OREGON STATE UNIVERSITY

ST 314

SUMMER 2019

Data Analysis Three
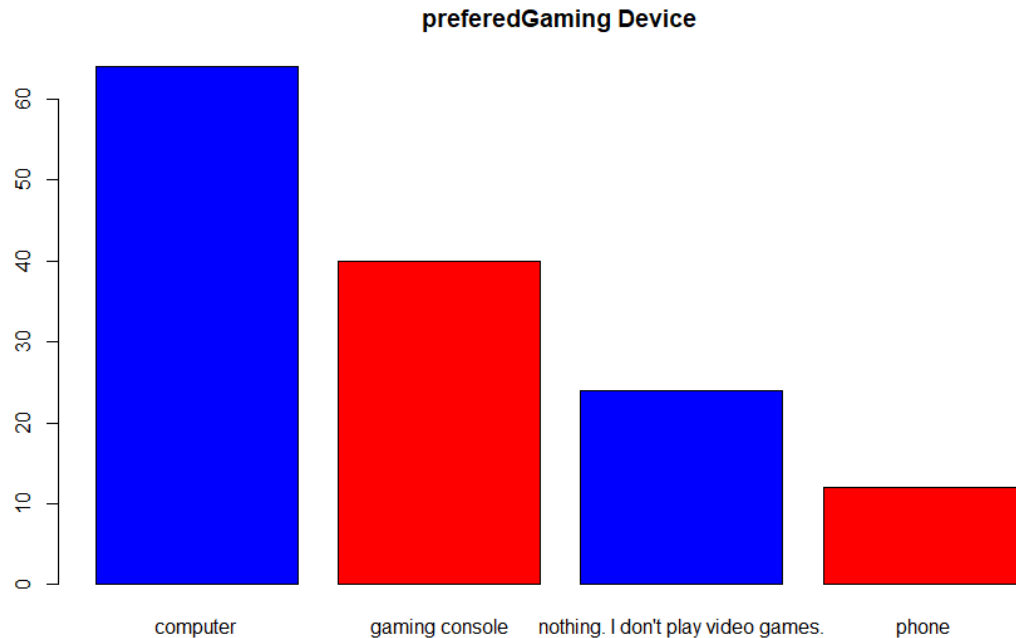
Author:
Thomas Noelcke

Instructor:
Katie Jager

I. PART I

A. Categorical Variable

The variable that I chose was the gaming device variable. I chose this variable because I'm interested in what proportion of people in this class are using which gaming devices. Below is a plot of the different gaming devices that people in this class are using.

preferedGaming Device

| Device	Count
Computer	64
Console	40
I don't Game	24
Phone	12

The above plot once sorted from highest to lowest generally looks like it is linear in distribution. The gaming devices listed in order by popularity are, computer at 64 people , the console at 40, I don't game at 24, and lastly phone at 12. I was surprised by the number of people who don't game at all in comparison to the number of people that game on their phone. I thought that this would have been flip flopped.

B. Quantitative Variable

For my second variable I chose Gaming time in hours. I chose this variable because it complimented the categorical variable that I chose earlier. Below are images of the plots requested.

ST314 Summer 2019 Students: Gaming Hours per week

ST314 Summer 2019 Students: Gaming Hours per week
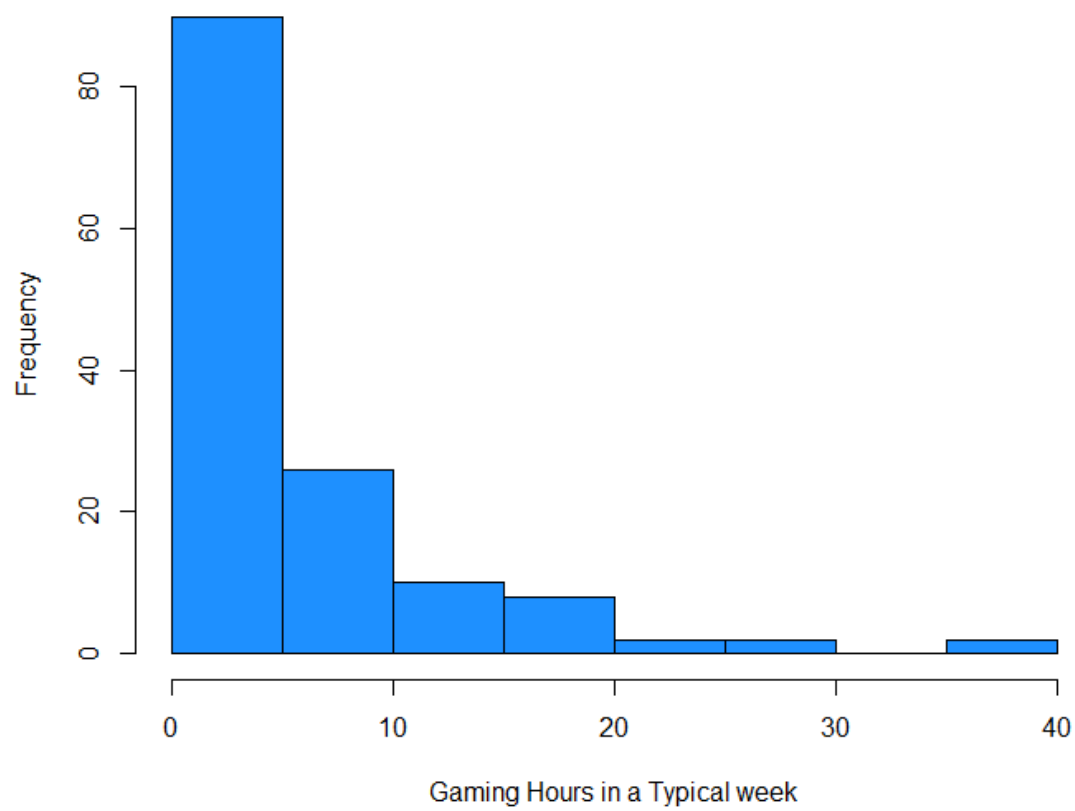

Gaming Hours in a Typical Week
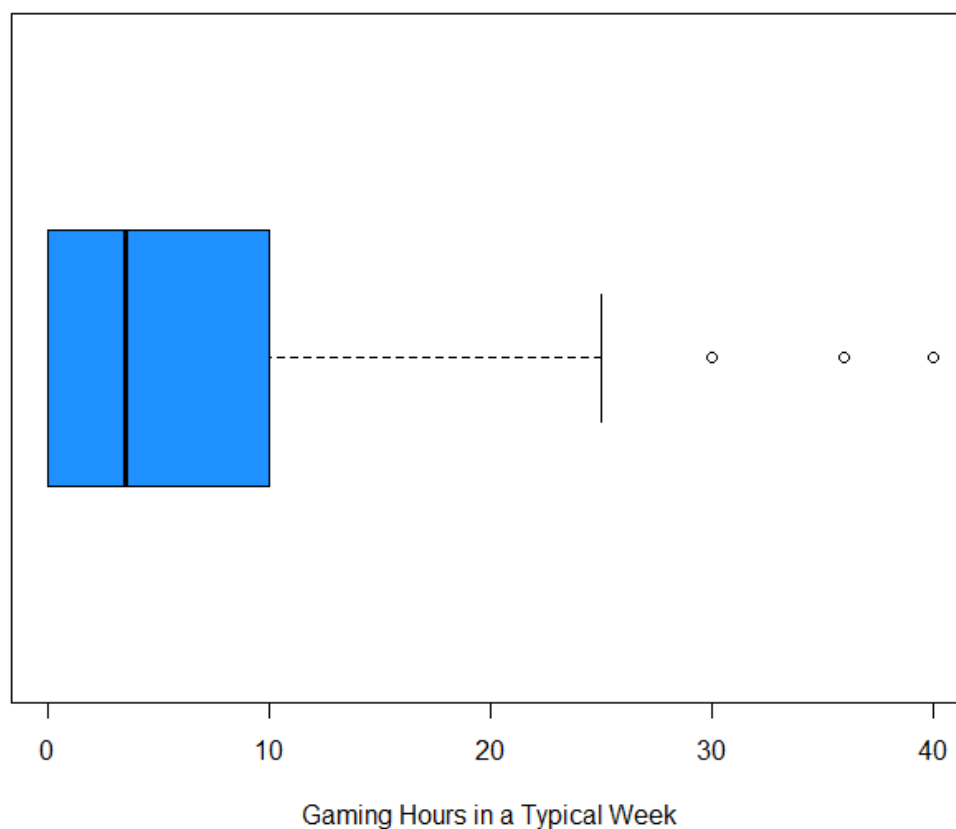
I prefer the Histogram to visualize the data. I prefer this plot because it is more intuitive for me. It is also easier for me to tell which direction the plot is skewed if any.

Name	Value
mean	6.125
standard deviation	7.768
Minimum	0.000
1st quartile	0.000
median	3.500
3rd quartile	10.000
maximum	40.000
IQR	10.000

With the given statistics and plots above we can say that the distribution of gaming hours is positively skewed. This is because the mean is much larger than the median but also because of the shape of the histogram. There were three outliers in this set of data that were between 25 and 40 hours. Given that we have skewed data with outliers that the median would be the preferred measurement for the center of the data. This is because the mean will be skewed higher due to the outliers in the data set.

II. PART 2

ST314 Sample: Gaming Hours Per Week
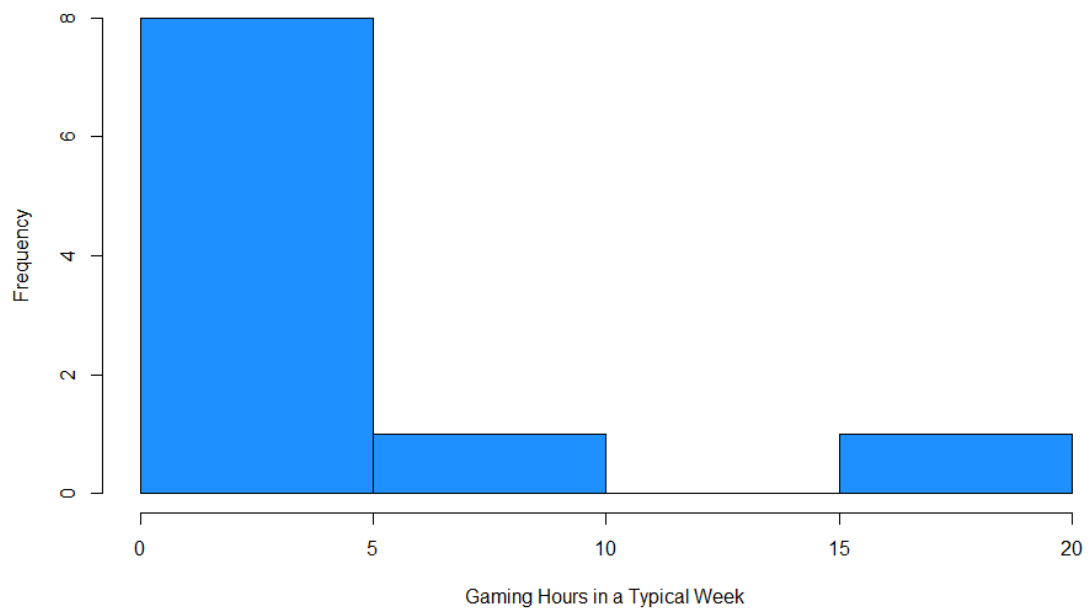
ST314 Sample: Gaming Hours Per Week
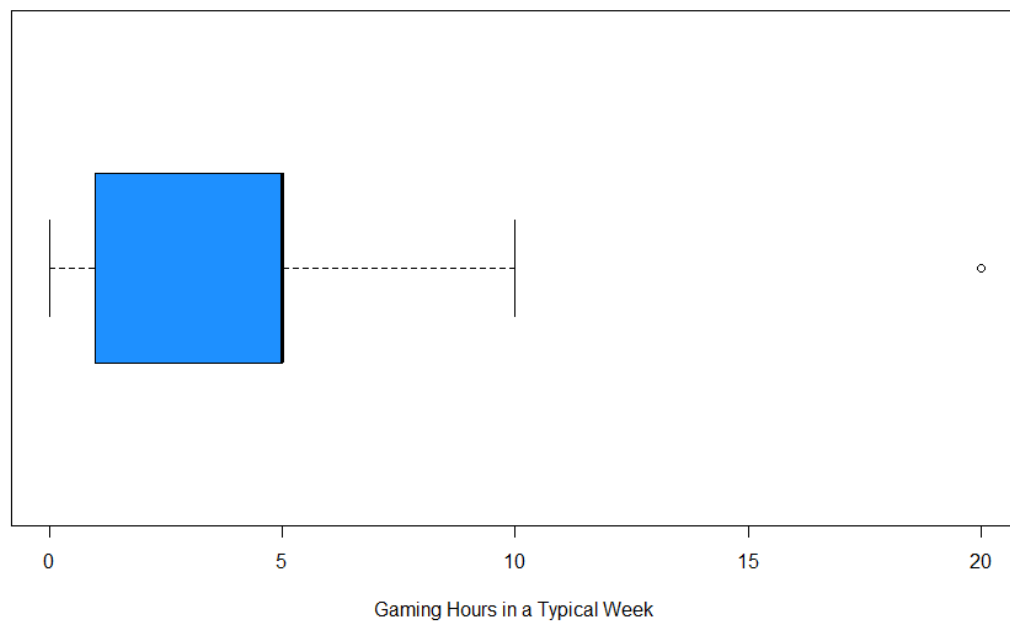
Name	Value
mean	5.6
standard Deviation	5.892
median	5.0

My sample values when compared to the population values were very different. The plot appeared to be positively skewed but looking closely and the mean and median they were fairly close for this sample. The box plot revealed that the mean was right at the top of the IQR indicating that the mean was skewed by an outlier. This is similar to the results for the population mean however the sample median and population median were very different. The standard deviation for the sample also differed from the population standard deviation.