**Homework 1**

**Name: Sushant Kotwal**

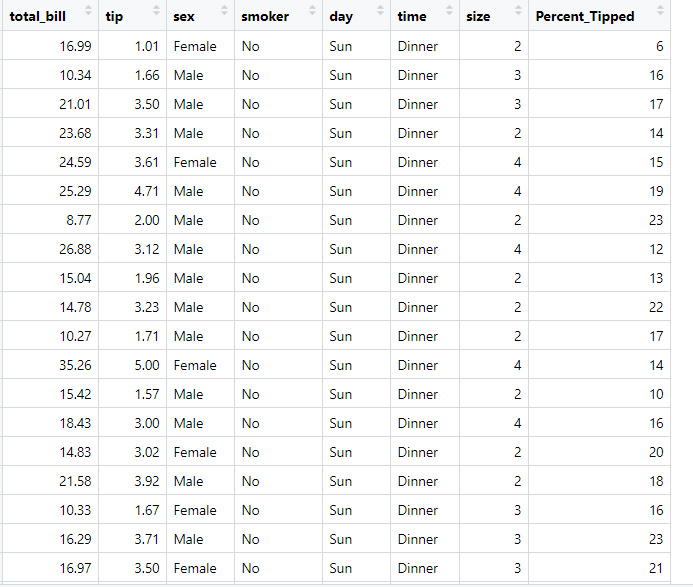
**Username: skotwal**

**University ID: 2000773474**

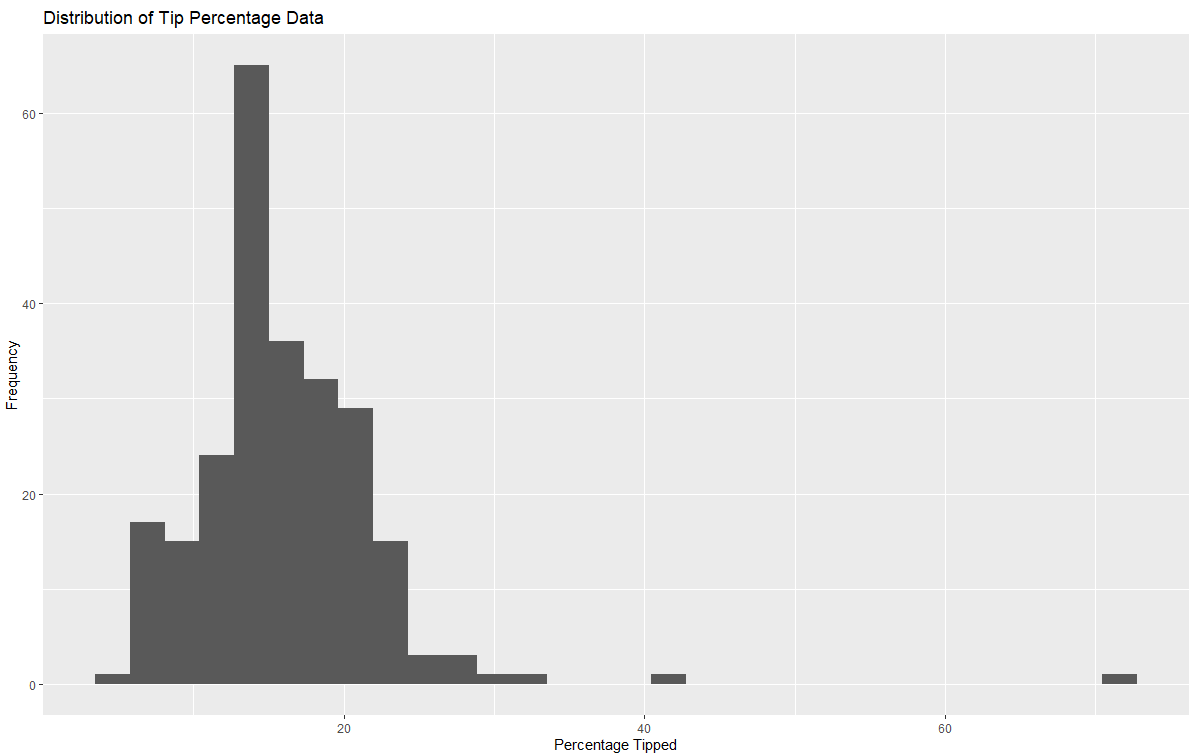
Q.1) Use ggplot to draw one graph of the percentage tipped for the sample that enables you to see the center, spread, and shape of its distribution. Make sure your graph looks nice (meaningful titles and labels, sensible choices of any adjustable parameters). Describe, numerically or in words, the center, spread, and shape of the distribution. Is it normal?

Answer:

1. We have been given the information regarding the total bill and the tip in dollars for the tips dataset.
2. We can find the percentage tipped from the above two attributes of the dataset by calculating: tip/total\_bill
3. After our calculation, we get the following dataset:



1. The following is the graph for the percentage tipped attribute:

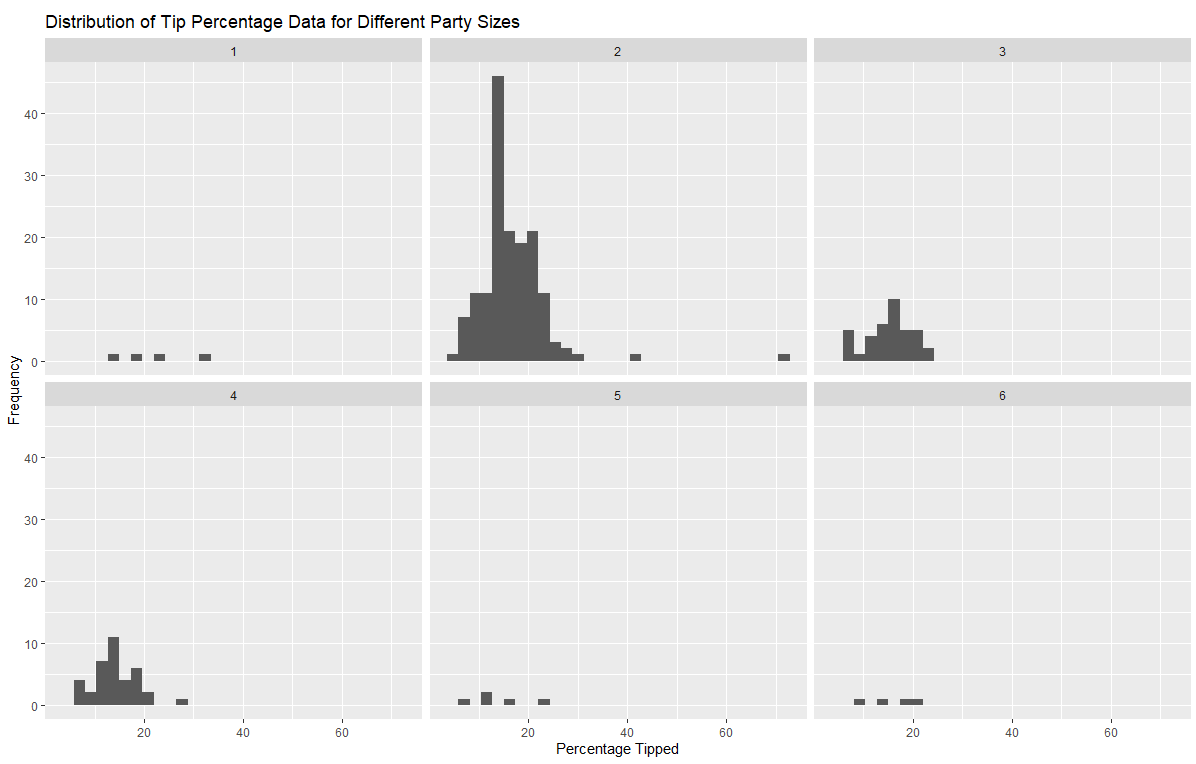


1. We can infer the following points from the above distribution:
   1. Center:
      1. The center of the distribution is the mean or the median of the distribution.
      2. From the graph, we can estimate that the median point for this distribution of data lies somewhere between 15% and 17% because most of the data is focussed around that region in the graph.
      3. After calculating the values for mean and median, we found that Median =15.5, mean=16.094
      4. Here, both are our centers of distribution however while considering we consider median as the true center of distribution over the mean because mean is affected by the presence of outliers in the data unlike median.
   2. Spread:
      1. From the above graph, we can easily estimate the spread of distribution by finding the range of data
      2. For finding the range we find the difference between the maximum value and minimum value of the percent tipped attribute of the dataset.
      3. From the graph, we can see that the maximum value is somewhere around 71 and minimum value is around 4. So our range becomes 71 – 4 i.e 67
      4. Another measure of spread that we can use is Mean Absolute Deviation.
      5. It is the absolute distance of each observation from the mean which in this case is: 5.1891
   3. Shape:
      1. We can state the shape of the distribution just by looking at the above histogram.
      2. We can see that the tail of the histogram is shorter on the left side and longer on the right side.
      3. As the longer tail is associated with larger data values, we can say that that distribution is skewed and since the longer tail is on right side, we say that the distribution is Right Skewed.
      4. Also for the above distribution, we see that data is piling up on a single mound, so we can conclude that our distribution is Unimodal
      5. The right skewness can be confirmed further by the observation that our mean is greater than the median and tends to shift more towards the right direction.
2. Hence we can say that this distribution is not a normal distribution, but is a Right Skewed and Unimodal distribution.

Q.2) Use a set of faceted to display the distribution of percentage tipped for each party size in a way that allows easy comparison. Are there clear differences between the distributions for different party sizes, or are they about the same? Or is it impossible to say? Explain

Answer:

1. From the dataset, we see that there are party sizes having number of people ranging from 1 to 6.
2. We then plot for different party sizes using facet plots to get the following distributions:



1. We can draw forth following conclusions from the above graph:
   1. From the number of percentage tipped observations, we can see that most of the party was having a size of 2,3 and 4 (the most being 2)
   2. There were very less parties for having size 1,5 and 6.
   3. There are some clear differences between the distributions for percentage tipped for different party sizes.
   4. When we see the distributions for party size 1 and 6, it is evident from the graph that the distribution of percentage tipped is uniform shaped and thus we can say that the data is uniformly distributed.
   5. The distribution for party size 5 is almost uniform but it is slightly right-skewed with it’s median or center being around 12.
   6. As for distribution with party size 2, most of the data is focussed in this party size. The median or center for this party size is 16. However this distribution is right-skewed mainly because of the presence of some outlier values
   7. For distribution with party size 3, at a glance this distribution seems symmetric, however after closely checking the graph especially after checking the median and mean values we see that the median is greater than the mean. The mean tends to shift towards left a little so we say that this distribution is left-skewed.
   8. Lastly for party size 4, the median or the center is around 15 and the distribution appears to be right-skewed.
2. So we can say that not all graphs have same shape. Some tend to be different and have different shape such as having uniform shape. Some are showing skewness mostly right-skewed and estimating one graph(party size 3) just by looking at it is little difficult. So in this case we had to calculate the values of measures of centers to correctly find out the shape of distribution. But it was not impossible to find the differences in any of the case.