Broadband GINI

Bryan Morgan & Hasnain Nazar

Project goals

The main purpose of this project was 2 fold. One, to demonstrate to users that substantial and meaningful internet access is not as simple as “do you have broadband?”. Two, we also wanted to let users explore the data for themselves. Initially we had thought about creating a very structured narrative to guide the users through but this proved to be extremely difficult for a number of reasons. First there are some elements of the story that many people are familiar with, such as rural areas having poorer access to the internet, and we felt that we might run the risk of losing user interest if we only highlighted things they already knew. Additionally, because this issue has been discussed by so many different groups, with differing agendas, we felt that we would run the risk of just being another voice amongst many. With these two concerns in mind, we decided that we wanted to allow users more freedom to explore the information, though we did still restrict the exploration to what we considered to be the most important factors for meaningful internet access. After much discussion, we decided to boil down our definition of meaningful internet access to three different components: 5 or more Internet Service Providers (ISPs), download speeds greater than 25 megabits per second and having access to the internet via multiple devices such as desktop/laptop and a mobile device. We felt that once an individual met all three of these criteria, they were meaningfully connected to the internet, that they were able to leverage the internet in whatever way they needed and were not excluded from any aspect of it.

Having 5 or more ISPs meant that there was actual choice in terms of which provider to go with. The reason there needed to be 5 or more ISPs was because there are typically 2 large providers such as Comcast and AT&T who dominate the market, subsequently there needed to be at least 3 other options in order for there to be reasonable competition in the market. By having competition, consumers are able to both have a choice in which provider they go with, as well as helping drive some of the cost down for consumers. While this was by no means a deciding factor, we felt that having a limited number of ISPs did act as a limiting factor in how people interacted with the internet and contributed to slower speeds as well as higher costs.

The next factor we considered was the download speed. The FTC has recently changed its definition of broadband to 25 megabits per second (mbps) to reflect the increasing bandwidth being used by modern web applications. We felt this was a critical component to identifying whether or not someone is adequately connected to the internet because without it many web applications are either completely unavailable or have difficulty functioning making them essentially unusable. This metric is related to the number of ISPs available, and while these are important indicators of being completely connected, we felt there was one final component that was also required; being able to connect in different ways via multiple devices.

Of the three metrics used to evaluate whether someone is completely connected, this may be the most contentious as many will say that having a mobile device is not a necessity.

Yet we argue that there are two main reasons that having both desktop/laptop as well as mobile devices to access the internet demonstrates unconstrained internet access. First there is a growing social expectation that people are constantly connected to the internet, and those who are not often miss out on opportunities. Additionally, there is a substantially different experience between mobile and desktop, and having only one of the two means that an individual is constrained to only what is possible on that platform. With this framework in mind we then went about creating visualizations for each component, as well developing a composite score and subsequent visualization.

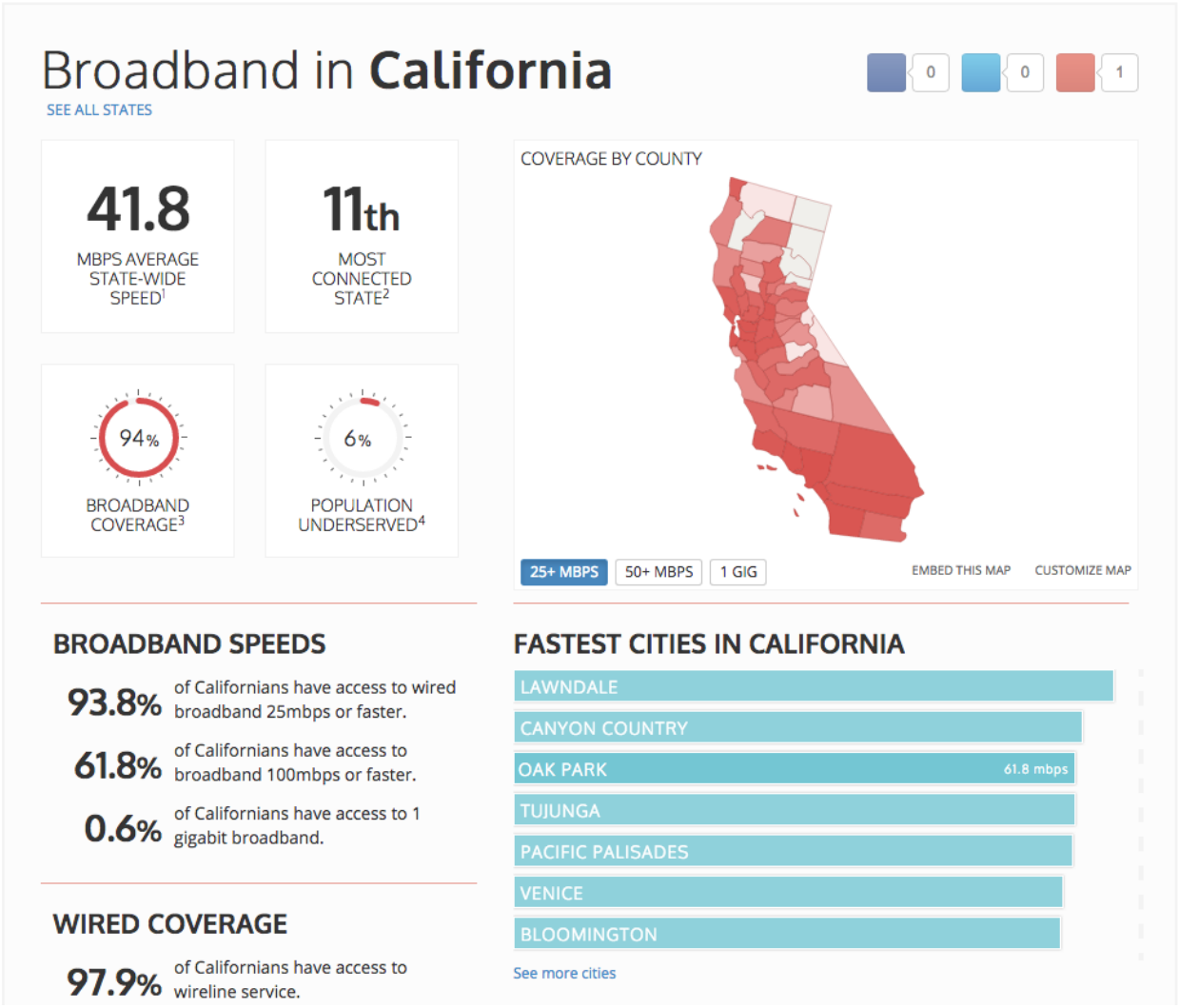
The visualizations, as previously mentioned, were designed to support user exploration. Within the confines of these three criteria, we wanted to allow users to be able to interact with the data while framing the interaction so that they could see patterns in terms of accessibility within geographic regions in California. In order to do this we integrated maps in to our designs, the goal of which was to allow users to understand different counties in context, meaning where they are located in California as well as what other counties are nearby. Additionally, when visualizing broadband speeds in California we broke California into geographic and cultural regions to illustrate general trends amongst counties in a similar region. In addition to the maps, we also used bar graphs to support a more direct comparison of the data being contextualized by the map. Once the user was more familiar with the geographic regions we switched to different types of visualizations, one radar graph and one treemap series. These two styles were intended to support very different interactions. The radar graph is being used to give users a quick view of what portion of the population are using only one type of device, essentially what percent are not getting the complete internet experience. While the treemap series is being used to organize geographic regions, then counties, by how well they are doing according to our computer “Broadband GINI”. The user is able to quickly see which regions are doing the best as they are organized (top to bottom, left to right) by their scores, as well as see on which criteria are the counties doing well or poorly (based on the area of the different criteria).

Each visualization was designed to support the specific data it was displaying as well as encourage user exploration. Additionally though, each visualization was intended to build upon the previous, allowing users to have a more holistic perspective when they reach the treemap. Our intent was to allow users to have enough information to make sense of the treemap and quickly compare regions and counties once they reach the end of the page.

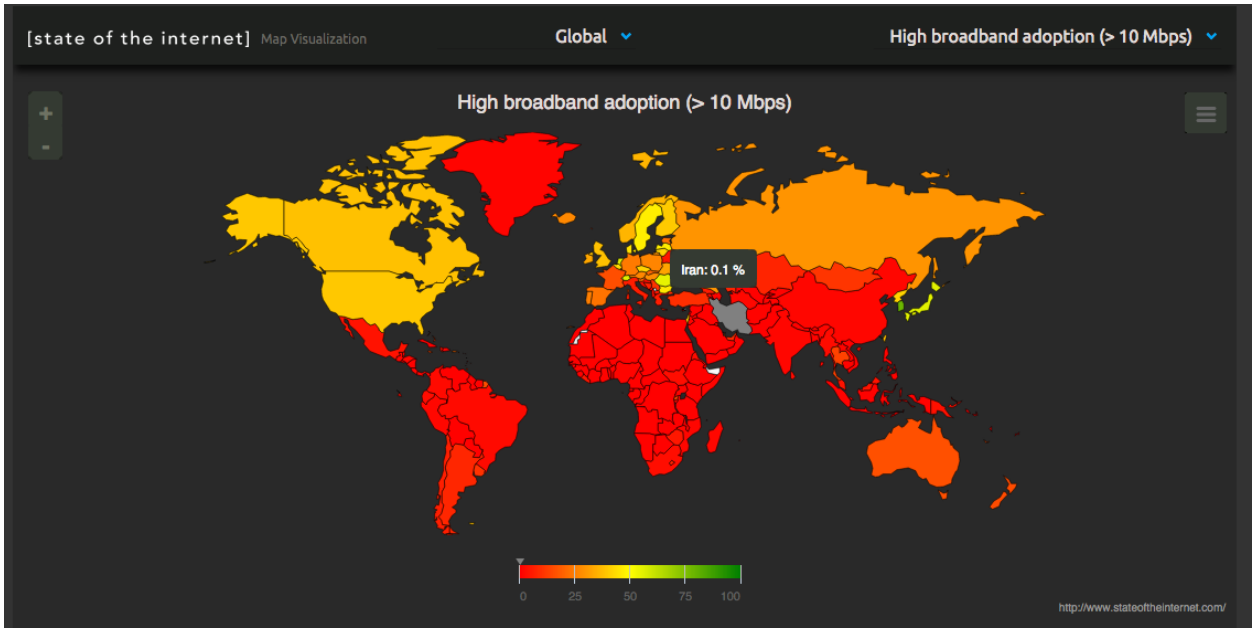
Discussion of related work

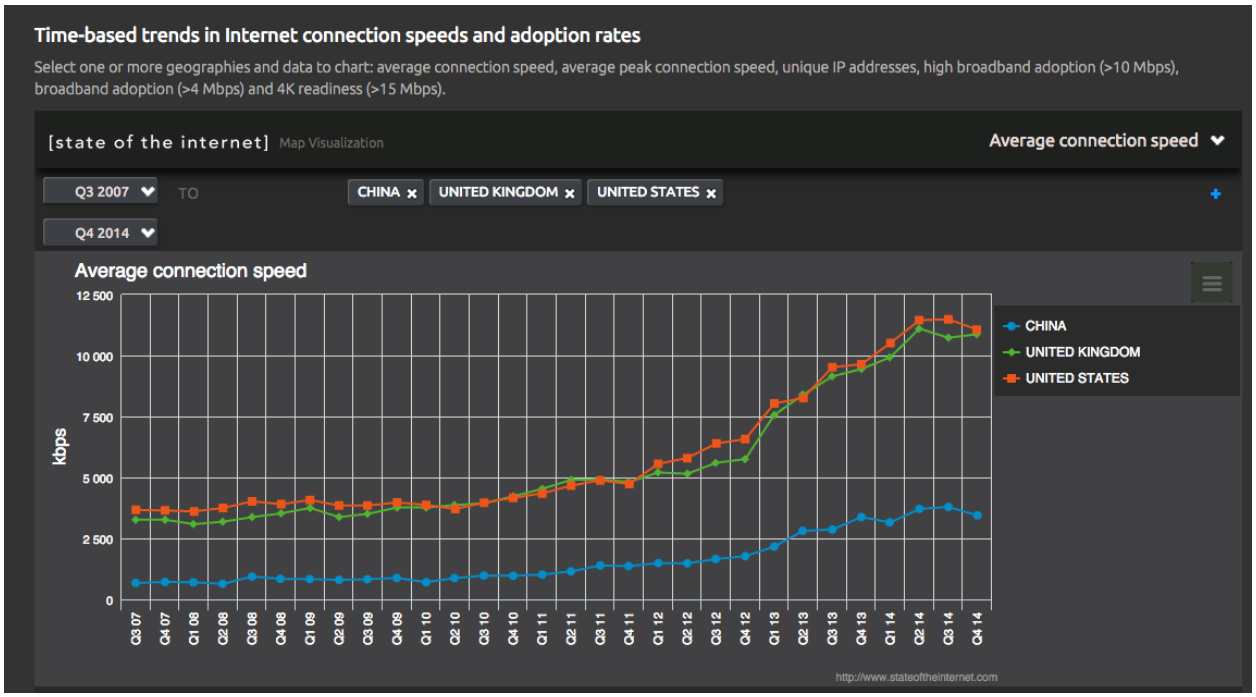
- You must discuss at least 3 pieces of related work for each student on the project. You must say how your project relates to these readings or projects, and provide an image or sufficient description of the related work for the reader to understand how it is related.
    - Related work:
        - Broadband Now
            - (http://broadbandnow.com/California/Berkeley?zip=94702)
        - Akami State of the Internet
            - (http://www.stateoftheinternet.com/trends-visualizations-connectivity-global-heat-map-Internet-speeds-broadband-adoption.html)
        - Geographies of Google Search
            - (http://geography.oii.ox.ac.uk/?page=geographies-of-google-search)
        - Digital divide exacerbates US inequality
            - http://www.ft.com/intl/cms/s/2/b75d095a-5d76-11e4-9753-00144feabdc0.html?segid=0100320#axzz3WwdmUgW6
        - Data Visualization: Breaking Down the Digital Divide in New York State
            - http://jeanniechoi.journalism.cuny.edu/data-visualization-breaking-down-the-digital-divide-in-new-york-state/
        - Is Broadband Internet Access a Public Utility?
            - http://business.time.com/2013/01/09/is-broadband-internet-access-a-public-utility/

Considering the fact that internet usage is a relatively common area of interest, we had quite a few related works we could have chosen from. Unfortunately a substantial portion of the visualizations that exist are not particularly helpful such as (http://wwwnui.akamai.com/gnet/globe/index.html). Fortunately we were able to find some sites that we found helpful. While they did not have direct applicability to what we were designing, they were useful both in terms of the content covered as well as the visualizations themselves, helping to contextualize how others had thought about similar problems.

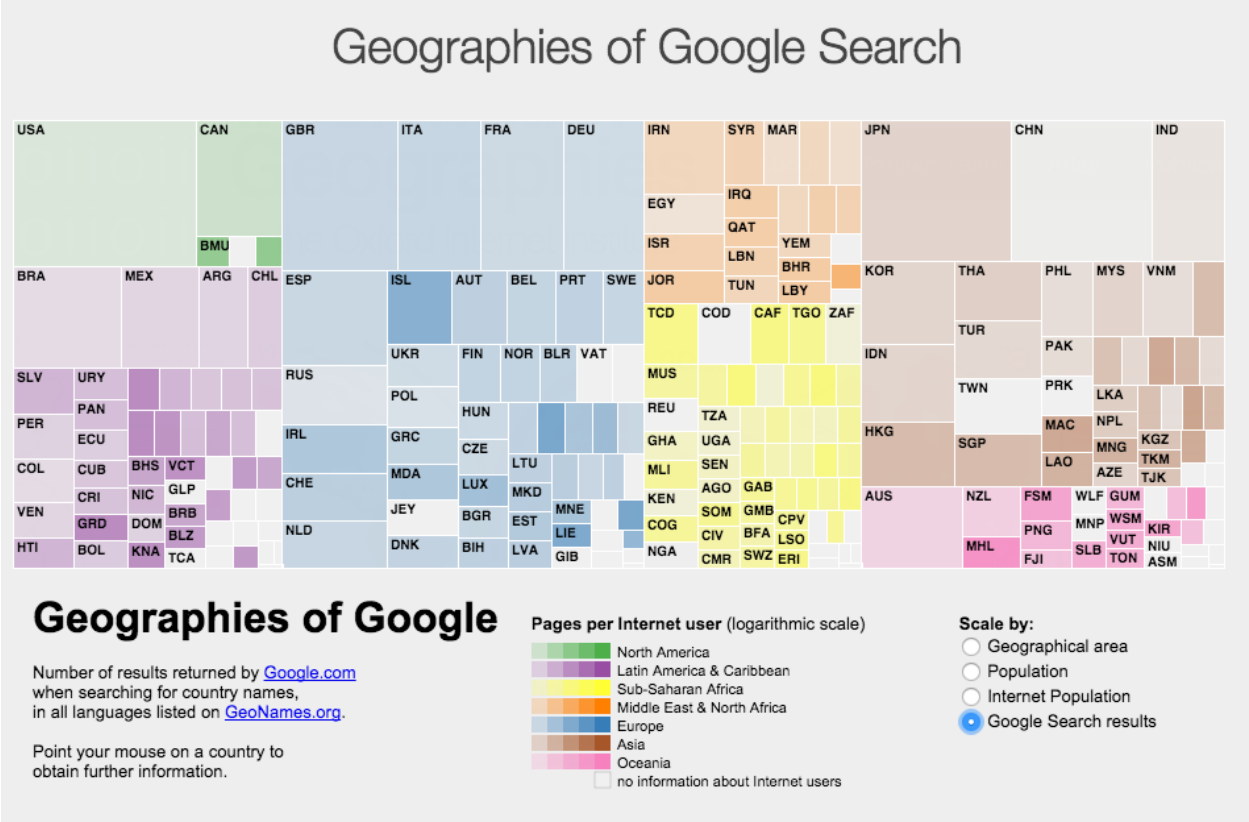


Broadbandnow.com was particularly helpful when we were trying to struggle with what information to represent and what not to. We felt that the California page (http://broadbandnow.com/California) of Broadbandnow.com did an excellent job of showing a large amount of information in a way that is conceptually quite easy to understand. One of the take aways was the the organization, having a central “take away” section as well as more detailed elements on the page. This was something that we used in our design, though instead of going from a summary to details, we went from details to summary. Additionally Broadbandnow.com did an good job of selecting important metrics such as speed and percent underserved.

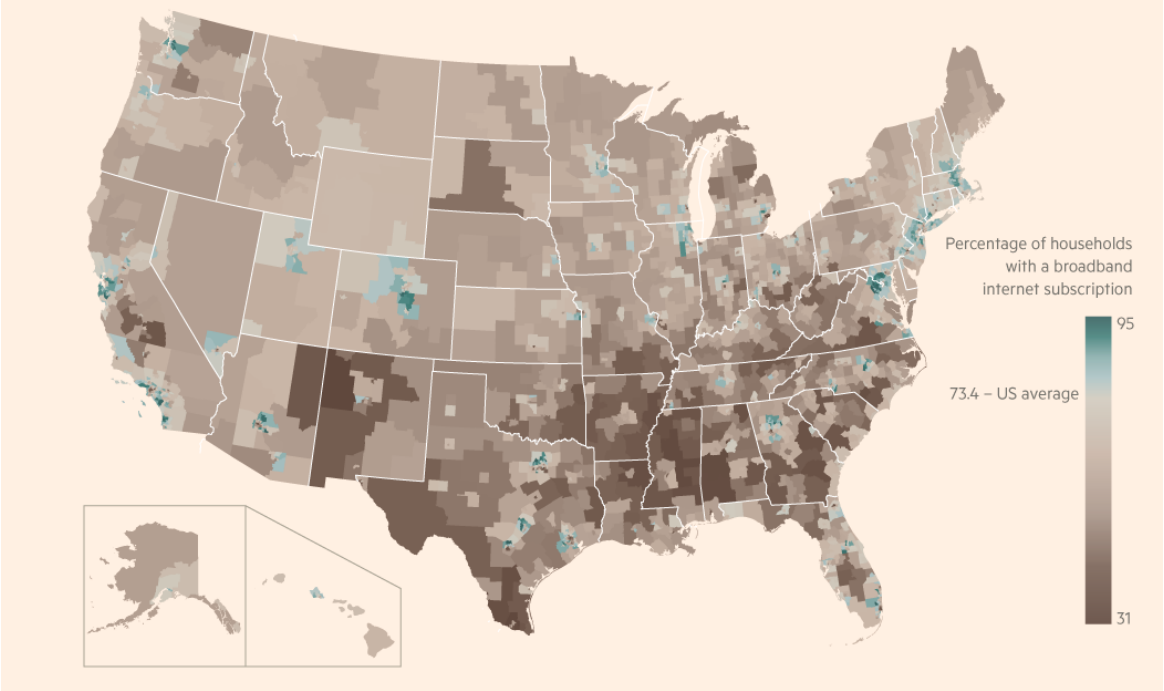




Akami State of the Internet we thought was both helpful but also illustrated the need for a more granular interpretation as well as more engaging visualizations. The benefit of Akami though is the fact that they quite a bit of data and are able to show multiple different perspectives. The interactivity of their visualization was very interesting and helped us in conceptualizing how much we wanted to restrict users. While we certainly wanted to encourage discovery, we also wanted users to view patterns in the data which is somewhat difficult to do with Akamai's site, outside of the line graph at the bottom. This demonstrated to use the need to focus on California, as originally we had considered contextualizing California within the US.

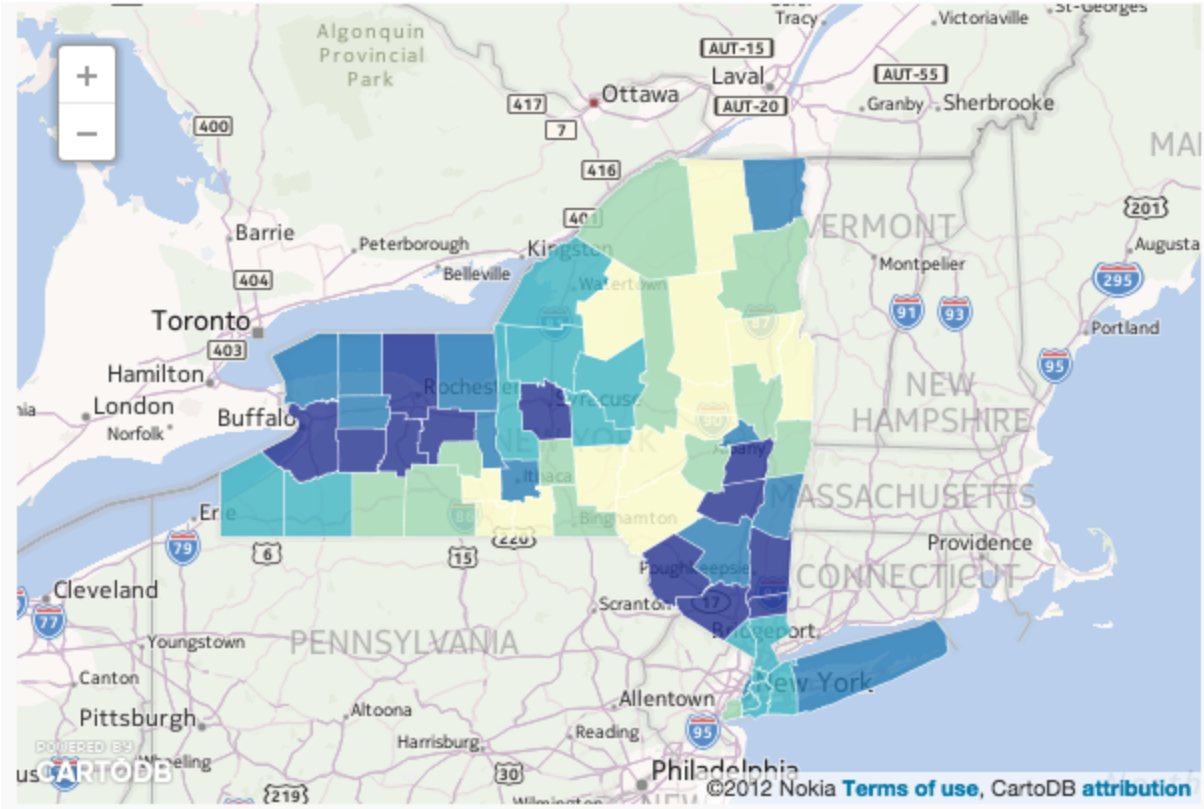


The Geographies of Google Search was quite useful in helping us understand how to effectively use the treemaps. While this visualization does not have a drill down feature, it does a very nice job of representing different aspects of Google searches based on geographic region as well as country which is similar to what we were attempting to create, with the regions within California and then the counties within each region.



Digital divide exacerbates US inequality
http://www.ft.com/intl/cms/s/2/b75d095a-5d76-11e4-9753-00144feabdc0.html?segid=0100320#axzz3WwdmUgW6

This article by the Financial Times assisted us in putting together our initial narrative. The article does a decent job with the visualization, as it highlights the percent of households with a broadband subscription quite well. The supplementary text with the article really helped us establish a framework. “The problem is most acute in urban areas where the typical cost for the most basic broadband packages is too expensive for some” helped us build a more nuanced narrative. We did feel that the map wasn’t helpful as it could have been, especially in terms of visualization and color schemes. Their color choice (blue) put an emphasis on those who already have a broadband subscription. For our visualization we decided to highlight the areas without subscription instead, as it not only helps the narrative but falls more in-line with what the user may expect to see.



Data Visualization: Breaking Down the Digital Divide in New York State
http://jeanniechoi.journalism.cuny.edu/data-visualization-breaking-down-the-digital-divide-in-new-york-state/

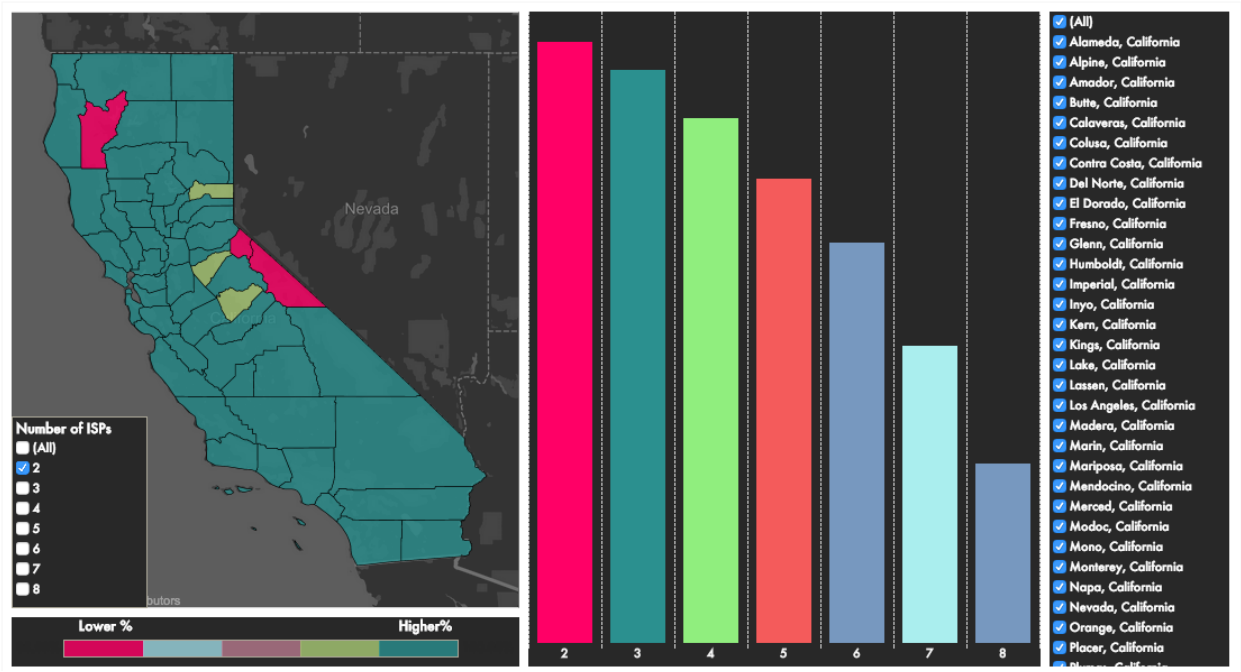
This visualization does a particularly good job using the NTIA data to map internet download speeds across the state of New York. The key aspect of this project that relates to ours is the way in which it is a specific case study of one state. As this project's focus was on download speeds and stats regarding population density, this helped us find a starting point to begin our visualizations. The visualizations in the project use multiple data sets all of which we scanned and included in our exploratory process. The website does a noticeable good job of fluidly constructing a narrative with both text and visualizations, which is something we wished to emulate with our project.

Is Broadband Internet Access a Public Utility?
http://business.time.com/2013/01/09/is-broadband-internet-access-a-public-utility/

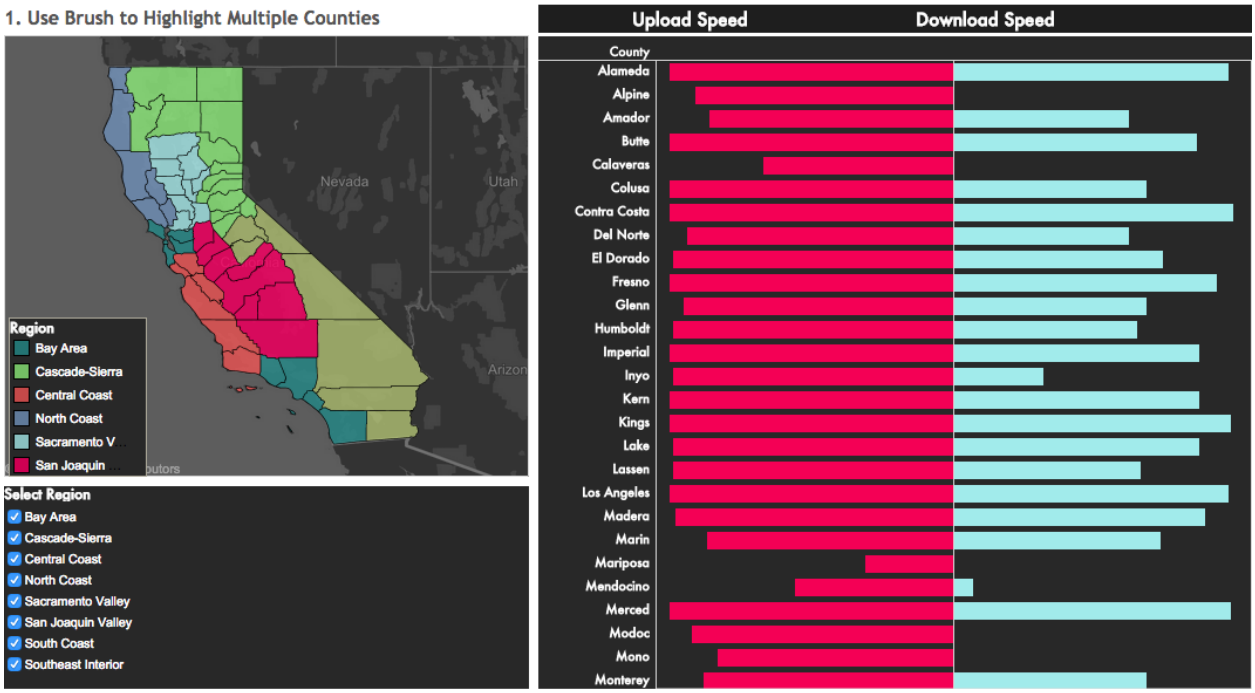
This article, alongside the book, "Beyond Broadband Access: Developing Data-Based Information Policy Strategies," truly assisted us in gaining a better understanding of the field and the complexities with our narrative. We struggled as a team to initially understand the "so what?" of our project and reading about different stances on the issue helped. These writings helped us see the spectrum of conversations occurring around broadband access. These allowed us to read about concepts that we were working around, such as how one can measure "change" and whether broadband should be seen as a public utility. Both readings helped us gain a holistic and multidimensional understanding of the subject matter which we were hoping to illustrate.

A description of the visualization, with adequate screenshots and/or illustrations

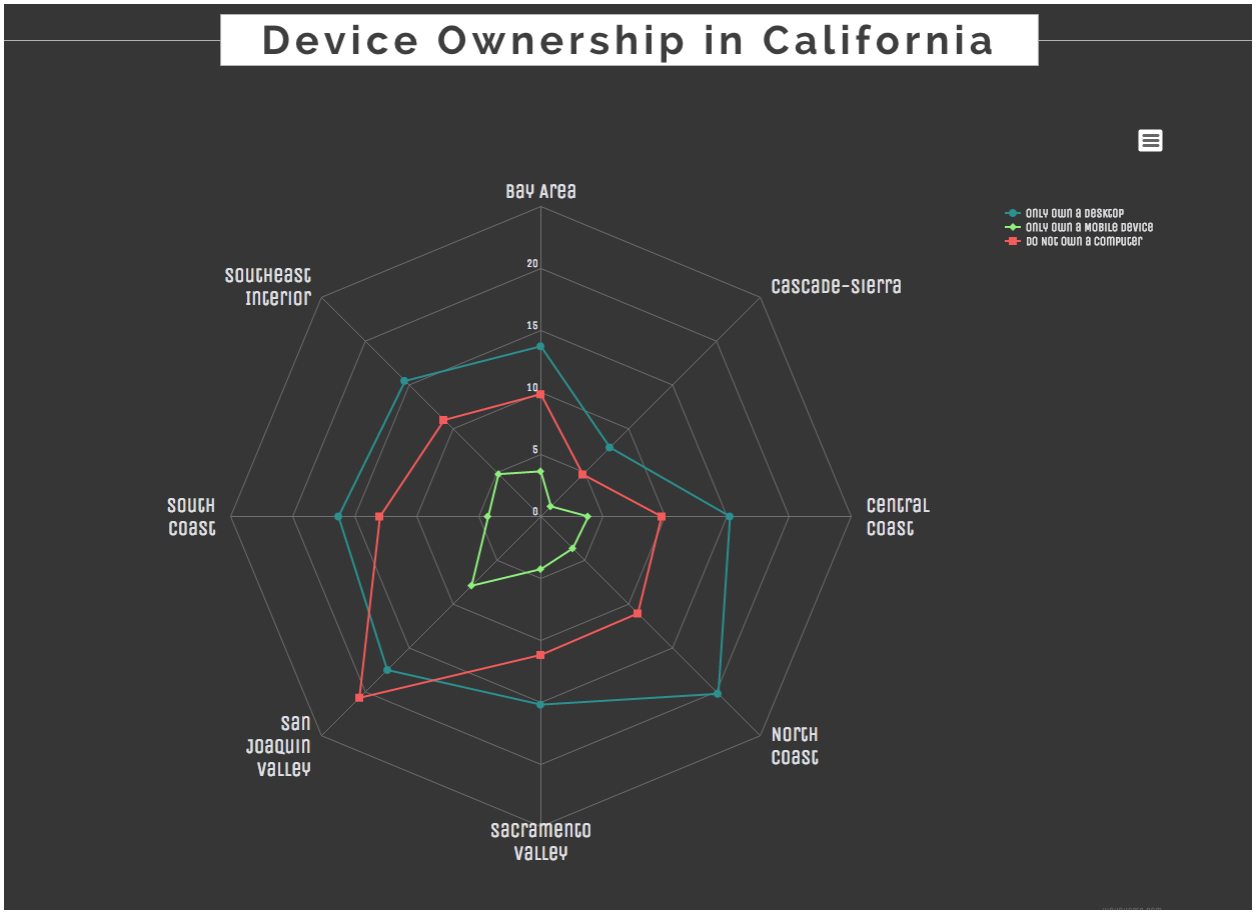
Our infographic begins with the quote, ""FCC STATES THAT 1 OUT OF 6 INTERNET USERS IN THE US DON'T HAVE BROADBAND ACCESS." We wished to start with this as it is an attention grabber and immediately sets the stage for what the website will we be focusing on. We then follow-up that quote with some quick facts about California as to narrow the scope of our discussion. After a quick framing of the issue, we wish to give the user an interactive visualization that allows for them to better understand the issues we are presenting. The first module in our project is that of "Internet Service Providers in California." We begin by quoting an article discussing the lack of competitors in the ISP space in California and its negative effects on the population.



This first visualization allows for the user to filter through the numbers of ISPs each county has access to via the map and filter. Each time the user filters through the possible number of ISPs the county has access to, the map changes colors in correspondence to the legend. If the user were to hover over any of the counties the map would display a tooltip that presents information about what percentage of the county has access to the number of ISP selected (by the filter). In addition the map is linked to a bar graph that further visualizes the specific counties' percentage of populations access to the number of ISPs (2-8).



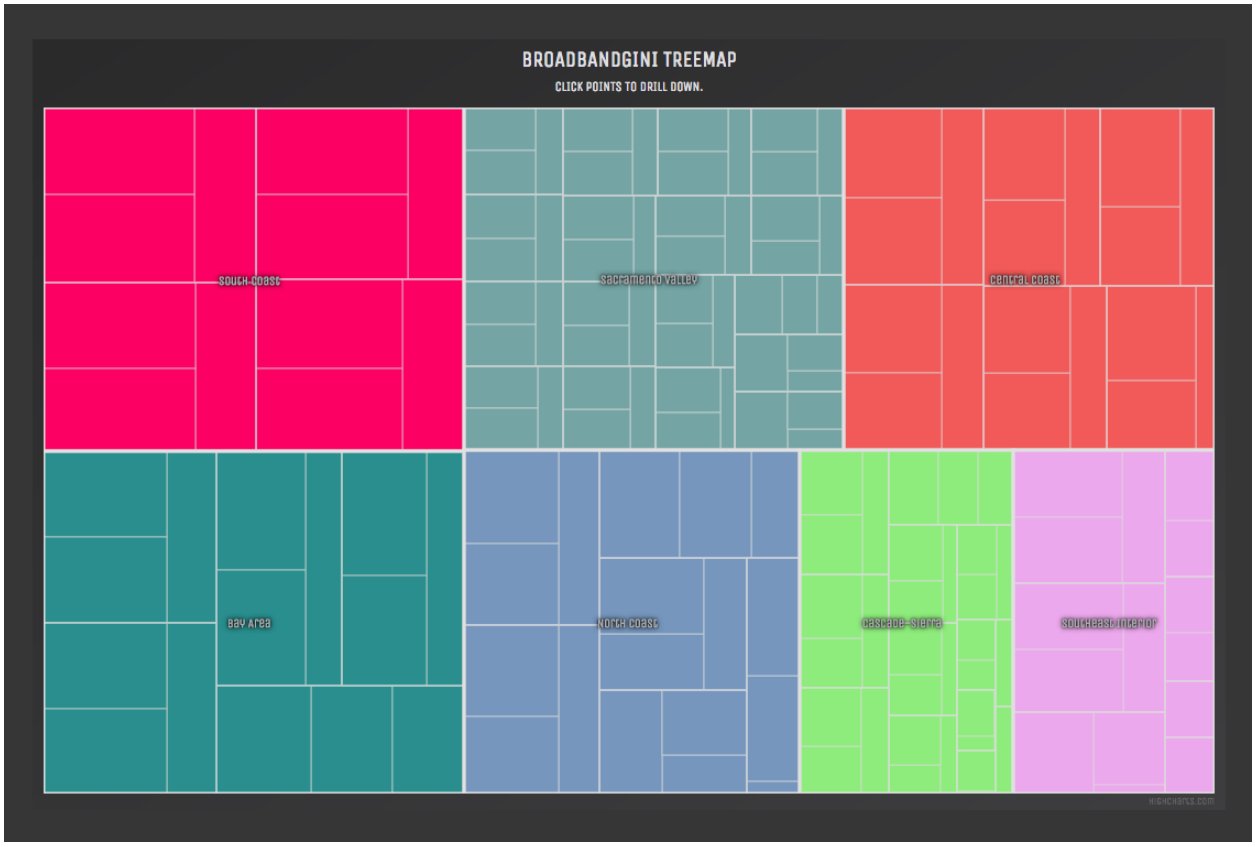
Our second module was focused around the specific upload and download speeds in California. We framed this module with some text speaking to the new definition of “broadband” in America. The new definition outlines that upload speeds are now 3 mpbs (previously 1.5 mpbs) and download speeds are 25 mpbs (previously 4 mpbs) to qualify for having broadband in the U.S. We thus created a visualization that provided the user with the power to see each countries’ upload and download speeds in relation to the new definition. This visualization was structured in a fashion that used the map as more of a navigational tool as opposed to an exploratory tool. The map was broken into various regions that allowed then allowed the user to select a cluster of states which would in turn visualize the right side of the visualization - upload and download speeds of each county. The graph displays the counties selected and what percentage of their population reach the new definition of having broadband in America.



Our third module aimed to visualize device ownership in California. The visualization is setup so that each of the regions outline the graph while the lines inside and it's proximity to the region reflects attributes about the respective region. We decided to have the user be able to see three levels of device ownership in California:

- Those who only own a desktop
- Those who own a mobile device
- Those who do not own a computer

Alongside the visualization, users may hover over any part of the graph which will then present them with a tooltip that will display those same three levels and the percentage of population that fall into that category of that respective region.



Our Fourth and final module is a treemap of a Broadband GINI we developed. The broadband GINI is calculated by adding the composite scores of the three different criteria (percent of the population in a county that meet the base level of each criteria) which is then averaged. This treemap allows the user to “drill” down in any of the regions presented before. Once the user clicks on any of the regions he/she is then presented with the counties within that region. The user then may select any of the counties and will be presented with a final layer with statistics about:

- The percentage of the county with access to 5 or more ISPs
- The percentage of the county a download speed of 25mpbs (or more)
- The percentage of county with desktop and handheld.

A region that has a high Broadband GINI will have a well balanced percentage across the three factors above which well result in a relatively even spaced graph.

What data were used to accomplish the goals

NTIA:

The National Telecommunications and Information Administration provided a wide-range of data that aided us greatly in our exploratory portion of our project. A few of the popular reports include “broadband availability urban vs rural” as well as “number of providers by speed tier”. This dataset was, although very large, extremely useful in that it provided us with a starting point. We were able to clean-up the data,extract what was useful for us and begin brainstorming and ruminating with the data. We also used this data to make our initial U.S. maps before we decided to narrow our scope to just California.

ACS:

The American Community Survey (ACS) is an ongoing survey of the American population done each year that is designed to help policymakers understand different communities so that they can plan for the future. The data from the was critical in understanding how what technologies different communities use to access the internet as it gave is detailed and reliable information across each county in California. Yet while this data was quite helpful, it was difficult to use as each county was encoded with a specific number and the geographical key was in a very messy csv file which required some scripting to clean up. Once we had the key set though it was quite easy to work with the other data.

Broadbandmap.gov

The National Broadband Map (NBM) is a tool to search, analyze and map broadband availability across the United States. The NBM data set - something created by the National Telecommunications and Information Administration (NTIA) and FCC - greatly assisted us in getting data on upload and download speeds in all of the counties across California. Although there were a few geographical and name errors in the data it was a primary source for our visualizations.

Which tools were used to accomplish the goals

Tableau

We used Tableau for its strengths in being able to take big sets of data, combining them and quickly creating visualizations for exploratory purposes. Once we sketched out what we wanted, it was a matter of time before we could make them work in Tableau. Tableau offered the ability to make custom color schemes, which we felt aided us in our goal of creating a pleasant user experience. Tableau was also great in helping us map our data quickly so that we could begin to cluster regions and further understand and validate our research and findings.

Highcharts

We used highcharts because it allowed us greater flexibility when building more exotic visualizations such as the radar graph and the treemap. We initially considered using Tableau for these as well, but Tableau is poorly suited to making radar graphs and we were having difficulty implementing a drill down feature as well. Subsequently we decided to use

highcharts which made implementing these elements quite straight forward, allowing us to focus on the actual visualization as opposed to fighting the program.

What steps were required to accomplish goals

Our process was most likely quite similar to other groups and was essentially a mixture of EDA as well as loose narrative construction, though with the caveat of redesigning the visualizations. This broke down to 7 major sections: data collection and processing, EDA, discussion, rough draft, discussion, redesign and composition of our site. Each of these stages built upon the last, allowing us to represent a more focused version of what we were trying to communicate.

Data collection was a relatively straight forward process as there are a number of different agencies and groups that measure internet activity. Ultimately we decided to focus on government sources as we felt they would be the most reliable and would be least likely to have data skewed data. Once we had the data though we need to go through and select the relevant information as we actually received more than we needed. Once this was done we began our initial analysis.

We conducted exploratory data analysis (EDA) on the data once we had selected what we needed. We had a number of different hypothesis that we wanted to explore such as whether rural populations were more disenfranchised or whether GINI was an effective tool for predicting a significant portion of the population being disconnected. In order to accomplish the EDA quickly we used Tableau. This was especially effective for us as we planned to use Tableau for part of our final visualizations as well, meaning that we could use some of what we found as a rough draft.

As we went through the EDA, we were able to start dropping different points of inquiry, such as the population density and GINI. This lead to discussions regarding what we wanted to represent and how we wanted to represent them. Initially we thought some of the interesting elements would be things such as percent of the population that has computer vs. percent of the population who has a computer but without internet. Additionally we thought about looking at distribution of broadband access by age as well as upload and download speeds by county. These elements found there way into our initial rough draft.

The rough draft was built entirely in Tableau and was focused on displaying the information that we thought was valuable at the time. We also wanted to set up the initial framework for the website that we were planning on using. This was very helpful as we were able to identify unexpected problems with the framework we were planning on using and solve them. Additionally we were able to get feedback from Marti which showed us that we were taking far too lose an approach and were not giving the user a “so what”. This led us to regroup and discuss exactly what we were trying to accomplish with the visualization.

From there we talked more about what the minimum message we wanted to convey was. We felt that the ultimate take away we wanted to communicate was that internet access is not as simple as “is broadband available”. From this we started to look at what the core elements of being meaningfully connected were and we were able to decide on our final 3 criteria: ISP choice, speed and access to multiple devices. From there we were able to redesign our visualizations.

Once we had our three criteria, we went about redesigning the visualizations to support user exploration of these elements. We had taken the comments about choropleths to heart and wanted to make sure that we did not overuse them, but we did feel they could

contribute to the story so we worked hard to incorporate one, supported by a bar graph. We also worked hard to incorporate other type of visualizations in order to both give the user variety as well as leverage the best type of visualization for each task. In order to ensure that we were on the right path we talked to people close to use who were not involved with the project to see if they were able to understand what we were trying to communicate.

Finally we put everything together in on the website. This we initially considered to be trivial but proved to be more significant than we expected as we wanted to contextualize the visualizations as much as possible. We worked on including additional facts that would help the user understand some of the key assumptions the visualizations made, as well as include interesting facts to demonstrate that this is an important issue. Additionally we suggested other materials the user may be interested in if they wanted to learn more about the subject.

What kinds of results you obtained, with a focus on usability tests or responses from prospective or real users

Usability Tests: Our initial usability tests in regards to our final product were somewhat minimal as we made a lot of last minute pivots and were thus only able to test our website on a few users that were able to meet us in the eleventh hour.

Prospective users: We are pleased to see that many of the prospective users at the showcase gave us positive feedback.

Points of refinement: Our initial usability tests and responses from prospective and real users gave us the following results:

Overall:

- We need to refine our color palette
- Some suggestions were to use a stepped color module.
- Others suggested reducing the “loudness” of some colors to match the other colors.
- Action item: We will experiment with the stepped color palette but will be sure to inverse the color scheme so that the pink/red color we have corresponds to lack there of.
- We need headers/titles in certain areas:
- We will add the word, “percentage” to various areas in our website - particularly module 3.
- We will have titles for both module 1 and 2 describing the scale and measurements of the bar graphs.

Module 1

- We will add radio select instead of current filter as user only needs to see one map version at a time.

Module 2:

- Rethink color palette and inverse color scheme.
- Be sure to add a header to the bar graph to show what each side of the bar graph is “reaching”

Module 3

- Icons in legend area will help the user understand the different lines.
- We will add additional descriptive text to frame the graph.

Module 4

- Redesign the color palette and text selection
- Add framing to this module and rethink how information is presented once the user "drills down."
- Possibly use the tool tip as a navigation method.
- Add scale for each of the different percentages and sizes.
- Possibly show/describe what will happen in the next level down before the user proceeds.

Links to demos, documents, or whatever is needed to show the visualization

- http://broadbandgini.info

A table showing which parts were done by each student, and in what proportion approximately

Name	Data Collection	EDA	Discussion	Rough Draft	Discussion	Rebuild	Compose Site
Bryan	60%	40%	50%	50%	50%	50%	50%
Hasnain	40%	60%	50%	50%	50%	50%	50%

A small thumbnail image (100x100 pixels) to be used to illustrate your visualization on the course final project web pages

- https://files.slack.com/files-pri/T02P95UBX-F04S5CU10/logo.png

Software created, to the degree this is possible.

Website Login:

User: info247
PW: finalproject1234

Appendix:

Extra Resources:

- http://www.pewinternet.org/2010/08/11/home-broadband-2010/
- http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/
- http://www.pewinternet.org/2013/08/27/broadband-and-smartphone-adoption-demographics/
- http://www.nytimes.com/2014/10/31/upshot/why-the-us-has-fallen-behind-in-internet-speed-and-affordability.html?abt=0002&abg=0
- https://www.newamerica.org/oti/the-cost-of-connectivity-2014/
- http://broadbandnow.com/California