

STATISTICS AND BOX-AND-WHISKERS PLOTS

Statistics can be organized and charted in a box-and-whiskers plot. Box-and-whiskers plots separate data into quartiles. Quartiles are based on the median of the data, the median of the upper half of the data, and the median of the lower half of the data. Pieces of data that are extremely different from the rest of the data are excluded from the quartiles and are called outliers.

Box-and-Whiskers Plots

Box-and-whiskers plots are used to separate data into four sections. The parts will differ in length in the graph, but each part will contain one fourth of the data, with the exception of the outliers.

Median - Separates the entire data set in half.

LQ – Lower Quartile - Median of the lower half of the data.

UQ – Upper Quartile - Median of the upper half of the data.

Interquartile Range – Difference between the UQ and LQ.

Outliers – Data that falls beyond the upper quartile or lower quartile and is more than 1.5 times the value of the interquartile range added to the UQ or subtracted from the LQ.

Lower Extreme – Smallest piece of data.

Upper Extreme – Largest piece of data.

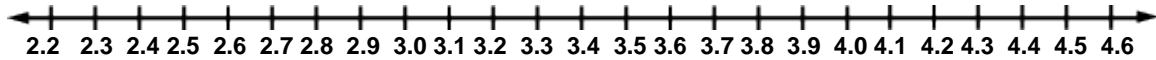
Follow the steps below to make a box-and-whiskers plot for the given data. The data represents the average monthly precipitation that occurred in Columbus, OH, in 2003. Also check the data for outliers.

Average Monthly Precipitation Columbus, OH (in inches) 2003			
2.5	2.2	2.9	3.3
3.9	4.1	4.6	3.7
2.9	2.3	3.2	2.9

Step 1: Put the data in **order** from **least to greatest**.

2.2, 2.3, 2.5, 2.9, 2.9, 2.9, 3.2, 3.3, 3.7, 3.9, 4.1, 4.6

Step 2: Draw a **number line** with a **scale** that fits the data.



Step 3: Find the **median**.

$$\frac{2.9 + 3.2}{2} = 3.05$$

Step 4: Find the quartiles (**LQ** and **UQ**).

Lower quartile - median of the lower half.

Upper quartile - median of the upper half.

Lower Quartile

$$\frac{2.5 + 2.9}{2} = 2.7$$

Upper Quartile

$$\frac{3.7 + 3.9}{2} = 3.8$$

Step 5: Calculate the interquartile range.

Subtract the lower quartile (LQ) from the upper quartile (UQ).

$$3.8 - 2.7 = 1.1$$

Step 6: Check for outliers.

Upper limit for outliers

Multiply the interquartile range by 1.5.

$$1.1 \times 1.5 = 1.65$$

Add that number to the UQ.

$$3.8 + 1.65 = 5.45$$

Compare to see if any numbers in the data set are greater than this sum.

The highest number in the data set is 4.6, so there is **no** outlier.

Lower limit for outliers

Subtract 1.65 from the LQ.

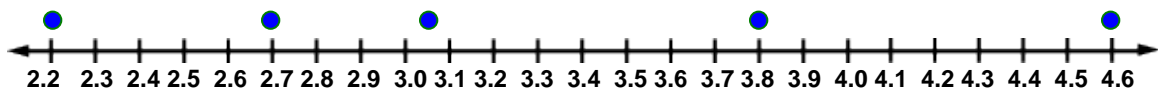
$$2.7 - 1.65 = 1.05$$

Compare to see if any numbers in the data set are lower than this difference.

The lowest number in the data set is 2.2, so there is **no** outlier.

~There are no outliers in this data set.~

Step 7: **Plot the points** for the following data points: **lower extreme, LQ, median, UQ, and upper extreme.**



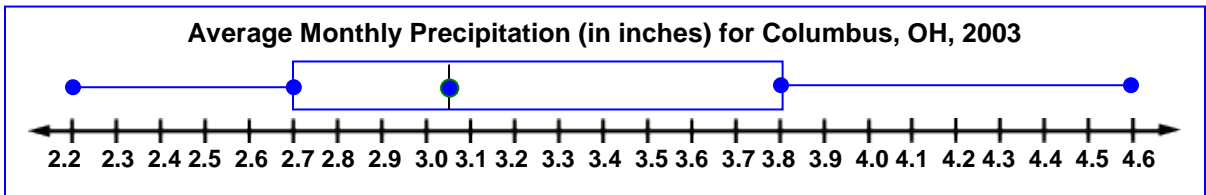
Step 8: Draw the **box-and-whiskers** graph.

A rectangle (box) extends from the lower quartile (LQ) point to the upper quartile (UQ) point.

A vertical line is drawn through the median point of the data set.

Lines (whiskers) extend from the LQ point to the lower extreme data point and from the UQ point to the upper extreme data point.

Add a title to the graph.



*Note:

- ❖ Data points 2.2, 2.3, and 2.5 are located under the left “whisker”.
($\frac{1}{4}$ th of the data points)
- ❖ Data points 2.9, 2.9, and 2.9 are located to the left of the median under the “box”. ($\frac{1}{4}$ th of the data points)
- ❖ Data points 3.2, 3.3, and 3.7 are located to the right of the median under the “box”. ($\frac{1}{4}$ th of the data points)
- ❖ Data points 3.9, 4.1, and 4.6 are located under the right “whisker”.
($\frac{1}{4}$ th of the data points)

Dot Plot

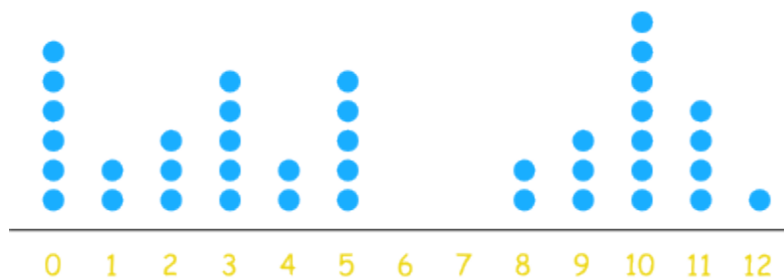
A Dot Plot is a graphical display of data using dots.

A survey of "How long does it take you to eat breakfast?" has these results:

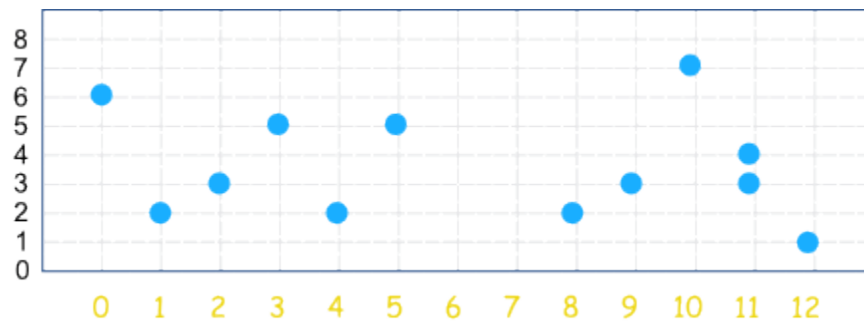
Minutes:	0	1	2	3	4	5	6	7	8	9	10	11	12
People:	6	2	3	5	2	5	0	0	2	3	7	4	1

Which means that 6 people take 0 minutes to eat breakfast (they probably had no breakfast!), 2 people say they only spend 1 minute having breakfast, etc.

And here is the dot plot:



Another version of the dot plot has just one dot for each data point. This has the same data as above:



But notice that we need to have lines and numbers on the side so we can see what the dots mean.