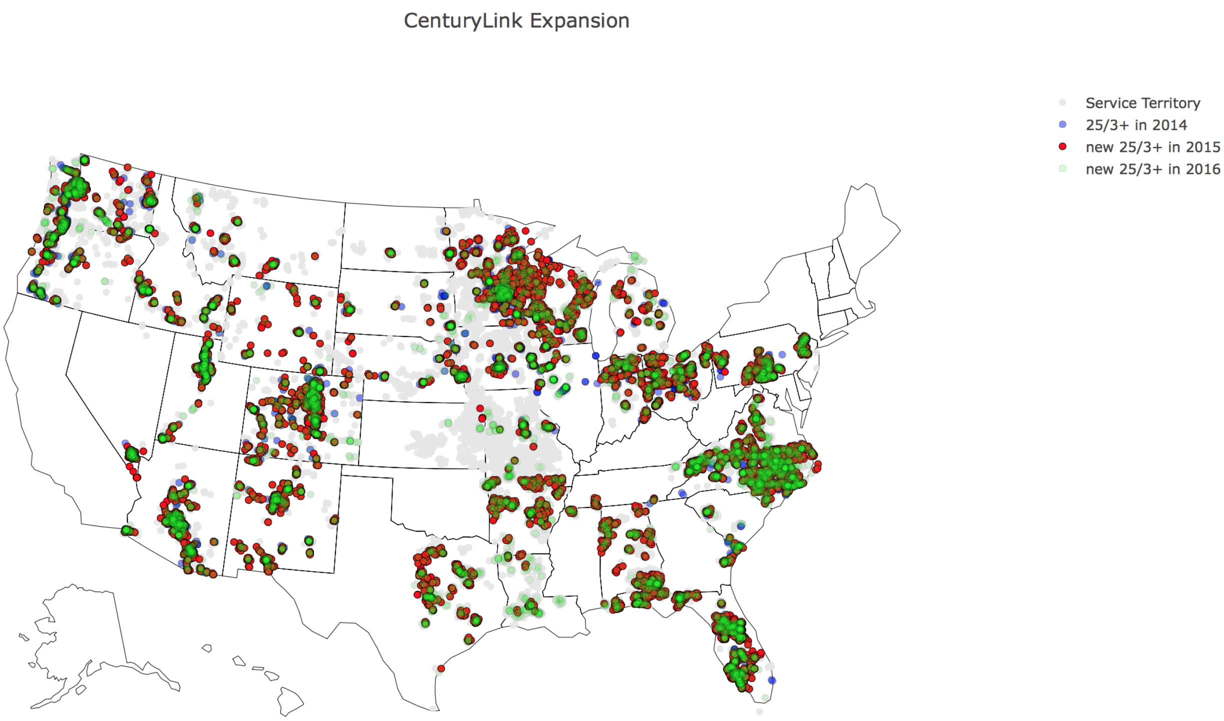

Figure 4: Graph for CenturyLink Expansion

CenturyLink is much more aggressive than AT&T in its bid to provide broadband Internet and as seen above it has expanded across the continental United States. CenturyLink does not expand its reach to new territories but continues to aggressively provide broadband Internet in its current territory region. CenturyLink appears to have expanded at a more rapid rate in 2015 than in 2016. Below are the quantitative results of our analysis regarding blockcodes serviced and the number of households in these blockcodes. Below are the quantitative results of our analysis regarding census blocks serviced and the number of households in these census blocks.

Quantitative Analysis

	Census Blocks	**25/3+ Households**	**Increase in 25/3+ Households Served**	**Subscriptions(all speeds) [17]**	**Total Blockcodes Served**
2014	243,677	7,380,903	-	6,063,000	1,075,235
2015	310,044	8,748,462	18.5 %	6,071,000	1,152,569
2016	341,297	9,455,039	8.07 %	5,950,000	1,158,510

Table 17: CenturyLink Expansion Data

Comcast Expansion

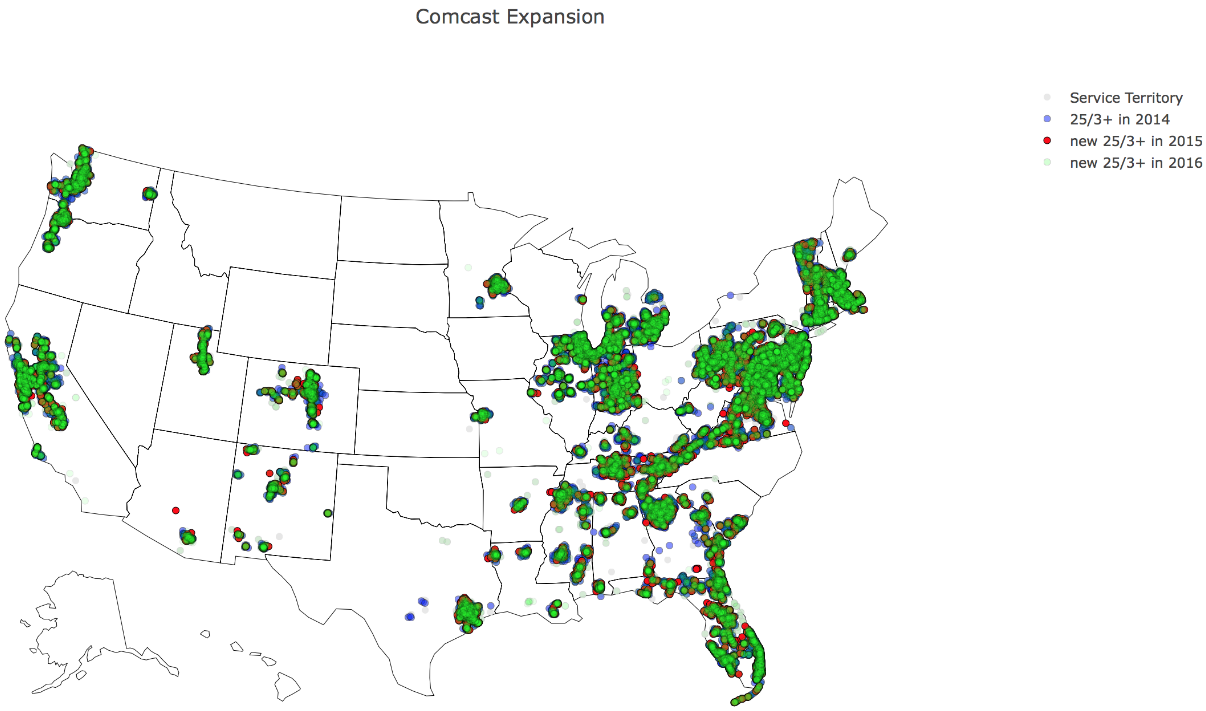

Figure 5: Graph for Comcast Expansion

Comcast has a spread-out reach across the United States with presence in the Mid-East, Northeast, South-East as well as the West Coast regions. However, Comcast does not have a heavy reach in the Mid-West region of the US. Below are the quantitative results of our analysis regarding census blocks serviced and the number of households in these census blocks.

Quantitative Analysis

	Census Blocks	25/3+ Households	Increase in 25/3+ Households Served	Subscriptions (all speeds) [17]	Total Blockcodes Served
2014	1,606,962	41,951,312	-	21,586,000	1,606,962
2015	1,607,456	42,257,594	0.7 %	22,868,000	1,607,456
2016	1,674,852	43,171,304	2.2 %	24,316,000	1,674,852

Table 18: Comcast Expansion Data

Cox Expansion

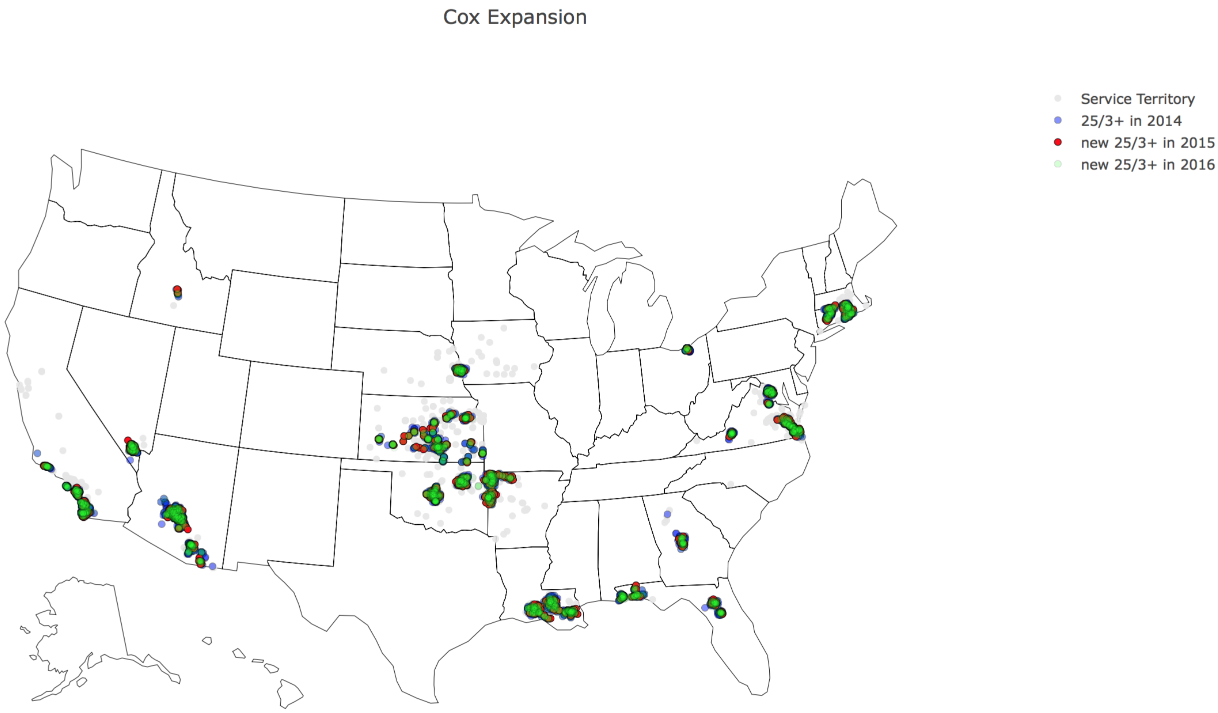

Figure 6: Graph for Cox Expansion

Cox is not as up to pace with the other providers in terms of population served and has a scattered reach across the US. Cox has a light presence in all the regions it services. It appears to have a concentrated presence in Louisiana, Kansas, and Oklahoma. Its growth along with reach seems to be far behind the other providers. Below are the quantitative results of our analysis regarding census blocks serviced and the number of households in these census blocks.

Quantitative Analysis

	Census Blocks	**25/3+ Households**	**Increase in 25/3+ Households Served**	**Subscriptions(all speeds) [17]**	**Total Blockcodes Served**
2014	286,285	7,942,981	-	NA	286,725
2015	286,704	7,992,330	0.62 %	NA	287,145
2016	277,443	7,981,569	-0.13 %	NA	277,885

Table 19: Cox Expansion Data

Verizon Expansion

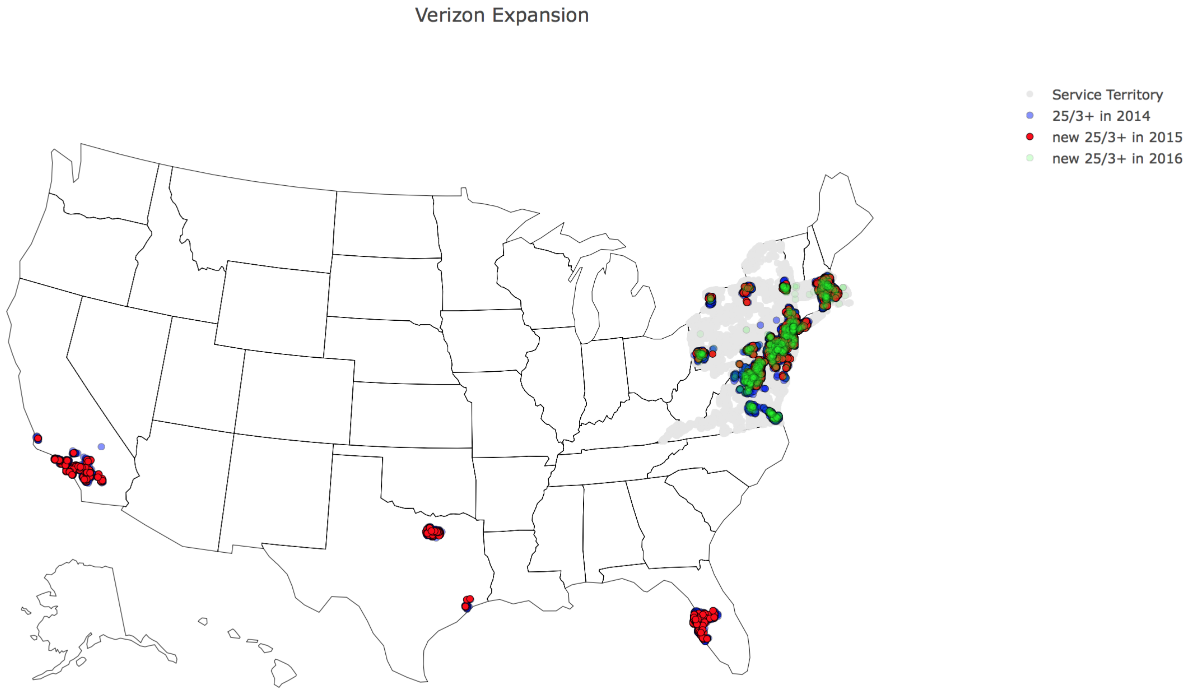

Figure 7: Graph for Verizon Expansion

Verizon arguably has the most interesting expansion pattern as it does not build on its preceding year's progress. It actively targets new regions across all years. In 2014 it provided broadband to concentrated regions in Texas, California and the Northeast belt. In 2015, it targeted completely new regions from the ones targeted in 2014 with the exception of the Northeastern belt. In fact, Verizon expands its broadband network in the Northeast across all the years. Below are the quantitative results of our analysis regarding census blocks serviced and the number of households in these census blocks.

Quantitative Analysis

	Census Blocks	**25/3+ Households**	**Increase in 25/3+ Households Served**	**Subscriptions(all speeds) [17]**	**Total Blockcodes Served**
2014	456,032	15,054,184	-	9,146,000	966,523
2015	432,415	14,205,888	-5.63 %	9,223,000	971,790
2016	375,776	12,792,356	-9.95 %	7,038,000	766,278

Table 20: Verizon Expansion Data

AT&T 100 Mbps vs Google Fiber

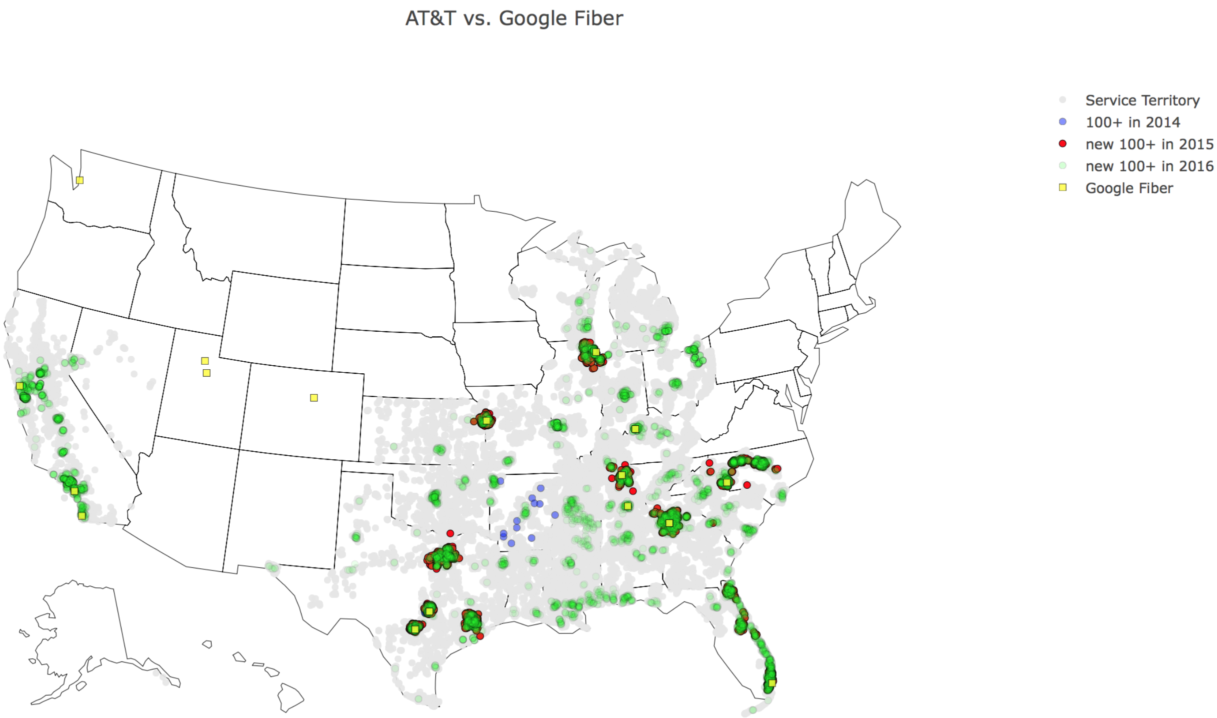

Figure 8: Graph for Expansion of Google Fiber and AT&T 100 Mbps

In the graph above we analyze the territory similarities between AT&T and Google Fiber in providing high speed Internet. First, Google Fiber announced its intention to provide 100 Mbps broadband service in its target cities after which At&T also upgraded its offerings in the area. We noticed that there is a similarly in the regions they operate. Google Fiber operates in the region where AT&T has the highest concentration of 100 Mbps households. Below are the quantitative results of our analysis regarding census blocks serviced and the number of households in these census blocks for AT&T 100 mbps.

Quantitative Analysis

	Census Blocks	**100+ Households**	**Increase in 100+ Households Served**
2014	4,495	147,347	-
2015	15,395	678,043	360.12 %
2016	44,443	958,142	41.30 %

Table 21: AT&T 100 Mbps data Expansion Data

7 Conclusion and Future Work

We analyzed various different demographic features and deployment metrics to determine and understand the growth of broadband service offerings across the contiguous USA. We identified key economic and education factors associated with broadband growth and also analyzed the patterns of different broadband providers as they expanded across the continental United States. Our findings indicated that there was a decreasing relationship between a household's wealth, education, and urbanity and their access to new broadband deployment. Currently, only 4% of Americans in urban areas do not have access to sufficient Internet speeds and this implies that new growth would have to be in rural or less urbanized sectors [1]. For comparison, our analysis shows that among the rural population, approximately 30 percent do not have access to 25/3 broadband and approximately 15 percent do not have access to 10/1 broadband.

A key finding of our research lies in the regression coefficients that we were able to identify between demographic variables and broadband deployment between 2014-2016. We identified that, while income and education maintained a relatively strong positive correlation with broadband availability, the correlation became less strong over time, suggesting that the expansion of broadband over the past 4 years has been allowing less affluent areas to get higher speed offerings. Running a k-Nearest Neighbor prediction algorithm resulted in a 76% accuracy for predicting new deployment regions for the year 2016, indicating that our identified demographic factors related to education levels and economic growth are strongly predictive. The analysis of predictive expansion of broadband is an intriguing prospect for further research.

In order to achieve viable and statistically relevant findings from our analysis, there were several corner cases and potential sources of error that we sought to address. One key flaw in the FCC Form 477 data we encountered was the fact that a statistically significant number of census blocks lose broadband coverage from year to year. We also found thousands of blocks where broadband disappeared and then re-appeared across a two-year time span. Because providers do not provide any supplementary commentary along with their reports, we are not able to identify a definite reason for these anomalies. More research on this issue will provide valuable insights into these patterns in the 477 data. An additional point of note was the presence of zero population census blocks, including a fraction with broadband coverage. Upon further analysis, we determined that many of these blocks contained part-time or vacation homes.

Our analysis additionally focused on the broadband expansion of large Internet Service Providers to determine what patterns existed. Providers generally expanded their customer base in one of three ways over the analyzed time period: first, they acquire customers from competing providers in the same area; second, they generate new household subscribe to the broadband service; and third, they expand into areas of the country that have not had broadband before [15]. Looking at the figures above in Section 6 for each provider we notice that the rate of broadband expansion year over year has increased for most large providers, indicating an overall increase in the broadband footprint and the rate of ISP expansion in the US. Interestingly, looking at the planned expansions of Google Fiber, we see that there is a strong geographical correlation between planned Google Fiber sites and ISP expansion, especially for AT&T.

Ultimately, the results of our research provide further insights into the nature of recent broadband deployments. This can potentially serve as a starting point to help guide corporate strategy and support informed government decision making and allocation of funding for broadband projects.

References

[1] Federal Communications Commission. 2018 Broadband Deployment Report. https://www.fcc.gov/reports-research/reports/broadband-progress-reports/2018-broadband-deployment-report, Feb 2018.

[2] Federal Communications Commission. 2016 Broadband Deployment Report. https://www.fcc.gov/reports-research/reports/broadband-progress-reports/2016-broadband-progress-report, Jan 2016.

[3] Kevin Barefoot. Defining and Measuring the Digital Economy. https://www.bea.gov/digital-economy/_pdf/defining-and-measuring-the-digital-economy.pdf, Mar 2018.

[4] Kevin A. Hassett and Robert J. Shapiro. The Impact of Broadband and Related Information and Communications Technologies On the American Economy. https://internetinnovation.org/images/misc_content/Report_on_the_Economic_Impact_of_Broadband_-_Hassett-Shapiro_-_Rev_-_March_23_2016.pdf, Mar 2016.

[5] Federal Communications Commission. 2015 Broadband Progress Report. https://www.fcc.gov/reports-research/reports/broadband-progress-reports/2015-broadband-progress-report, Feb 2016.

[6] Federal Communications Commission. Connect America Fund Phase II Auction (Auction 903). https://www.fcc.gov/auction/903, Aug 2018.

[7] United States Census Bureau. What are Census Blocks? https://www.census.gov/newsroom/blogs/random-samplings/2011/07/what-are-census-blocks.html, Jul 2011.

[8] United States Census Bureau. Geographic Terms and Concepts - Block Groups. https://www.census.gov/geo/reference/gtc/gtc_bg.html, Dec 2012.

[9] United States Census Bureau. Geographic Terms and Concepts - Census Tract. https://www.census.gov/geo/reference/gtc/gtc_ct.html, Dec 2012.

[10] United States Census Bureau. The Urban and Rural Census Classifications. https://www2.census.gov/geo/pdfs/reference/GARM/Ch12GARM.pdf.

[11] United States Census Bureau. American Community Survey Methodology. https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html, Jun 2018.

[12] Federal Communications Commission. Fixed Broadband Deployment Data from FCC Form 477. https://www.fcc.gov/general/broadband-deployment-data-fcc-form-477, 2018.

[13] Joshua Aizenman. Real Estate valuations and Economic Growth: The Cost of Housing Cycles. Vox EU, Oct 2016.

[14] Charles Ka Yui Leung. Economic Growth and Increasing House Prices. Pacific Economic Review, Dec 2003.

[15] Wireline Competition Bureau. Internet Access Services: Status as of December 31, 2016. Federal Communications Commission, Feb 2018.

[16] Katura Jackson. Modernizing the FCC Form 477 Data Program. https://www.federalregister.gov/documents/2017/08/24/2017-17901/modernizing-the-fcc-form-477-data-program, Aug 2017. The Federal Register.

[17] Leichtman Research Group. Research Notes. https://www.leichtmanresearch.com/research/, May 2018.

[18] Federal Communications Commission. Staff Block Estimates. https://www.fcc.gov/reports-research/data/staff-block-estimates, 2016.