ISyE 2028 – Basic Statistical Methods - Fall 2015
Bonus Project: "Big" Data Analytics
Final Report

Since the late 19th and early 20th centuries, television has become increasingly popular. Since its inception, the concept of television has been constantly evolving. What started off as transferring a still picture through a wire turned into broadcasting motion pictures through the use of cables; cables eventually lost popularity when satellites became able to broadcast TV shows and movies at increasing locations. With the invention of the internet, the concept of television has expanded even further. Now, many companies such as Netflix and Hulu have taken advantage of the versatility of the internet in order to provide a larger percent of the population with a different way to watch TV shows and movies. The evolution of how people choose to watch their shows caused me to investigate whether people prefer to watch shows on the internet or to watch shows through more traditional methods such as cable and satellite.

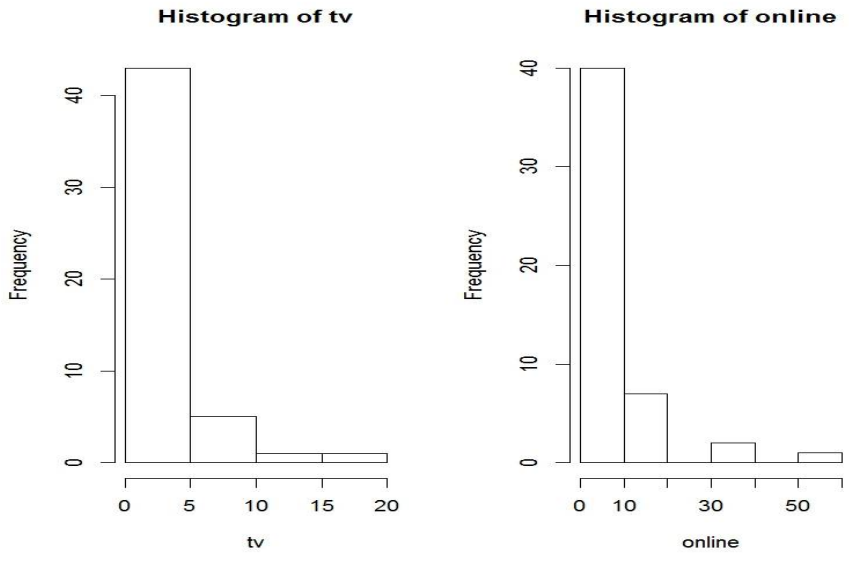

Figure 1

To begin, I needed to collect data from students as Georgia Tech. I chose to collect data through an online poll posted on Georgia Tech Facebook pages and physical surveys around campus. I used two forms of data collection in order to account for the portion of the population that does not spend a great deal of time on the internet. After I polled the Georgia Tech students, I had a sample size of 50. Approximately half of the sample consisted of results found through an online survey; the other half consisted of physical surveys taken around campus. The sample size was large enough to assume a normal distribution, but the histogram for the sample was extremely skewed as you can see in Figure 1.

In order to analyze the results of the poll, I found the numerical summary of the hours spent watching shows online and the numerical summary of the hours spent watching shows using the traditional methods. I did so using R. The results are shown below. Figure 2a shows the summary for the hours spent watching shows using a traditional method. Figure 2b shows the summary for the hours spent watching shows online.

```
> summary(tv)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.00    0.00    1.00    2.22    2.00   20.00
```

Figure 2a

```
> summary(online)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.00    1.25    5.00    7.92    9.50   59.00
```

Figure 2b

Using the same program, I created a boxplot for the hours spent watching shows online and a boxplot for the hours spent watching shows using traditional methods. The boxplots can be seen in Figure 3. The numerical summaries of the data shows that on average, the number of hours spent watching shows online is larger than the number of hours spent watching shows using a traditional method. The boxplot shows a similar result. The boxplot, however, reveals that there are outliers in both sets of data. To verify these results, I also created a ninety-five percent confidence interval for both sets of data using R. Because I will be used the sample standard deviation, I used the Student's t-distribution in order to find the population mean instead of the normal distribution. The confidence for the online hours was from 4.84 to 11.00 hours; the interval for the traditional hours was from 1.05 to 3.39

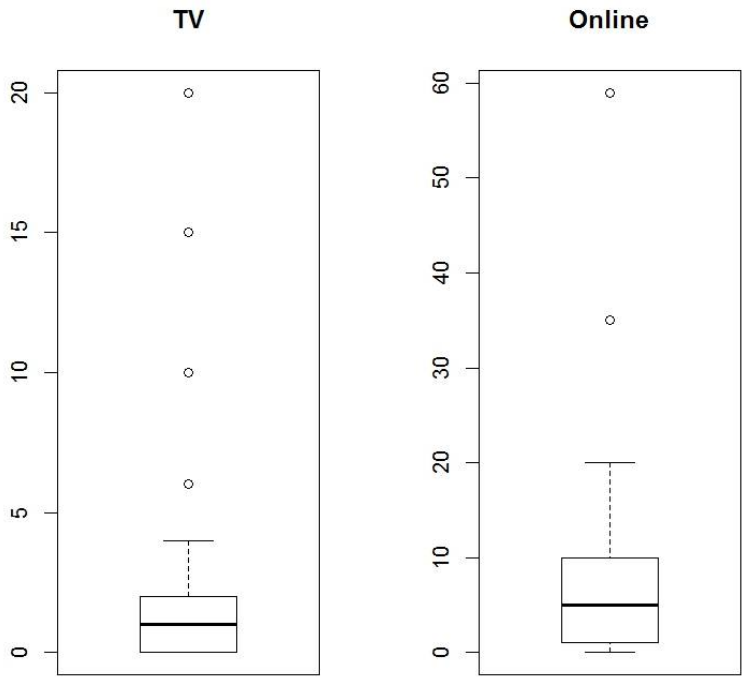

Figure 3

hours. It is clear from this data that the number of hours spent watching shows online is higher than the hours spent watching shows using traditional methods.

My hypothesis was that the number of hours spent watching shows on television will be substantially lower than the number of hours spent watching shows online. Though my results show that the hours spent watching shows on television is lower than the number of hours spent watching shows online, the results were not as extreme as I expected. This may be because I had a small sample size or because my survey focused solely on Georgia Tech students. Because I collected data from a technical school, my results are also biased. The results could be improved if this experiment was done on a larger scale. If this experiment had been done on a larger scale, the results could be used in the world of marketing. If companies knew which source of entertainment was the most popular, they could fund marketing in the right way.