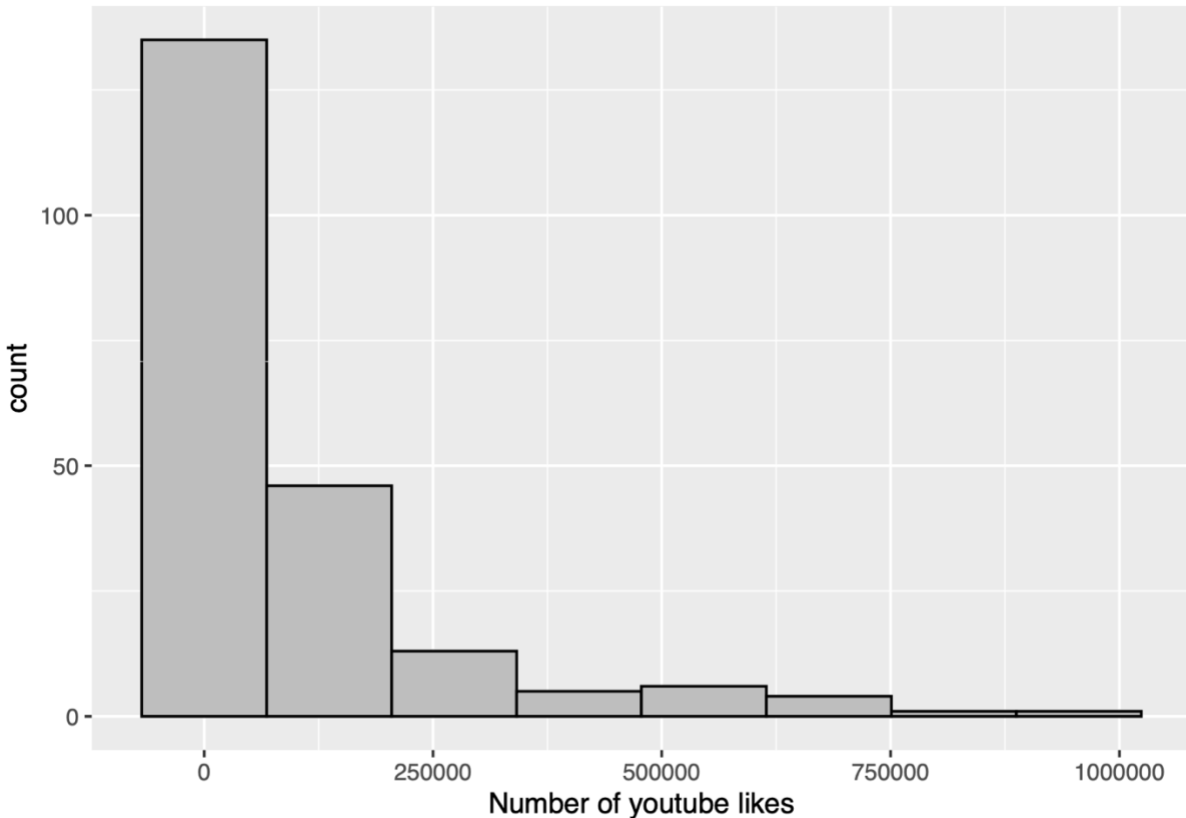


Tutorial 2

Xuanqi Henry Wei



Hi, my friend! How's it going? I would like to share a graph I learnt during the class today.

Based on the data regarding the number of youtube likes for Super Bowl ads from the '.rmd' document of STA130 Problem Set 2, I'm using a histogram to reveal the data. To point out, for histogram, I mean a popular graphing tool summarizing data on an interval scale illustrating the major feature of the data. For each interval, a rectangle is constructed with a base length equal to the range of values in an interval and a height equal to the observations falling into the group.

A histogram has the measurements of the number of youtube likes in its x-axis. These measurements are grouped, for example, from 1 to 1.5 is one group, which is called a bin. Each bin has a bar associated with it that shows the count of observations in that bin. These bins are next to each other with no spaces. The more the count the taller the height of the bar. This way the Y-axis shows the count - the number for ads in each youtube likes categories. Moreover, the overall range for x-axis is from 0 to 1,000,000, which means each bins has a base length for 125,000 youtube likes. The data is cantered at the left side of the graph at the bin from 0 to 125,000 likes where the mode of the whole graph possibly located. The number of ads than has more than 250,000 of youtube likes is much lower than the ones below 250,000 youtube likes.

The histogram follows right-skew distribution, which means most data falls to the left and the right-side tail is longer than its left-side tail. The histogram is heavy tailed, which means the tail is long and it basically implies a larger variance. We usually use it to describe the spread of a

data. Besides, mean is often greater than the median in a right skewed graph. By means, we mean the total number of ads divided by 8 bins in this circumstance; and, by median, we mean list all the youtube likes for each ad incrementally and the middle one is the median.

So, in summary, I think I managed to explain that a histogram lets me see the distribution of the number of youtube likes related to an ad.