David Ly
CIS 320
Winter 2016

Computer Ownership and Internet Access across Different Characteristics

Objective

Computer and internet use have become an integral part of our academic, professional and social life. As students who grew up alongside the advent of the "internet" in the 1990s, we have developed a curiosity of how this revolutionary new technology affected our generation, those that came before us, and those that came after us. However, before we even begin to answer the question of how, we must analyze how well people of different characteristics have adapted to computers and the internet. This brings up the question: how does computer and internet ownership and access differ amongst people in different social, economic, and geographic statuses. Ultimately, the scope of this investigation is to analyze data of individuals who do or don't have ownership along with access to a computer and the World Wide Web.

Background Information

Internet and computer access has become increasingly important in the 21st century that some may even call it an essential utility comparable to that of water, electricity or gas. We rely on it to connect us to others and its ability to access the treasure trove of data generated by the billions of electronic devices out there. The U.S Census Bureau has begun surveying about computer and internet use since the early 2000s through the use of the American Community Survey (ACS) to measure and monitor nationwide development of broadband technologies. In turn, state and local government may use this data to help address areas of low computer and internet access by focusing resources on areas deprived of the service.

Research Questions

We begin this project with assumptions that income, education, and age are probably key factors when looking at computer and internet trends. It is not our intention to make inflammatory statements but there is a general consensus that the wealthy (income), the intelligent (education), and the young (age) are more likely to have these privileged resources at the palm of their hands. In short, it is our expectation that they more readily have access to these technologies. At the end of this project we hope to have the appropriate visuals, numbers, and stories to answer the following questions and see what other factors contribute to computer ownership and internet access.

- Is computer and internet access becoming more accessible as time goes on?
- Does income affect computer ownership or internet access?
- Does age affect computer ownership or internet access?
- Does education affect computer ownership or internet access?
- Does gender affect computer ownership or internet access?

References to Data Source

We will be accessing four excel files on census.gov. Each excel file charts different measurable characteristics. Amongst the characteristics are: year of survey, age, race, education, state, region, gender, income, and employment status. Although most of the files are conveniently in excel format, we will at times draw from pdf files. During the initial extract-

transform-load process we have found that the pdf to excel conversion isn't as clean as we would like it; so there may be times where we will have to manually copy and paste sections at a time to recreate the table in Microsoft Excel.

- www.census.gov/hhes/computer/files/2012/table1.xls
- www.census.gov/hhes/computer/files/2012/table2.xls
- www.census.gov/hhes/computer/files/2012/table3.xls
- www.census.gov/hhes/computer/files/2012/table4.xls
- http://www.census.gov/content/dam/Census/library/publications/2014/acs/acs-28.pdf
- http://www.pewinternet.org/files/oldmedia//Files/Reports/2013/PIP_CellInternetUse2013.pdf

Data Cleansing

Now that the objective and research questions of this project are defined to guide us, we will document our progress starting with our first dive into the data. You would think the extract, transform, and load phase to be as simple as textbooks describe it to be, but it was soon apparent it was not. Although the data was charted in excel files with meaningful labels, there were often missing values, invalid entries, and miscellaneous notes scattered across random cells and random sheets. This part of the investigation helps in understanding the data as you attempt to organize it into sets of data that may be related to each other. As we clean up the data and move things around, we catch glimpses that reinforce our current understanding as well as facts that challenge existing beliefs.

First Glance

Using SPSS Statistics we generate the output for our first data set. This data set includes the states and how many households in that state have internet access in percentage value. The descriptive mean tells us that 79.62% of the United States has access to internet as of 2012. This can be transformed into a statement such as: nearly 8 out of 10 households in the United States have access to the internet. Next up, we will look at the maximum value of 87.9% (Oregon) versus our lowest value of 64.8 %(Mississippi). The range or difference between the highest and lowest value is 23.1% which is quite a huge difference. I couldn't imagine living without the internet so this variance across states is worth investigating. I would also like to take this moment to say that that aim of this report is not to argue that this caused that; rather it will only analyze the given data sets through a variety of tools and merely speculate why it is the way it is.

Descriptives

[DataSet1]

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
State with internet access	52	23.1	64.8	87.9	79.625	5.2961	28.048	-.567	.330	-.030	.650
Valid N (listwise)	52										

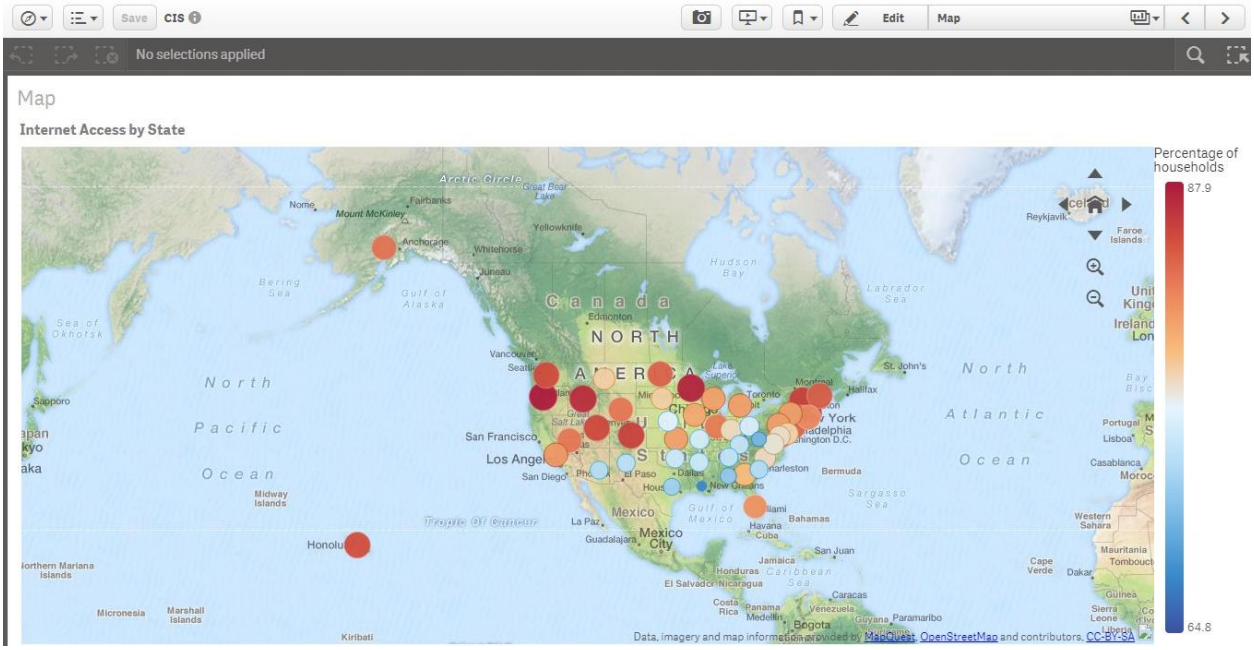
Next we tried to reproduce the above in 'R'. We would load the data into R using the read.csv command. Although you can extract it directly from a website, data sets may sometimes be unclean; and so this particular demonstration will use the 'file.choose' feature to grab it locally from the desktop. The 'summary (data)' command crunches through the data and tells us everything that SPSS told us and more. We try the'table (data)' command for the next several sets to produce a semi structured understanding of the data set. Although it is in the form of pivot table/cross tabulation, there is much more to be discovered. At this point, we can manipulate the output using operations to give us better insights. We were able to answer basic questions like the number of observations, proportions of a whole, and 'R' was able to distinguish between different categories and data types.

```
> data <-read.csv(file.choose(), header=T)
> summary(data)
        State    State.with.internet.access    X             X.1
           : 2   Min.   :64.80              Mode:logical   Mode:logical
 Alabama   : 1   1st Qu.:75.50              NA's:53        NA's:53
 Alaska    : 1   Median :80.70
 Arizona   : 1   Mean   :79.63
 Arkansas  : 1   3rd Qu.:83.85
 California: 1   Max.   :87.90
 (Other)   :46   NA's   :2
> |
```

```
> data <- read.csv(file.choose(), header=T)
> table(data)
      Internet
Gender   N   Y
     F 107 433
     M  98 381
> table(Gender)/length(Gender)
Gender
        F         M
0.5299313 0.4700687
> table(Internet)/length(Internet)
Internet
        N         Y
0.2011776 0.7988224
```

While we are on the topic of states and regions, a heat map of the above data will be included to better illustrate my next points. Perhaps age, income, or education had nothing to do with computer and internet access. Or perhaps the data point was an outlier. Using QlikSense's geo mapping, we graphed all 52 states and their associated percentage value in hopes of a more complete view of the surrounding area. Our results indicate that the west, the east and the north had the highest percentages of households with internet and computer access, while the southern region held the least amount of households. Not surprisingly, Oregon, who maintains the highest percentage, fell in the west; while Mississippi, who holds the lowest percentage, resided in the south. After looking at the age, income and education of the population in that region and comparing it to the rest of the country, we now know that being in that region doesn't automatically diminish your chance at computer ownership or internet access. Rather, it is a mix of many characteristics that put them below the national average.

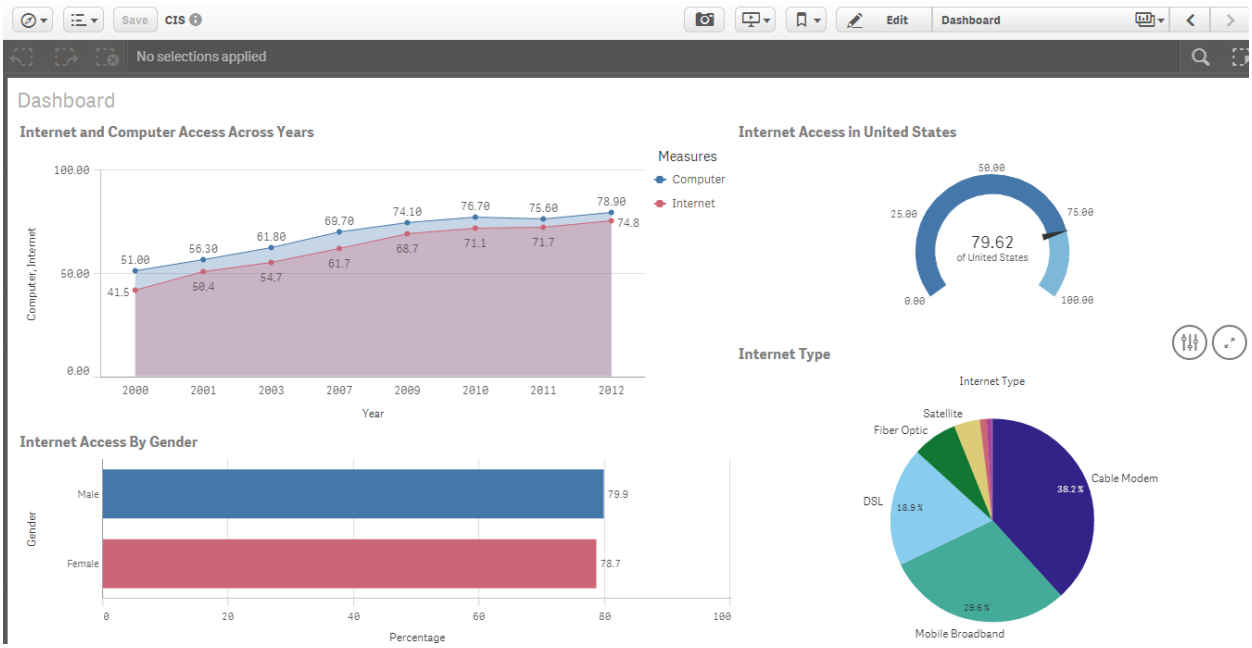
David Ly
CIS 320
Winter 2016



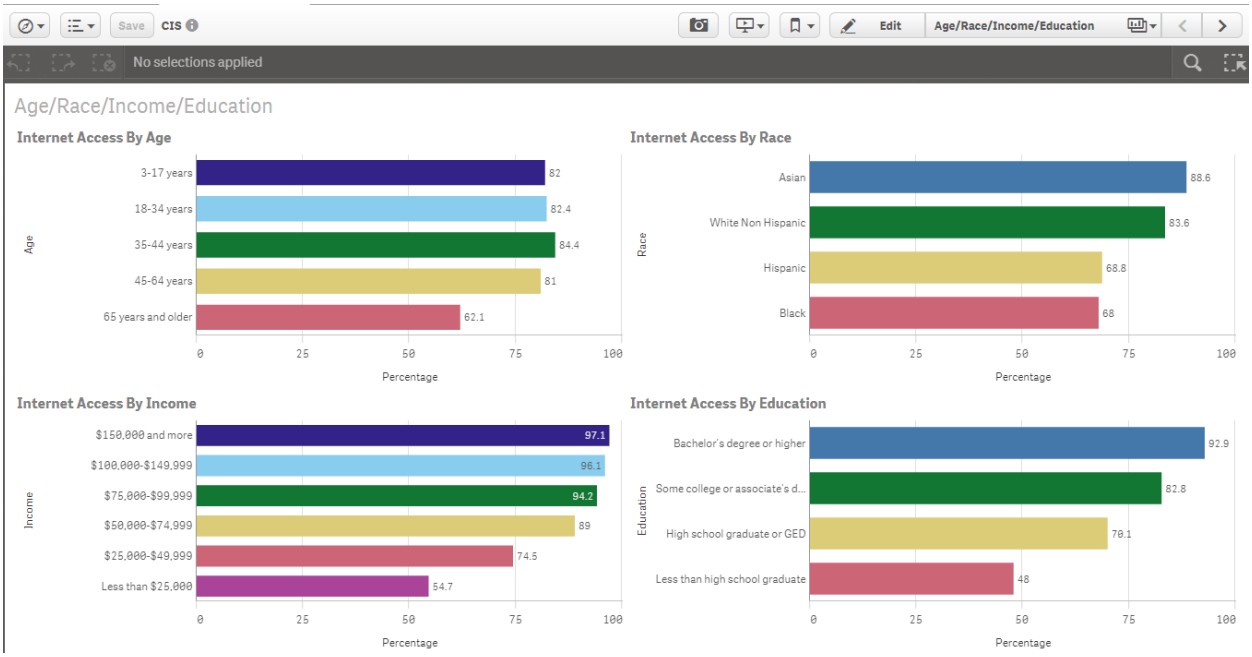
Key Highlights

An analysis of computer ownership and internet access wouldn't be complete without analyzing the gender component of it. Surprisingly, both male and females reported similar results. While 79.9% of males indicate that they own a computer and have access to the internet, females were not far behind with 78.7%. Media portrays the geeky male nerd to be at the computer conducting their own business while the well manner house wives tend to the house. This stereotype is further reinforced when we see little to no females in CIS classes and the ones that we do come across aren't mind-blowingly skilled when using it. We then begin to question if they are at all familiar with the internet or even have access to it. The data shown here indicate that both genders are fairly equal when it comes to access of it which come to debunk our beliefs.

Our next visual helps us understand internet access and computer ownership across a given time period. At first glance we can see not everyone who owns a computer or similar device will have an internet connection, but it is dead close (78.9% versus 74.9% in 2012). We can also see that the line representing computer is higher than that of internet, which is logical since what is the point of an internet connection if you don't even have a device in the first place. Even with incomplete data we can safely say that that the two lines are growing rapidly from 51% and 41.5% in 2000 to 78.9% and 74.9% in 2012 for computer and internet ownership respectively. It is on the same graph that we can see the area between the two curves is lessening. To put this into words, you can say that those who own a computer are almost always going to have an internet connection too. This data may also prove to be of use in the commercial environment seeing as how these two products and services share such a solid relationship. Perhaps two industries giants may form a partnership and refer customers after they buy this or that, which is similar to Amazon's "people also bought this" advertisements.



Our final visual takes a look at the how internet access differs across age, and income, and education. The age graph depicts that the age group from three to the early sixties are steady at 80%. We then experience a sharp decline when we enter the age group of sixty five and older. The income graph shows a similar story when the lower income households are less likely to have internet access. Only when it hits the $75,000 income bracket does all subsequent brackets are at a consistent 90%. We see this story again in the education graph as a higher education level reflects a higher percentage of internet access in the household. You may be thinking, "This was obvious; you took three pages to come to that conclusion?" I would have had the same reply; the average person would have had some idea of this conclusion but the data science process allows us to confirm and see relative to where each assumption stands relative to another.



David Ly
CIS 320
Winter 2016

Final Thoughts

Research questions are not always there to guide us in the real world and good data scientists are able to make sense and uncover meaningful insights with any data set. After looking at the visuals, it is clear that there are areas of low internet access across many social characteristics. As stated in the introduction, technology is an increasingly integral part of our life. Those that are less fortunate and unable to access information or connect with others will be undoubtedly worse off in life than those that do have access. By exposing individuals and households to the internet, we take a step forward in educating them and closing the gap between those that are better off and those that are less fortunate.