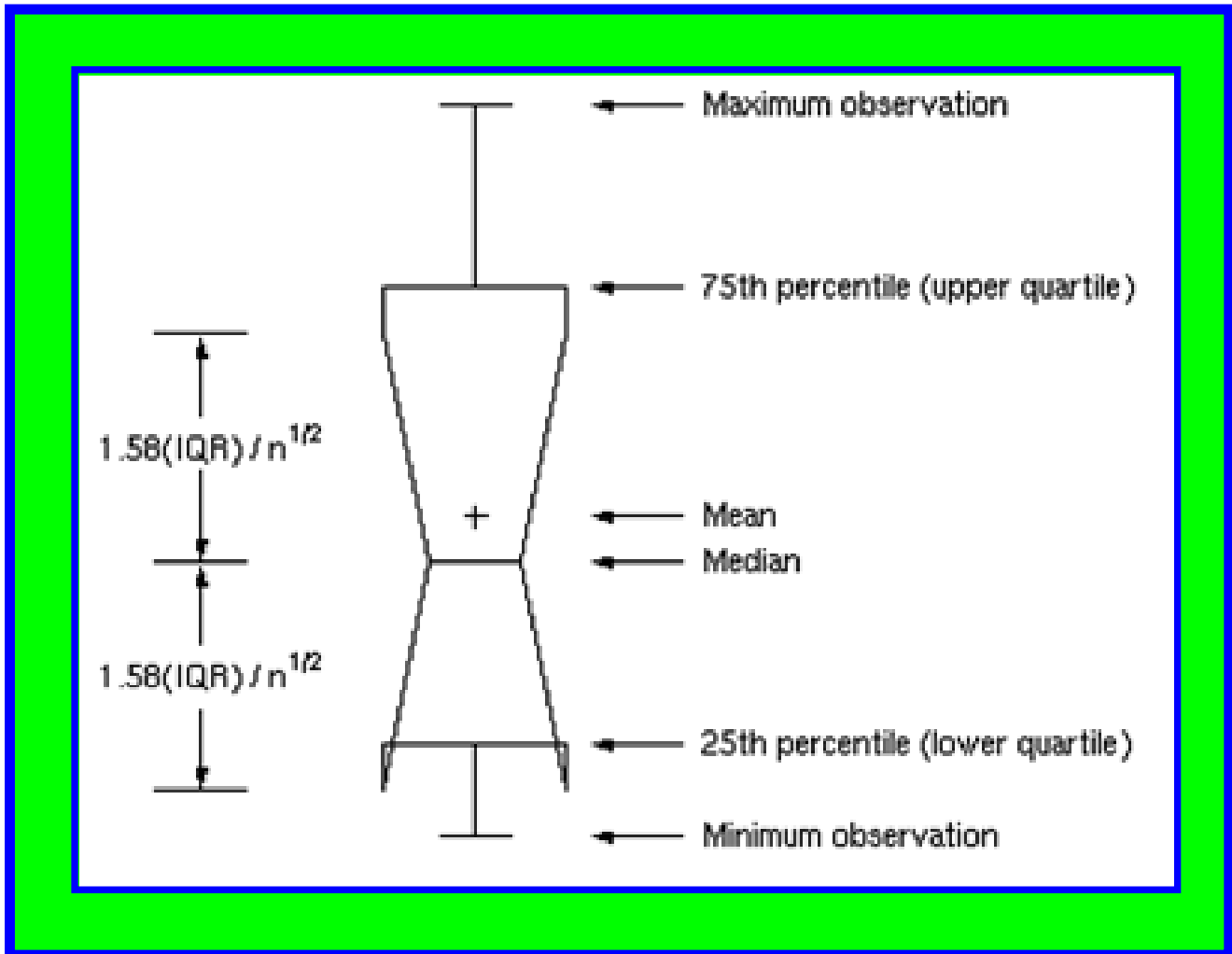
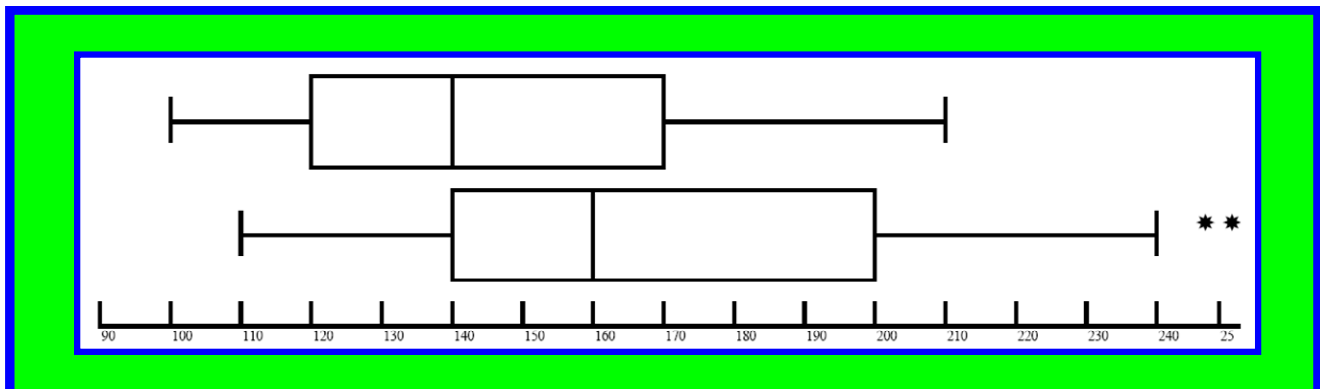


## Box Plots



Box-and-whisker diagrams, or Box Plots, use the concept of breaking a data set into fourths, or quartiles, to create a display. The *box* part of the diagram is based on the middle (the second and third quartiles) of the data set. The *whiskers* are lines that extend from either side of the box. The maximum length of the whiskers is calculated based on the length of the box. The actual length of each whisker is determined after considering the data points in the first and the fourth quartiles.



Although box-and-whisker diagrams present less information than histograms or dot plots, they do say a lot about distribution, location and spread of the represented data. They are particularly valuable because several box plots can be placed next to each other in a single diagram for easy comparison of multiple data sets.

### What can it do for you?

If your improvement project involves a relatively limited amount of individual quantitative data, a box-and-whisker diagram can give you an instant picture of the shape of variation in your process. Often this can provide an immediate insight into the search strategies you could use to find the cause of that variation.

Box-and-whisker diagrams are especially valuable to compare the output of two processes creating the same characteristic or to track improvement in a single process. They can be used throughout the phases of the Lean Six Sigma methodology, but you will find box-and-whisker diagrams particularly useful in the *analyze* phase.

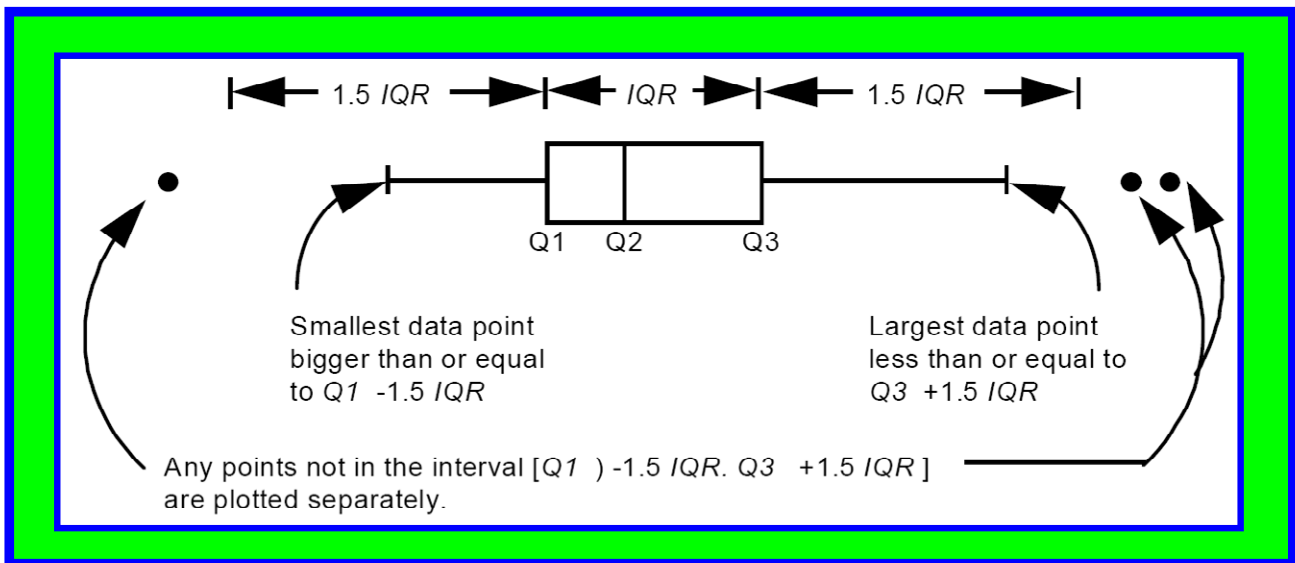
### How do you do it?

1. **Decide which Critical-To-Quality (CTQ) characteristic you wish to examine.** This CTQ must be measurable on a linear scale. That is, the incremental value between units of measurement must be the same. For example, time, temperature, dimension and spatial relationships can usually be measured in consistent incremental units.
2. **Measure the characteristic and record the results.** If the characteristic is continually being produced, such as voltage in a line or temperature in an oven, or if there are too many items being produced to measure all of them, you will have to sample. Take care to ensure that your sampling is random.
3. **Count the number of individual data points.**
4. **List the data points in ascending order.**
5. **Find the *median* value.** If there are an odd number of data points, the median is the data point that is halfway between the largest and the smallest ones. (For example, if there are 35 data points, the median value is the value of the 18th data point from either the top or the bottom of the list.) If there is an even number of points, the median is halfway between the two points that occupy the centermost position. (If there were 36

Order	Value	Boundary
1	27.75	
2	37.35	
3	38.35	
4	38.35	
5	38.75	
<b>Second Quartile</b>		<b>39.250</b>
6	39.75	
7	40.50	
8	41.00	
9	41.15	
10	42.55	
<b>Third Quartile</b>		<b>42.725</b>
11	42.90	
12	43.60	
13	43.85	
14	47.30	
15	47.90	
<b>Fourth Quartile</b>		<b>48.025</b>
16	48.15	
17	49.86	
18	51.25	
19	51.50	
20	56.00	
<b>Data Table divided into quartiles</b>		

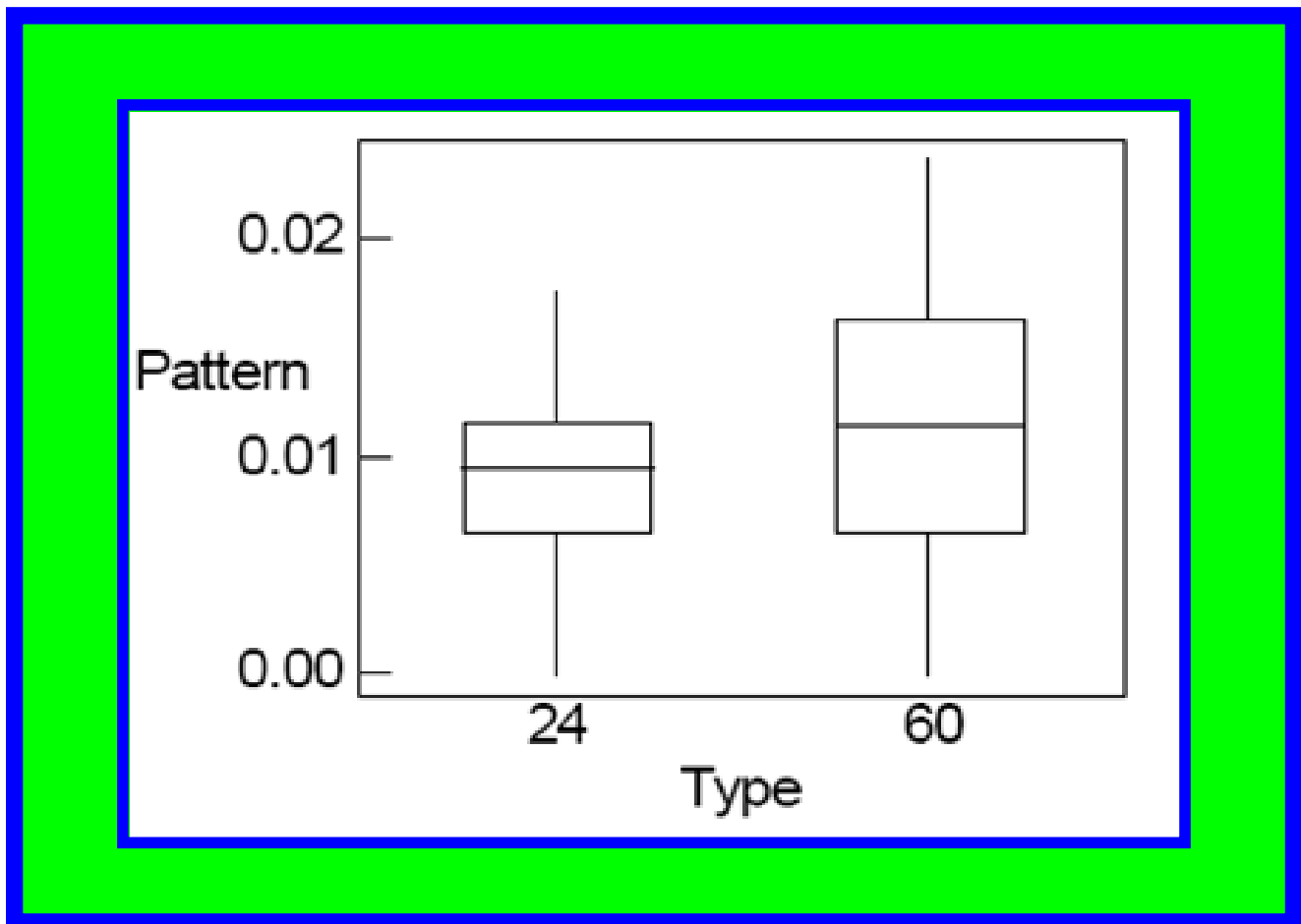
points, the median would be halfway between point 18 and point 19. To find the median value, add the values of points 18 and 19, and divide the result by 2.) If you think of the list of data points being divided into quarters (quartiles), the median is the boundary between the second and the third quartile.

6. The next step is to find the boundaries between the first and second and the third and fourth quartiles. The first quartile boundary is halfway between the last data point in the first quartile and the first data point in the second quartile. (If one data point is on the median, that data point is considered to be the last point in the second quartile *and* the first point in the third quartile.) In a similar way, find the third quartile boundary, the halfway point between the last value in the third quartile and the first value in the fourth quartile.
7. Draw and label a scale line with values. The value of the scale should begin lower than your lowest value and extend higher than your highest value. The scale line may be either vertical or horizontal.
8. Using the scale as a guideline, create a box above or to the right of the scale. One end of the box will be the first quartile boundary; the other will be the third quartile boundary. (The width of the box is somewhat arbitrary. Boxes tend to be long and thin. As an option, if you have multiple data sets with different numbers of data points in each set, make the width of the boxes so that they correspond roughly with the relative quantity of data represented in each box.)
9. Draw a line through the box to represent the median (second quartile boundary).
10. The next step is to draw the *whiskers* on the ends of the box. Find the *inter-quartile range* (IQR) by subtracting the value of the first quartile boundary from that of the third quartile boundary.



- a. Smallest data point is bigger than or equal to  $Q1 - 1.5 IQR$
- b. Largest data point is less than or equal to  $Q3 + 1.5 IQR$
- c. Any points not in the interval  $[Q1 - 1.5 IQR; Q3 + 1.5 IQR]$  are plotted separately.

11. **Multiply the IQR by 1.5.** (The use of 1.5 as a multiplier is a convention that has no exact statistical basis. Multiplying by this constant helps take into consideration the fact that the first and fourth quartiles will naturally have a somewhat wider dispersion than the second and third quartiles.)
  12. **Subtract the value of 1.5(IQR) from the value of the first quartile boundary.** Find the smallest data point in your list that is equal to or larger than this value. Make a tick mark representing this data point to the left of your box (or above, if you used a vertical scale). Draw a line, the first *whisker*, from the side of the box to the tick mark.
  13. **Add the value of 1.5 (IQR) to the value of the third quartile boundary.** Find the largest data point in your list that is equal to or smaller than this value. Make a tick mark representing this data point to the right of your box (or below, if you used a vertical scale). Draw another *whisker* to this tick mark.
  14. It is possible that some data points in your list will lie outside of the ends of the whiskers you determined in steps 12 and 13. These points are called *outliers*. **Plot any outliers as dots beyond the whiskers.**
- [Note: steps 3 through 14 happen automatically if you use Excel, Minitab, or JMP to create your box-and-whisker diagram. If you are familiar with these software packages, their use can greatly simplify the process of making effective box-and-whisker diagrams.]*
15. **Title and label your box-and-whisker diagram.**



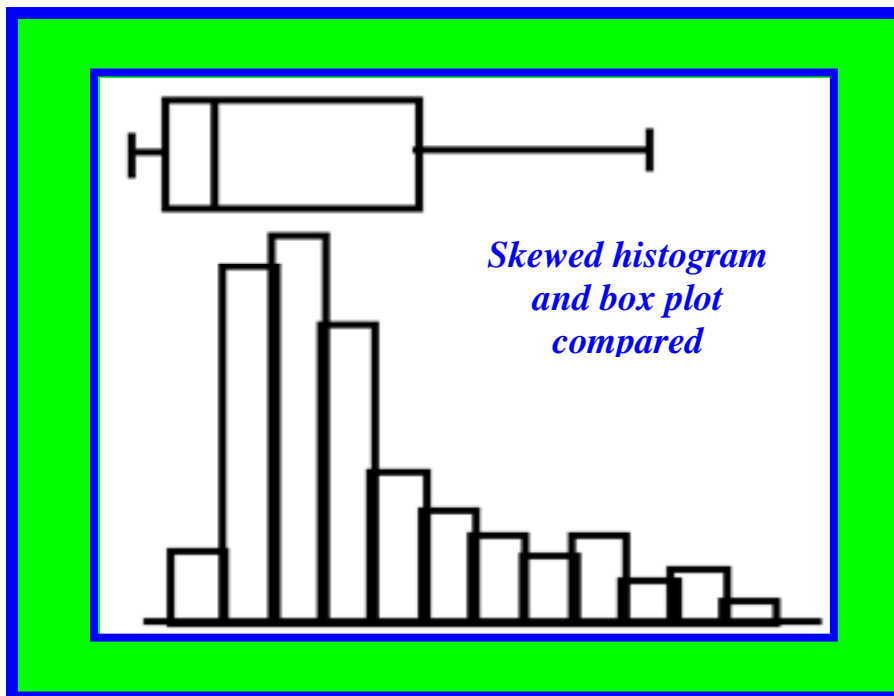
## Now what?

The shape that your box-and-whisker diagram takes tells a lot about your process.

One way to help you interpret box plots is to imagine that the way a data set looks as a histogram is something like a mountain viewed from ground level and a box-and-whisker diagram is something like a contour map of that mountain as viewed from above.



In a *Skewed histogram and box plot compared*

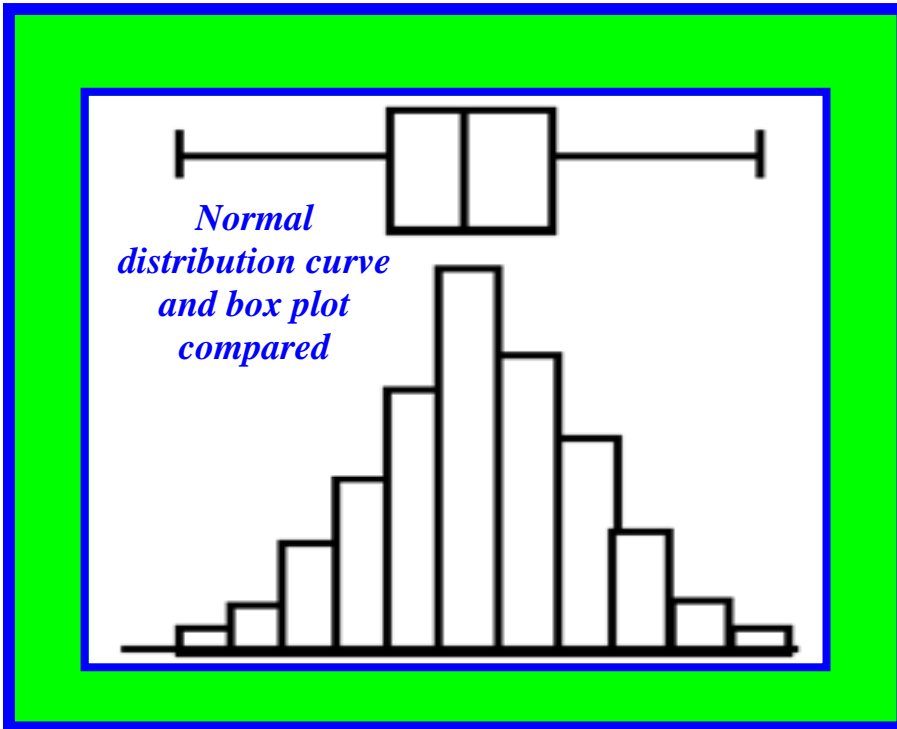


The second-quartile box is considerably larger than the third-quartile box, and the whisker associated with the first quartile extends almost to the end of the 1.5 IQR limit. An outlier beyond the 1.5 IQR limit of the whisker further emphasizes the fact that the data is strongly skewed in this direction. On the other side of the distribution, the whisker associated with the fourth quartile is well within the 1.5 IQR. In fact, the fourth-quartile whisker is shorter than the third-quartile box.

A histogram of this data would show a strongly skewed distribution verging on a precipice that fell off at the high end of the values. This kind of data set often occurs when there is a natural limit at one end of the distribution or a 100% screening is done for one specification limit.

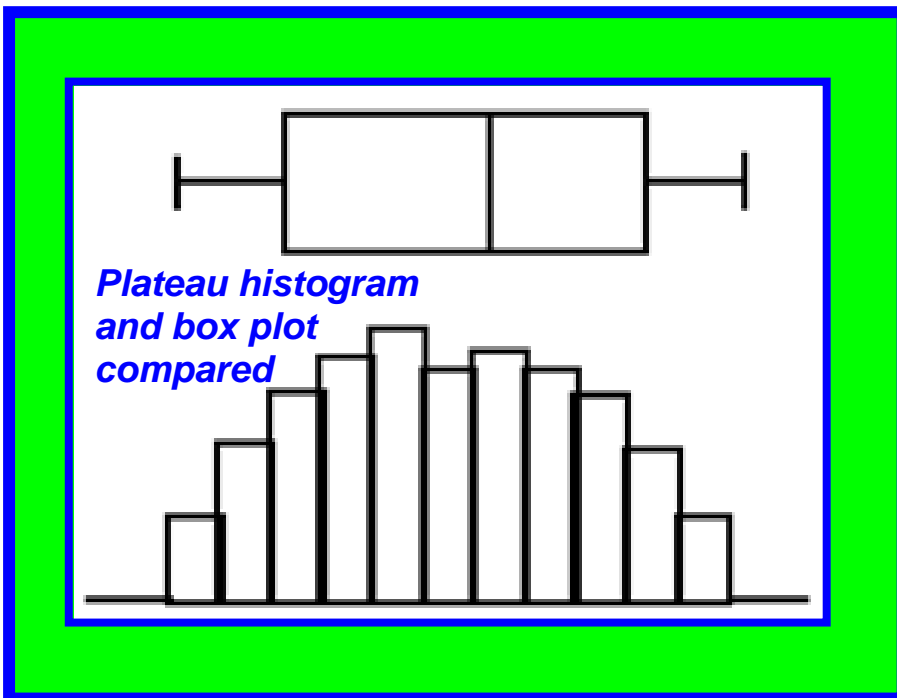
Although box-and-whisker diagrams can be oriented horizontally, they are more often displayed vertically, with lower values at the bottom of the scale.

The next example is what a normal distribution might look like as a box plot.



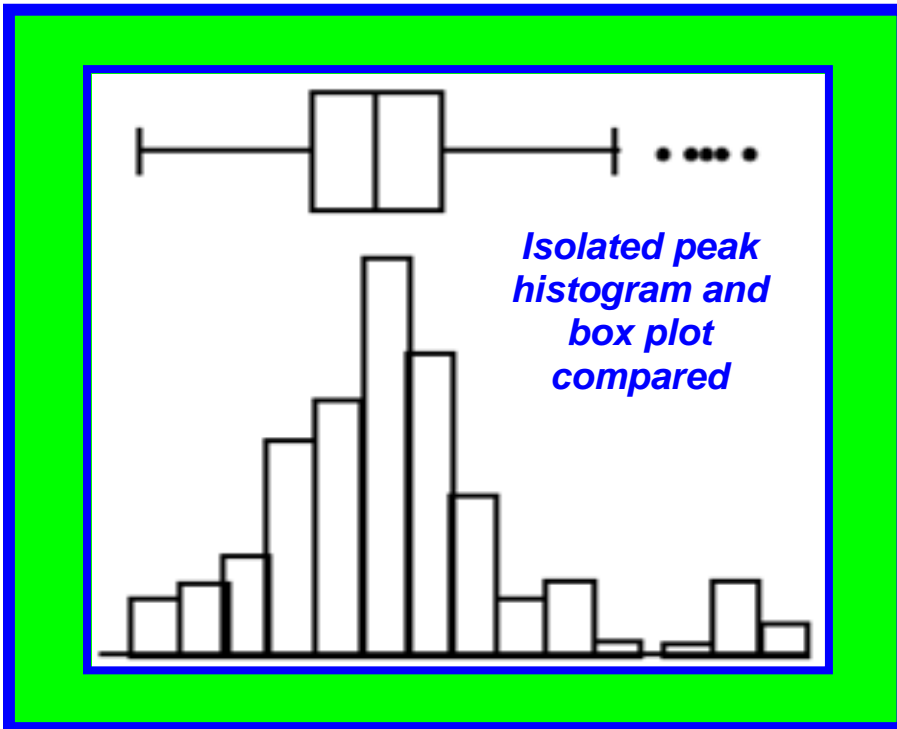
The second- and third-quartile boxes are approximately the same size. The whiskers are similar to each other in length and extend close to the 1.5 IQR limit.

If the data set were actually a combination of two different distributions, for example, material from two suppliers or two machines, it might form a histogram that looked like a plateau or a mountain with twin peaks.



The box plot would show an even distribution, but would have relatively large boxes and relatively short whiskers.

If there were a small amount of data from a different distribution included in the data set, for example, if there were a short-term process abnormality or a data collection error, the histogram formed would look like a mountain with a small isolated peak.

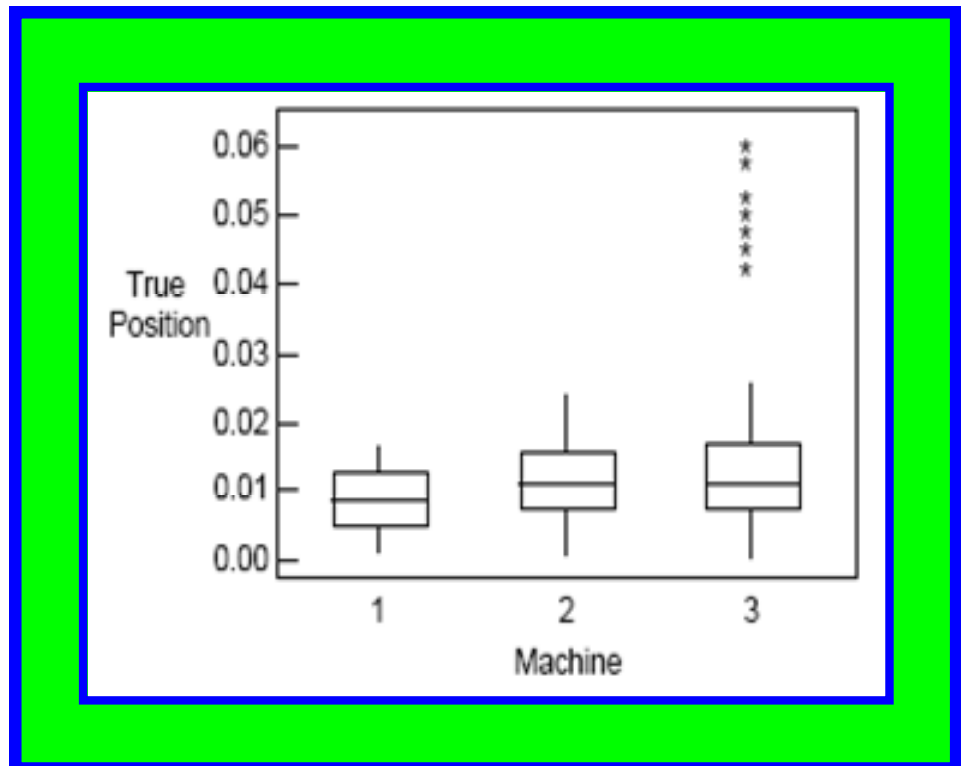


The box plot for that data set would look like one for a normal distribution but with a number of outliers beyond one whisker.

### ***Some final tips***

A box-and-whisker diagram is an easy way to compare processes or to chart the improvement process in one process.

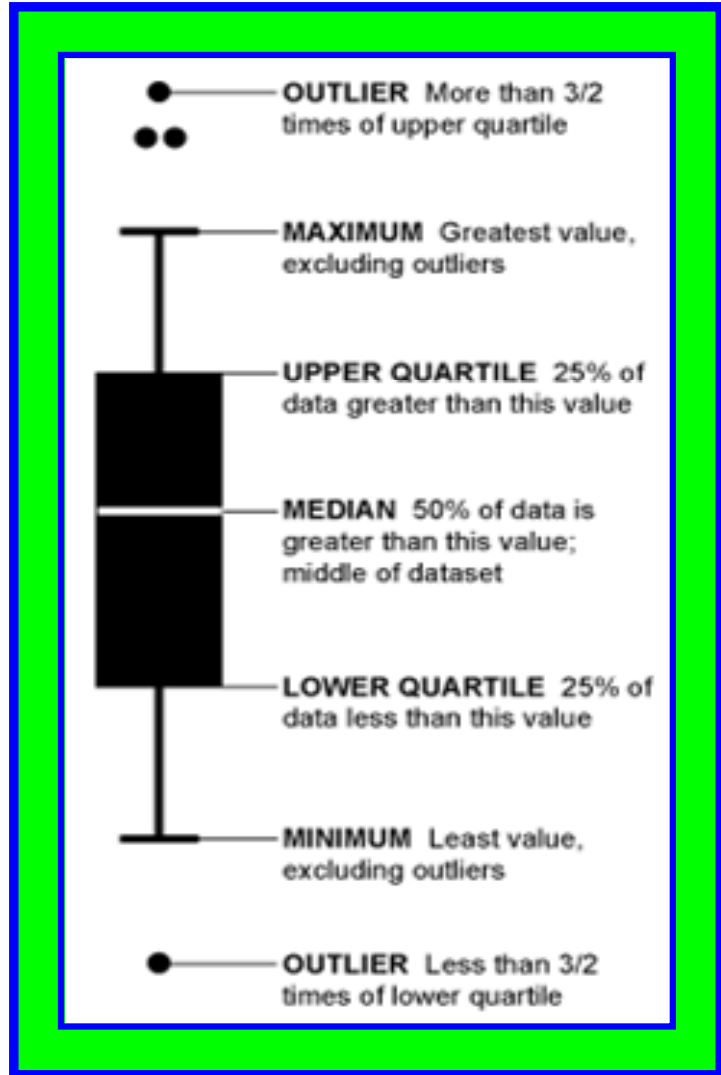
Box-and-whisker diagrams can quickly give you a comparative feel of the distribution of sets of data. They show the distributional spread through the length of the box and the whiskers.



Some idea of the symmetry of the distribution can also be gained by comparing the two segments of the box and the relative lengths of the whiskers. The existence and displacement of outliers gives some indication of the level of control in the process.

Two or more box-and-whisker diagrams drawn side by side to the same scale are an effective way to compare samples in a way that is compact and uncluttered. Many box plots can be added to a diagram without creating visual overload.

Not only can box-and-whisker diagrams help you see which processes need improvement, by comparing initial box-and-whisker diagrams with subsequent ones, they can also help you track that improvement. If specification limits or improvement targets are involved in your process, they can be added to the diagram to help visualize progress.



Steven Bonacorsi is a Certified Lean Six Sigma Senior Master Black Belt instructor and coach. Steven Bonacorsi has trained hundreds of Master Black Belts, Black Belts, Green Belts, and Project Sponsors and Executive Leaders in Lean Six Sigma DMAIC and Design for Lean Six Sigma process improvement methodologies. He has led some of the largest deployments in the world and has saved hundreds of millions with his project teams results.

The AIT Group, Inc.  
Steven Bonacorsi, Vice President,  
MBA, MS-CIS, PMP  
Lean Six Sigma Master Black Belt  
3135 South Price Road, Suite 115  
Chandler, AZ 85248-3549  
Phone: +(1) 888.826.2484

[sbonacorsi@comcast.net](mailto:sbonacorsi@comcast.net)  
<http://www.linkedin.com/in/stevenbonacorsi>  
<http://www.theaitgroup.com>

