**Part 1. Some visuals**

Figure 1. Broadband speed options available to 95% of population, per county.

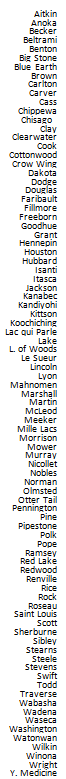


Figure 2. Big city vs. Small town Broadband median speed growth.

Shows that while for larger cities available internet speeds grow exponentially (with R2>97%), for smaller cities the trend is linear. For the smallest cities with population less than 1000 the growth is slowing down, with natural log function as the best fit for the data.

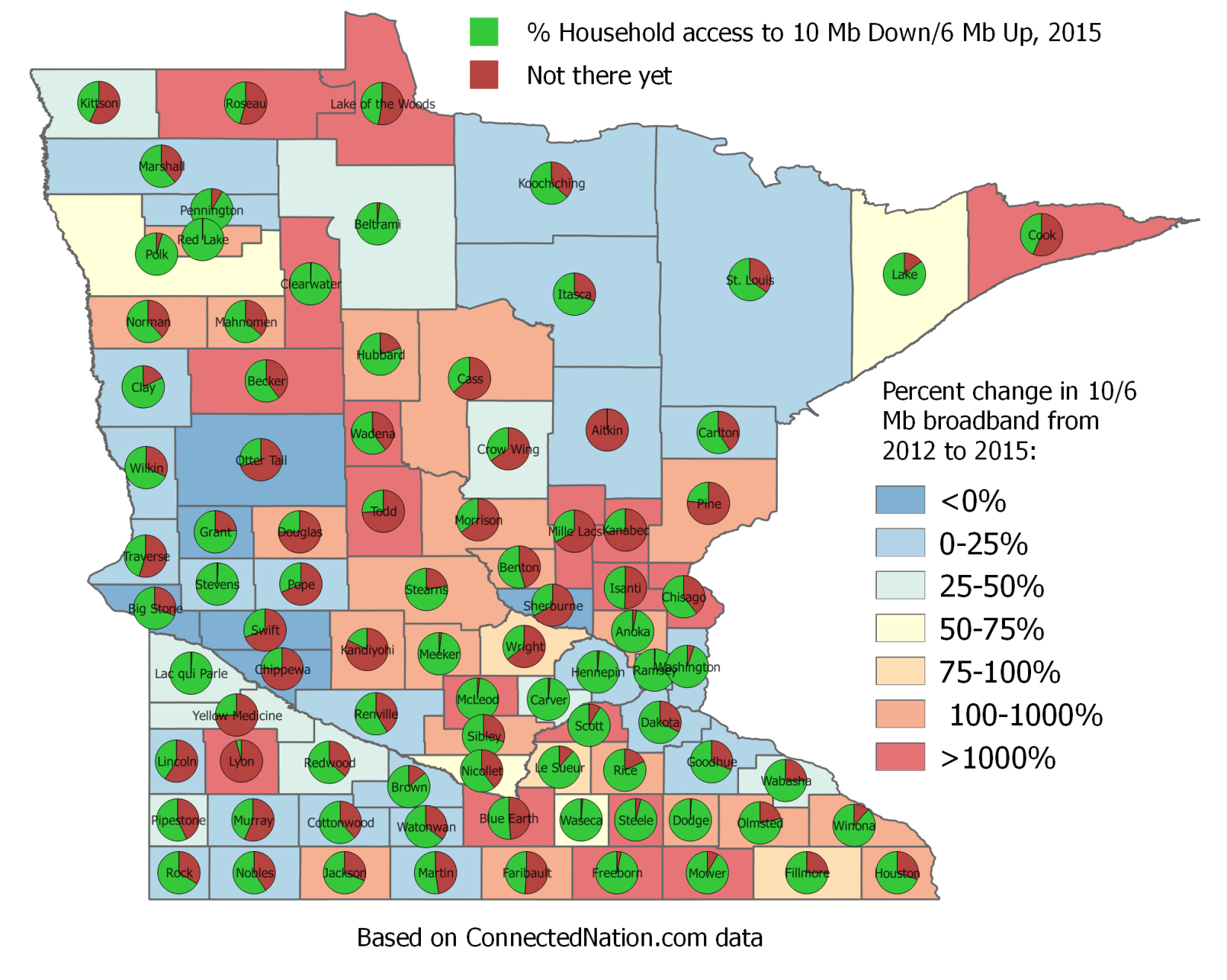
Data source: Akamai 2015

Figure 3. Big town median internet speed change from 2008 to 2015, Mb/second.

Figure 4. Small towns.

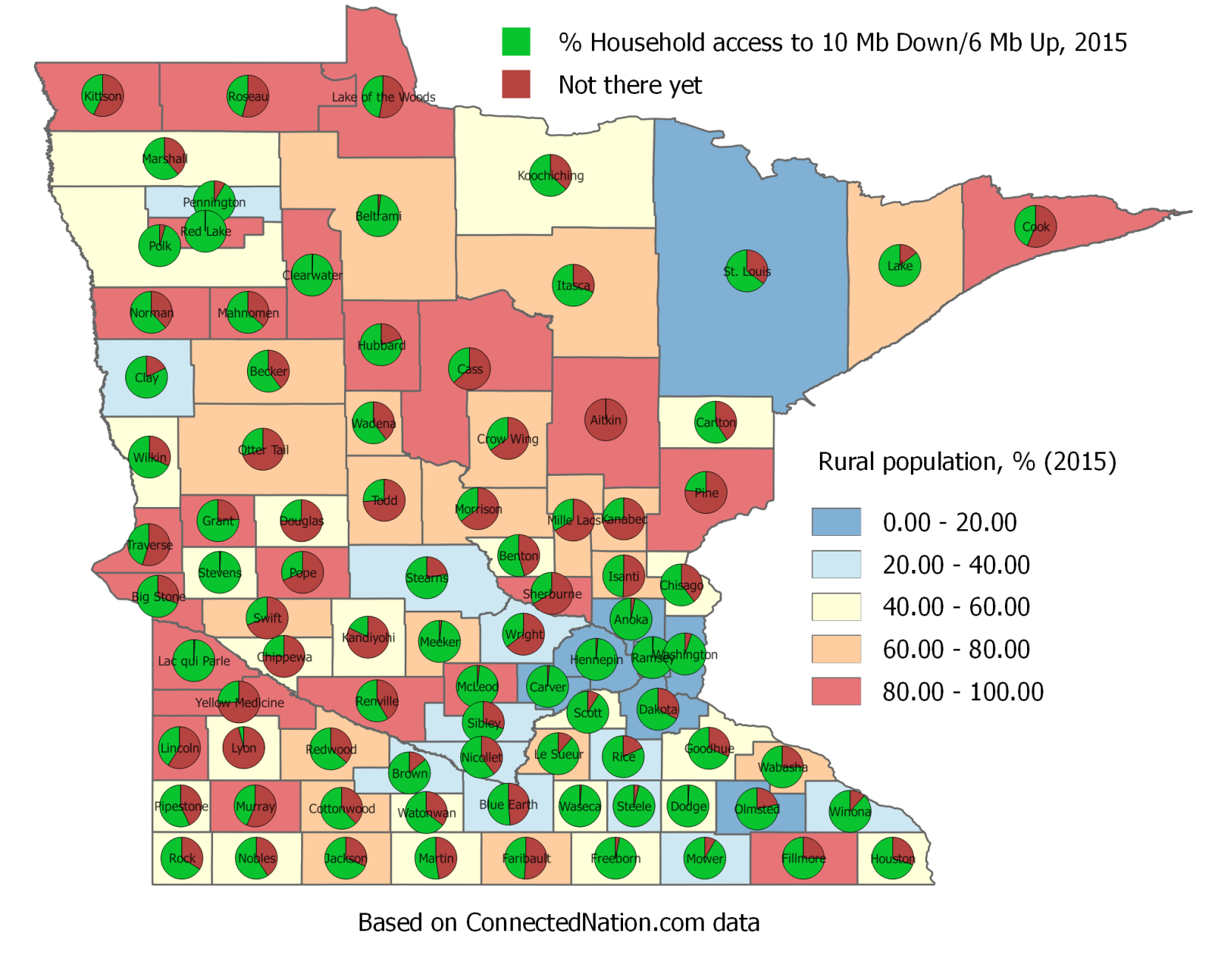
Figure 3 and 4 tell the story that while small cities can boast a few successful stories, median internet speeds are improving much faster in big cities.

Figure 5. Broadband coverage and trends in Minnesota, 2012 – 2015.



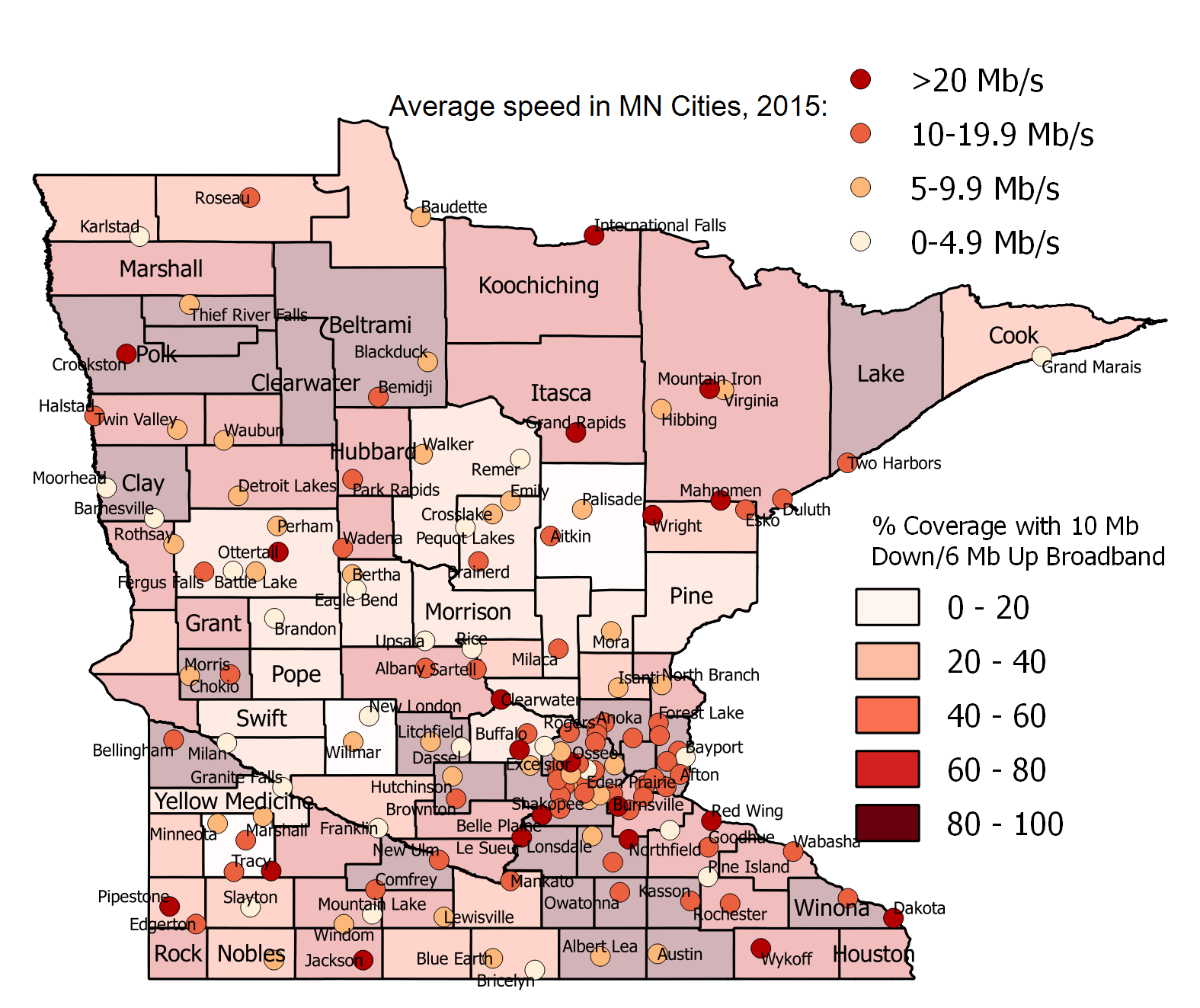
2012 and 2015 data on 10/6 internet is from Connected Nations, is in open access. Used as-is, to show changes. The map shows % households that have access to 10/6 internet in that county (to meet federal guideline), and what % does not. The background shows how much coverage has changed from 2012 to 2015.

Figure 6.



The map shows % households that have availability to 10/6 internet in that county (to meet federal guideline), and what % does not. The background shows % rural population in that county as of 2010 Census.

Figure 7



I sorted out the raw data a bit, and was able to build a map. I did not use 2007, because it is very incomplete and somewhat erroneous I though. Also I did not include cases where there is less than 10 IP in the analysis.

The map shows our “champion” and “not-so-champion towns” in implementing 10/6 Mb internet. As a background I showed % households that have access to 10/6 Mb internet.

Reference: 2012 Connected nation data on 6/10 Mb internet was taken from here, only place I could find it on the internet. I copied values from the map to Excel. https://mn.gov/commerce/images/2012\_Broadband\_Annual\_Report.pdf

**Part 2. Statistical analysis**

**Methods**

1. Since this is not an article, but a “drafty” report, I’ll be brief. All data and a couple of statistical analyses are available at your disposal. If you have any questions, please ask.
2. The first goal was to see what factors impact annual broadband speed change (e.g. acceleration), in MN cities. The second goal was to see what characteristics are common for towns with fastest internet speeds.
3. Original dataset had 354 towns. Only towns that had at least one quarterly observation during each year from 2008 to 2015 were used. The rest of the data feels too truncated and incomplete. Final dataset has 125 towns out of 68 MN counties. The towns are in Excel and on the map.
4. I ran 200 alternative models with different factors and interaction combinations to get a good feel of the data.
5. The variables tested are shown in the table below.

|  |  |  |
| --- | --- | --- |
| code | description | source |
| X | latitude | MN DNR |
| Y | longtitude | MN DNR |
| name\_city | city name | MN DNR |
| spzb\_zones | StatePlane Zone (1-North, 2-Central, 3-South) | MN DNR |
| landareameters\_city | city land area in meters | MN DNR |
| waterareameters\_city | city water area in meters | MN DNR |
| population\_city | total city population | Census 2010 |
| malepop\_city | male city population | Census 2010 |
| femalepo\_city | female city population | Census 2010 |
| whitepop\_city | white city population | Census 2010 |
| indianpop\_city | indian city population | Census 2010 |
| HHnumber\_city | number of city hoseholds | Census 2010 |
| agemedian\_city | median city age | Census 2010 |
| agemedianmale\_city | median male age | Census 2010 |
| agemedian\_female | median female age | Census 2010 |
| HHsize\_city | average household size | Census 2010 |
| landareamiles\_city | city land area In miles | calculated from GIS data |
| popdensity\_city | population density | calculated from Census 10 and MN DNRdata |
| HHdensity\_city | household density | calculated from Census 10 and MN DNRdata |
| percindian\_city | peccent indian population | calculated from Census 10 and MN DNRdata |
| percentmal\_city | percent male population | calculated from Census 10 and MN DNRdata |
| ave2008 | average speed in 2008 | Akamai |
| ave2009 | average speed in 2009 | Akamai |
| ave2010 | average speed in 2010 | Akamai |
| ave2011 | average speed in 2011 | Akamai |
| ave2012 | average speed in 2012 | Akamai |
| ave2013 | average speed in 2013 | Akamai |
| ave2014 | average speed in 2014 | Akamai |
| ave2015 | average speed in 2015 | Akamai |
| percchange0809 | speed change from 2008 to 2009 | calculated from Akamai |
| percchange0910 | speed change from 2009 to 2010 | calculated from Akamai |
| percchange1011 | speed change from 2010 to 2011 | calculated from Akamai |
| percchange1112 | speed change from 2011 to 2012 | calculated from Akamai |
| percchange1213 | speed change from 2012 to 2013 | calculated from Akamai |
| percchange1314 | speed change from 2013 to 2014 | calculated from Akamai |
| percchange1415 | speed change from 2014 to 2015 | calculated from Akamai |
| medianannualchange106\_city | mediana annual speed change | calculated from Akamai |
| aveannualchange\_city | average annual speed change | calculated from Akamai |
| max2008 | maximum speed in 2008 | Akamai |
| max2009 | maximum speed in 2009 | Akamai |
| max2010 | maximum speed in 2010 | Akamai |
| max2011 | maximum speed in 2011 | Akamai |
| max2012 | maximum speed in 2012 | Akamai |
| max2013 | maximum speed in 2013 | Akamai |
| max2014 | maximum speed in 2014 | Akamai |
| max2015 | maximum speed in 2015 | Akamai |
| avemaxnnualchange\_city | average max annual speed change | calculated from Akamai |
| name\_cnt | county name |  |
| areameters\_cnt | county area in meters |  |
| perimmeters\_cnt | county perimeter in meters |  |
| 106perccover2012\_cnt | percen coverage with 10/6 internet in 2012 | ConnectedNation |
| 106perccov2015\_cnt | percen coverage with 10/6 internet in 2015 | ConnectedNation |
| aveannual106change\_cnt | annual difference between 2012 and 2015 10/6 coverage | calculated from ConnectedNation |
| percchangeannual\_cnt | percent annual difference between 2012 and 2015 10/6 coverage | calculated from ConnectedNation |
| incomeperpers\_cnt | income per capita, county | Census 2010 |
| medianHHincome\_cnt | income per household, county | Census 2010 |
| population\_cnt | county population | Census 2010 |
| citytocountypopperc | percent population, city to county | Census 2010 |
| Hhnum\_cnt | number of hoseholds, county | Census 2010 |
| percentruralpop\_county | pecent rural population, county | Census 2010 |
| popdensity\_cty | population density, county | Census 2010 |
| medianage\_cty | meian age, county | Census 2010 |
| numproviders\_cty | dnumber of providers per county | ConnectedNation |
| blandin\_cty | Blandin activity in county (1-yes | Ann Tracy |
| cooppresent\_cnty | coop present in county (1-yes) | Ann Tracy |
| college\_city | college precent in city (1-yes) | Wikipedia |
| meetfedgoal | 10/6 internet is available for 95% population in 2015 (1-yes) | calculated based on Akamai |

1. The raw data, variable description and their statistical properties (like standard deviations) are provided in the attached Excel file. Please feel free to play with it further. All data and its sources were carefully checked and double-checked for error, so I believe you can skip that step.
2. Box-Cox was used as guidance for variable transformation. Most of the variables were converted into natural log function. After transformation, scatterplots and P-P plots were examined to make sure distributions and residuals look acceptable. The visuals are also provided in the Excel file. VIF tests for multicollinearity showed there is no multicollinearity problem.

**Story**

The correlation coefficients between most of the variables are very small and insignificant. The reason why it took me much longer to finish this project is because I am not used to giving up before I find at least some associations. Well I got some positive results, but underlying statistical evidence is pretty weak. So I would use these results with caution.

When carefully controlled for multicollinearity, the models adjusted R 2 never exceeded 30%. It is not a big problem, however, as we are more interested in a single factor effects. The bigger problem is that the significant variables (>0.05) in the model have signs I did not expect them to have.

Good things are that the variables in the model are all statistically significant, and have very low collinearity - which means they do have “some” effect on the dependent variable.

I guess the more important and sound findings of my modeling were specifying the variables that have nothing to do with internet “acceleration” or internet speeds under any condition.

I left some of the statistical analysis results in Excel. If you want deeper understanding please check there or ask me. Here I will provide the summary of results.

**Results and Conclusions**

Model 1. The data analysis demonstrates that the largest relative annual average increase in internet speeds in the last 8 years is observed in the smaller rural cities that have at least a few internet providers available in the area (i.e. stronger competition). The average internet connection speed in 2015 is also an important factor. Basically it is getting faster where it is already faster. There was also some evidence that the cities where coops are present usually demonstrate slower average speed when compared to the cities without the coops. This is a surprising finding which can probably be explained by the fact that coops usually operate in the country side.

Ln Population (Y) vs ln Average Annual Speed Change 2008-2015 (X)

# of providers (Y) vs Average Annual Speed Change 2008-20015 (X)

Ln Average Speed in 2015 (Y) vs ln Average Annual Speed Change 2008-2015 (X)

The factors that never showed any link to average speed increase include median age, race and gender composition (like White vs. Native American, household size, median household or per capita income, presence of college in town, town population density, a ratio of town population vs. county population, and whether the county in which the town is located is predominantly rural or urban.

Model 2. The characteristics of towns that feature the fastest internet in Minnesota in 2015 usually are larger population, college, and that the city is located in a county with a high percent of urban population. These towns also continuously demonstrate faster than average increase in internet speeds on an annual basis. The data also showed some (very) limited evidence that a higher medium income and lower age result in faster internet. But the latter two statements really require more data/research.

Ln Population (Y) vs ln Average Speed 2015 (X)

% Rural Population in the County (Y) vs ln Average Speed 2015 (X)

College City (1=Yes) (Y) vs ln Average Speed 2015 (X)

Factors that showed no effect on broadband speeds include city location, population or household density, gender or race composition or number of providers.

The important factors are summarized in Table 1.

|  |  |
| --- | --- |
| Factors of Broadband acceleration | Factors of Broadband speed |
| City population  Average internet speed in the area  Competition among providers | City population  Internet infrastructure and availability  College in town  % Rural population in county |

**Model 3.** In additional to the models above, I also ran some models for counties only. The goal was to see what are the common characteristics for the counties that in 2015 had the highest coverage of 10 Mb Up/6 Mb Down speeds (as per federal guideline). The data I used is publicly available, mostly from Census 2010 and ConnectMN/ConnectedNation. The details are in Excel if interested, feel free to explore more.

The finding was somewhat interesting, if not important. The only significant characteristic found to affect broadband availability is how much rural vs urban population lives in a specific county. The variable (percrural) depicts the percent of rural to urban population, based on Census 2010 data. It changes from 0.18% for Ramsey County to 100 % in 16 counties like Red Lake, Cass, Norman and Lincoln. Other details of the variable provided below.

|  |  |
| --- | --- |
| *percrural* | |
|  |  |
| Mean | 61.34274 |
| Standard Error | 3.023243 |
| Median | 62.15933 |
| Mode | 100 |
| Standard Deviation | 28.19893 |
| Sample Variance | 795.1796 |
| Kurtosis | -0.86998 |
| Skewness | -0.13289 |
| Range | 99.81283 |
| Minimum | 0.187166 |
| Maximum | 100 |
| Sum | 5336.818 |
| Count | 87 |
| Confidence Level(95.0%) | 6.010007 |

The model is cov2015 = 101.342433550561-9.09429127193845\*lnpercrural.

While R2 is still very low, the Pr > |t| is 0.002, which makes the variable very significant.

The regression mathematically shows that a 1% increase in Percrural is associated with an average 0.1 % decrease broadband availability at 10/6 Mb.

As an example, a county with 30% rural population will on average have 20% (!) more broadband coverage at 10/6 Mb speeds than counties with 100% rural populations (as (100-30)/30\*100=233%).

The scatterplot below shows the linear trend between the two. R2 of 11% is pretty good, since we are talking about a single variable.

Hope this report will have some use. Good luck.

**P.S.S.** I had fiber optics internet installed in my house today, so I had a chance to talk to the guys about this. In their opinion, by far the most important variable that drives internet speeds up is a completion between providers.