MOD 9: MEASURES OF SPREAD ABOUT THE MEDIAN

(This is a modified version of a Los Medanos College activity developed for their Path2Stats program.)

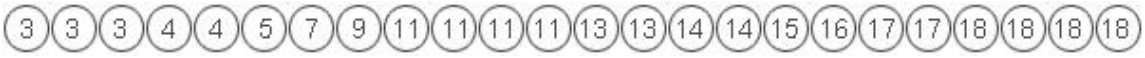
Learning Goals

- Use a five-number summary and a boxplot to describe a distribution.
- Compare distributions from two or more groups.

A statistician will characterize a distribution of data by describing its shape (balanced and symmetrical or skewed) and listing some of its ***descriptive statistics*** such as a measure of central tendency and a measure of variability.

We have discussed two ways to measure central tendency (a.k.a. typical value) for a distribution of data: the mean and the median. Also, if we have chosen to use mean as the typical value, we can measure variability using either the average distance from the mean (ADM) or standard deviation. However, there's a reason we have not discussed how to measure variability when we must use the median as the typical value instead of the mean. Technically statisticians do not measure variability about the median. However we can use the five-number summary and boxplots to quantify spread with regard to the median.

1) Find the median for this set of data. Mark the median on the data set, and then find the "median" of the lower half of the data and the "median" for the upper half of the data.



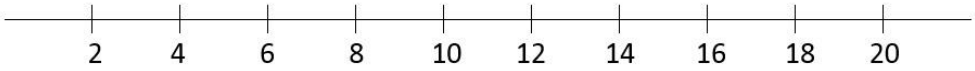
5-Number Summary	
Minimum	
Q1	
Q2	
Q3	
Maximum	

You just found the quartiles for this distribution of data. The quartiles always divide the data into four groups, with equal counts of data values in each group. The median is the second quartile Q2.

Fill in the table to find the 5-number summary for this distribution of data.

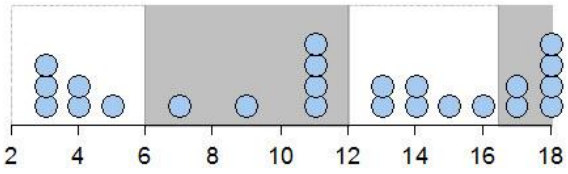
On the next page, we'll use the five-number summary to create a boxplot for this data on the next page.

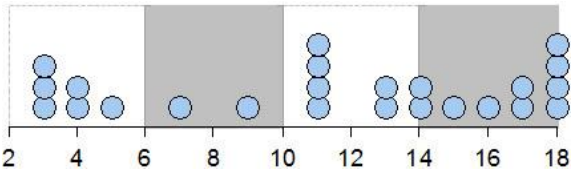
To create a boxplot for the distribution of data on the previous page, mark the quartiles on the number line. Then draw a box above the number line with the left side of the box at Q1 and the right side of the box at Q3. Draw a vertical line inside the box at Q2. Next, use whiskers on either side of the box to mark the minimum and maximum data values.



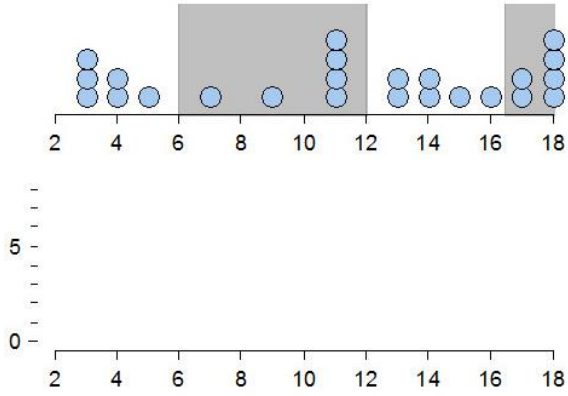
2) When we have a distribution of data, we can use dividers with *equal widths* or *equal counts* to either draw a boxplot or histogram as a visual display of the data.

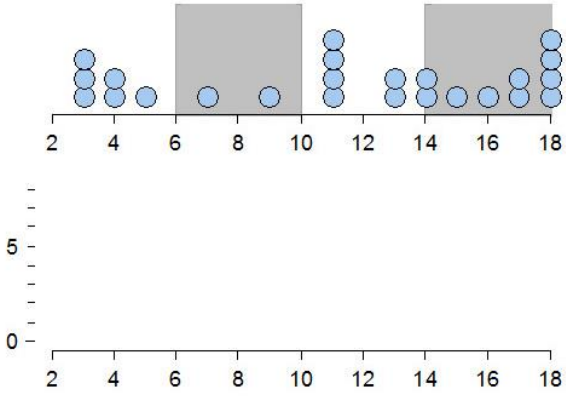
For each of the following graphs, determine whether the distribution is divided into *equal widths* or *equal counts*? Label each graph accordingly. Finally, write the percentage of data in each of the four shaded sections for both graphs.



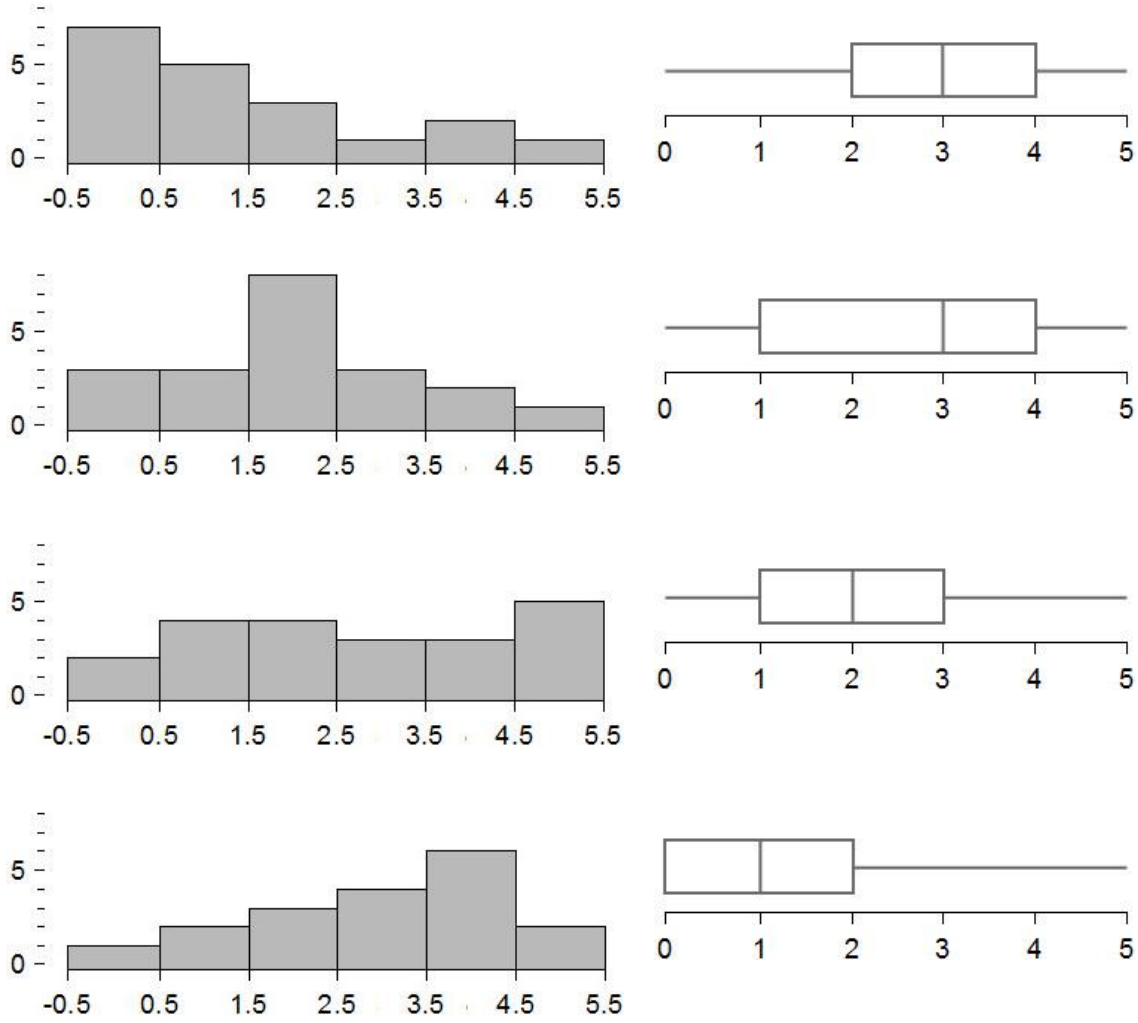


3) Create a histogram under the distribution with the correct type of dividers. Also create a boxplot under the distribution with the correct type of dividers.





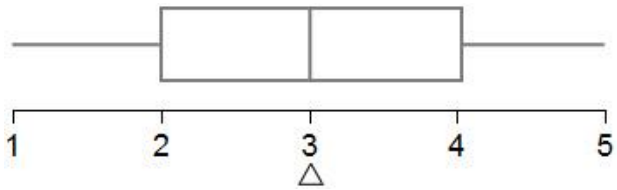
4) Match each histogram with its corresponding boxplot (i.e. the boxplot that was created from the same data distribution as the histogram).



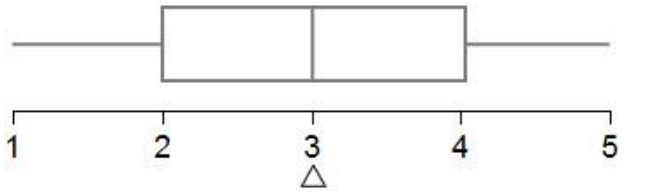
Explain the strategies you used to match each histogram with its corresponding boxplot.

5) Make up a data set (n=11) that fits the boxplot and has the most amount of spread possible about the mean. Then make up a second data set (n=11) that fits the boxplot and has the least amount of spread possible about the mean.

Most amount of spread possible:

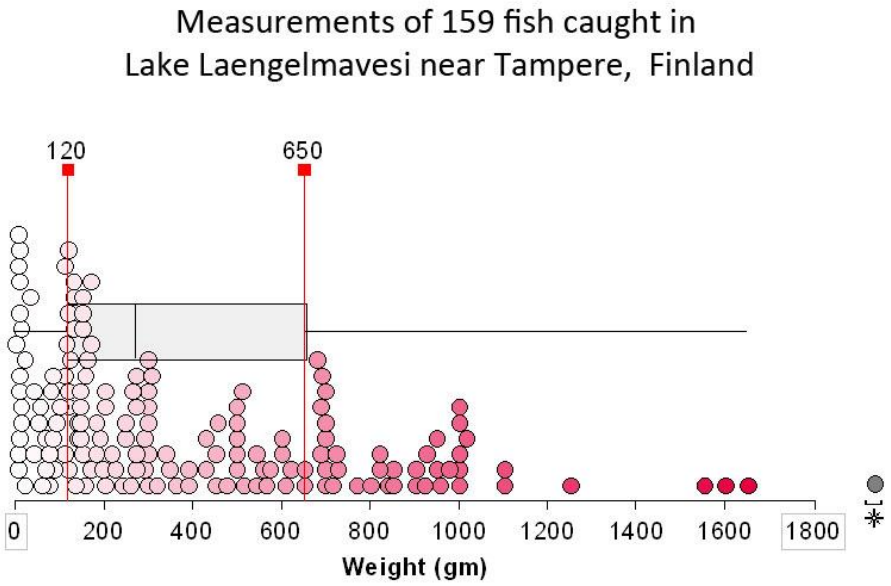


Least amount of spread possible:

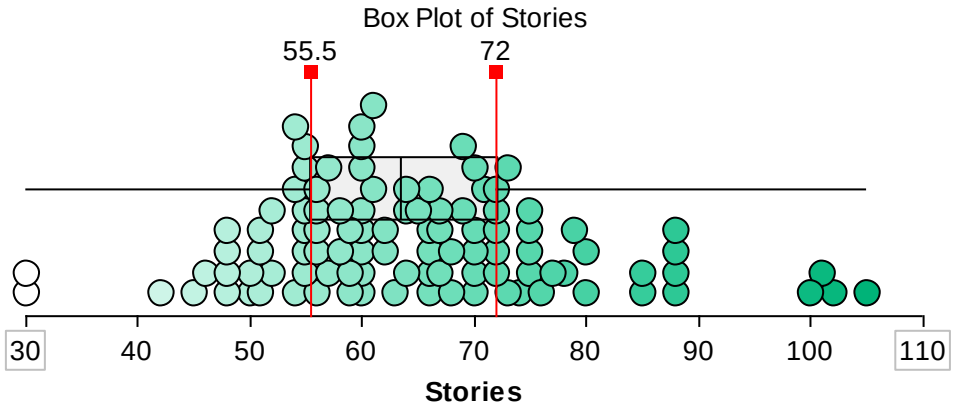


6) For each boxplot below, describe what the histogram for the distribution of data might look like. While you cannot know for certain what the exact shape of each histogram would be, you should be able to make some generalizations about the shape of each histogram.

7) **WAIT! We'll work through this one together as a class**. Use the 1.5 IQR criterion to find the outliers (if any) for the following data set.



8) Use the 1.5 IQR criterion to find the outliers (if any) for the following data set.



DO YOU KNOW YOUR GEOGRAPHY?

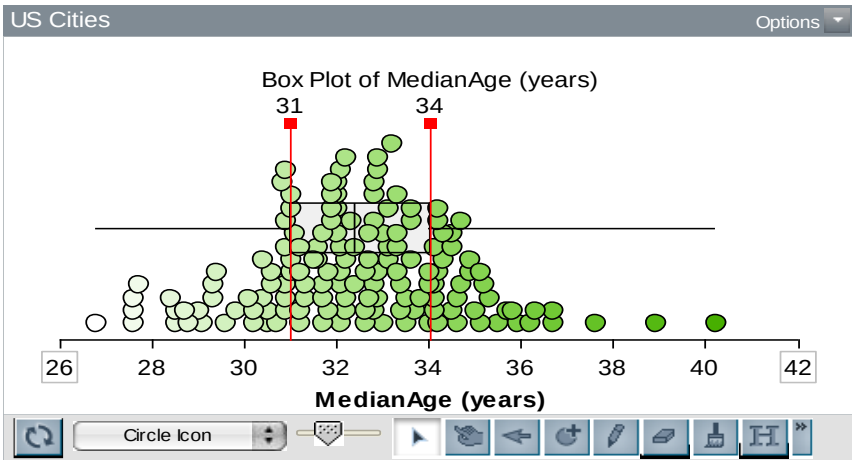
Please do not access any technology for this activity. Your instructor will distribute an ungraded geography quiz. Please use your best estimate to answer the questions.

9) When the activity is complete, use the space below to draw the two boxplots the class created. Be sure to label the five-number summary on each boxplot and indicate the outlier boundaries on each boxplot as well. We will refer to these boxplots in subsequent lessons.

HOMEWORK: IDENTIFYING OUTLIERS WITH THE 1.5 IQR CRITERION

Remember – neatness and completeness count. Also, you must show your work. The correct result without a sufficient amount of correct and appropriate work is worth zero points.

1) The data distribution represents the median population age for 146 U.S. cities. Use the 1.5 IQR criterion to find the outliers (if any) for the following data set. Be sure to show your work and indicate the boundaries for the outliers on the graph.



Continued on the next page …

2) The data distribution represents the number of larcenies, burglaries, and motor vehicle thefts per 100,000 residents for 146 U.S. cities. Use the 1.5 IQR criterion to find the outliers (if any) for the following data set. Be sure to show your work and indicate the boundaries for the outliers on the graph.