

Choose a graph.

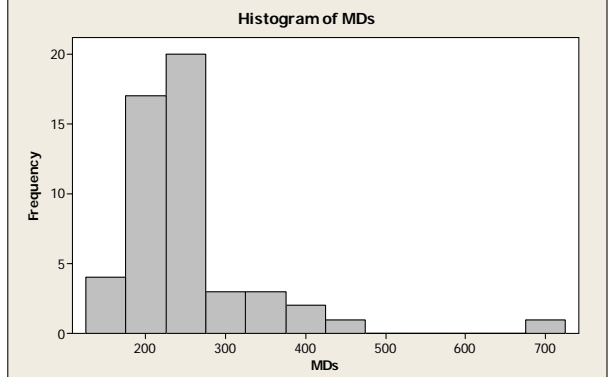
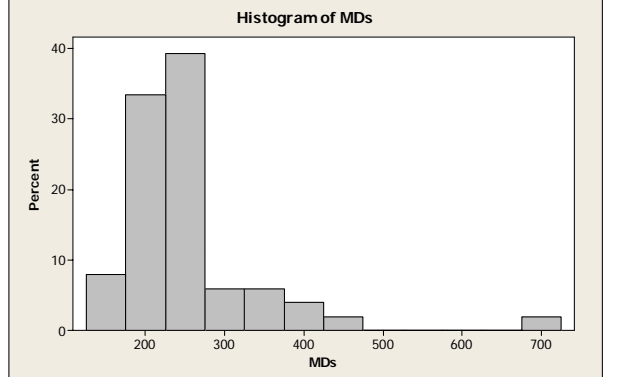
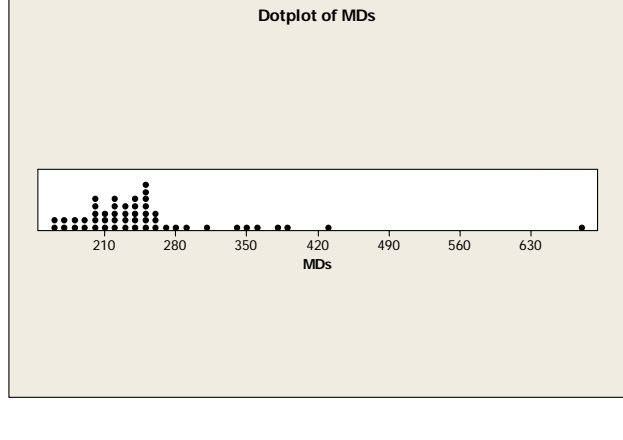
Suppose you have quantitative data and you are asked to choose a graph to find the center of a distribution, or the spread, or the shape.

Notice that here, in the graph, we will ignore what state each observation is from and just look at the distribution of the numbers.

When we make a graph with a summary of the data to answer some question, often we do not use all the information in the dataset. We do not use the information that is not relevant to the question.

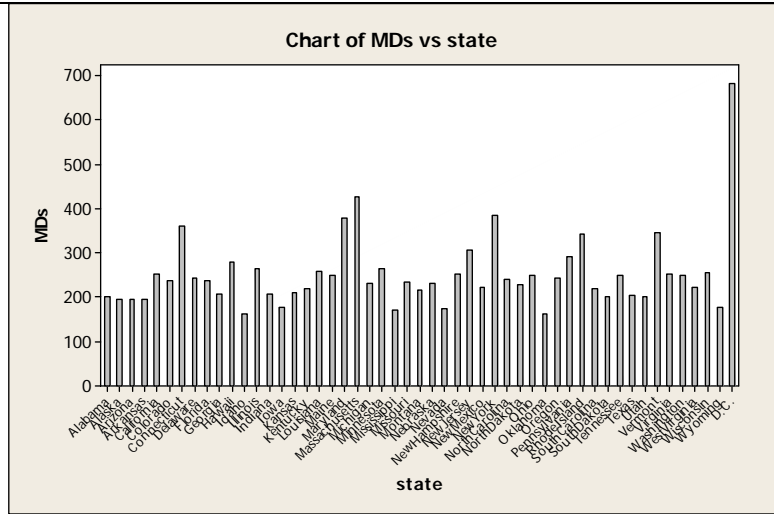
Example: In Table 1.4, we see the number of active medical doctors per 100,000 residents in each state. Make a graph to see the shape of this distribution and estimate the center and spread.

Several choices are possible to correctly answer this. Notice that all of these show the same basic shape, center, and spread.

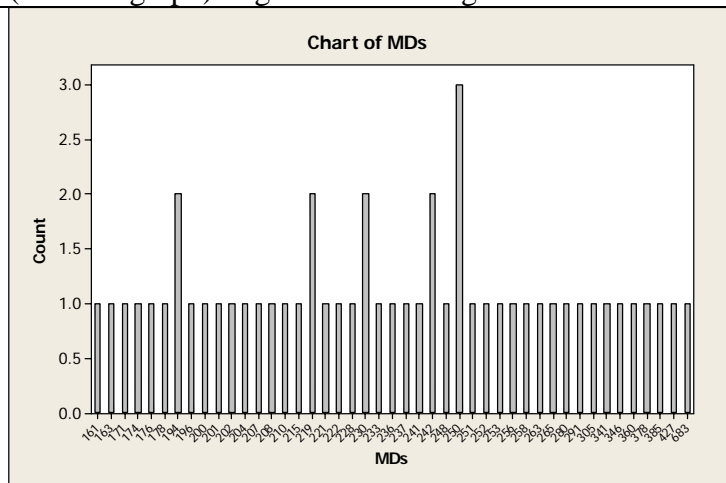
Histogram with actual frequencies	Histogram with relative frequencies (Choose Scale > Y-scale type > Percent)
	
Stemplot	Dotplot
<p><b>Stem-and-Leaf Display: MDs</b></p> <p>Stem-and-leaf of MDs N = 51 Leaf Unit = 10</p> <pre> 9      1  667777999 (22)  2  000001111222333334444 20    2  555555556689 8      3  044 5      3  678 2      4  2 1      4 1      5 1      5 1      6 1      6  8 </pre>	

Each of the following is NOT an appropriate graph to answer that question.

Below is a chart which gives the number of MDs by state. This is not a very useful graph to answer any kind of question. While you can see what the heights of the bars represent, the fact that there is no really meaningful ordering of the categories (states) here and the fact that there are so many of them, just about the only thing obvious from this graph is that one has a lot higher number of MDs than others. You could just as easily have seen that from the numerical data or a frequency graph (histogram, stemplot, or dotplot.) And the frequency graph would tell you more useful information too.



Below is a chart which gives the number of instances of each possible count of MDs. It is treating each possible value of “count” as one value of a categorical variable and showing how many times that value appears. That is useless here. If there were only a few possible values for the counts of MDs, so we could think of the counts of MDs it was reasonable to think of as a categorical variable, this might be an useful graph. (Recall problem 1.1 where the individuals were cars and one variable was the number of cylinders. A bar graph of the number of cylinders (like this graph) might be interesting for that dataset.



## Summary:

**Suppose you have quantitative data and you are asked to choose a graph to find the center of a distribution, or the spread, or the shape.**

The only appropriate choices are frequency graphs, which have the values of the variables along one axis and the frequencies, or relative frequencies, on the other axis. (Histograms have values of the variable on the horizontal axis and frequencies up the vertical axis, stemplots have values of the variable vertically and the “leaves” extend horizontally, giving a visual display of the frequency, dotplots can be oriented in either direction and aren’t shown in our text, but MINITAB will make them.)

It is true that these frequency graphs often do not use all the information in the dataset, such as which individual each observation is from. But that information may not be relevant to answer the particular question asked, so that’s why we don’t need a type of graph that includes it.

Suppose you have categorical data and you want to choose a graph to illustrate the distribution of the data. Again, you need a frequency graph. One such graph is a pie chart, and another is a bar graph, as described in our text. Notice that a bar graph has values of the variable along the horizontal axis and the frequencies, or relative frequencies, on the vertical axis.