2/14/2023

Happy Valentine's Day

Percentiles
Data Summaries
Pie Charts, Bar Graphs, Line Graphs

My YouTube channel: Statistics in Excel playlist:
https://www.youtube.com/playlist?list=PLc023Wgf4h8n34C9CnssfNHMEF5DIUePS

Percentiles are considered a measure of location.
0th percentile, then this value is the minimum value of the data/distribution
100th percentile is the maximum value
50th percentile is the median

Interpret a percentile: eg. 30th percentile: 30% of the data is equal to or smaller that value.

How do we calculate the percentile of a value in a data set?

1. Sort the data from smallest to largest.
2. Find the value in the list (find its position in the list) … if there are 100 values in the list, what position is our value at in the sorted list?
3. Take that position number and divide by the total number of observations in the list.
4. Convert to a percent. Round to a whole percent unless we are less than 1% or greater than 99%.

How do we identify a value in the dataset that corresponds to a certain percentile?

P is the percentile, and n is the number of observations in the dataset. Multiply p*n = location in the sorted list that corresponds to the p-percentile. Round to a whole number to get a value from the list, or we can use the decimal part to estimate between values. Count through the sorted list to find the indicated value.

Bar graphs should start at 0.
Descriptive title.
Axis titles.
If the data is complex, then you also should have legend.
Do not include totals.

For the pie chart, descriptive title, display percentages on each slice, and either a legend or category labels on each slice.

Line graphs are used typically for time-based data. Time goes on the horizontal axis. If the data is ordered, we can use this graph type. We do not need a pivot table to make this graph. A column of times, and the measurements at each point in time. Descriptive title and axis titles.

What makes a graph:
Missing important things like titles, axis labels.

3D: perspective makes the graphs harder to read and can be misleading.
Rescaling by area is bad.
Graphs can be too busy: too many lines, or too many pie slices…
Bar graphs not starting at 0.