

## Box Plots

### Prerequisites

#### [Percentiles](#)

We have already discussed techniques for visually representing data (see [histograms](#) and [frequency polygons](#)). In this section we present another important graph, called a [box plot](#). Box plots are useful for identifying outliers and for comparing distributions. We will explain box plots with the help of data from an in-class experiment. Students in Introductory Statistics were presented with a page containing 30 colored rectangles. Their task was to name the colors as quickly as possible; their times were recorded. We'll compare the scores for the 16 men and 31 women who participated in the experiment by making separate box plots for each gender. (Such a display is said to involve [parallel box plots](#).)

There are several steps in constructing a box plot. The first relies on the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles in the distribution of scores. Figure 1 shows how these three statistics are used. For each gender we draw a box extending from the 25<sup>th</sup> percentile to the 75<sup>th</sup> percentile. The 50<sup>th</sup> percentile is drawn inside the box. Therefore,

the bottom of each box is the 25th percentile,  
the top is the 75th percentile,  
and the line in the middle is the 50th percentile.

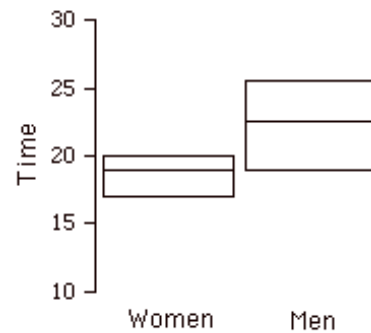
The data for the women in our sample are shown in Table 1.

Table 1. Times (in seconds) for women to name the colors.

14	17	18	19	20	21	29
15	17	18	19	20	22	
16	17	18	19	20	23	
16	17	18	20	20	24	
17	18	18	20	21	24	

For these data, the 25<sup>th</sup> percentile is 17, the 50<sup>th</sup> percentile is 19, and the 75<sup>th</sup> percentile is 20. For the men (whose data are not shown), the 25<sup>th</sup> percentile is 19, the 50<sup>th</sup> percentile is 22.5, and the 75<sup>th</sup> percentile is 25.5.

Figure 1. The first step in creating box plots.



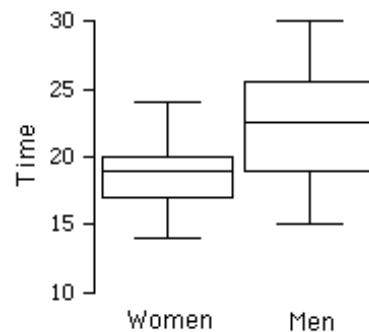
Before proceeding, the terminology in Table 2 is helpful.

Table 2. Terminology.

Name	Formula	Value for Women's Data
Upper Hinge	75th Percentile	20
Lower Hinge	25th Percentile	17
H-Spread	Upper Hinge - Lower Hinge	3
Step	1.5 x H-Spread	4.5
Upper Inner Fence	Upper Hinge + 1 Step	24.5
Lower Inner Fence	Lower Hinge - 1 Step	12.5
Upper Outer Fence	Upper Hinge + 2 Steps	29
Lower Outer Fence	Lower Hinge - 2 Steps	8
Upper Adjacent	Largest value below Upper Inner Fence	24
Lower Adjacent	Smallest value above Lower Inner Fence	14
Outside Value	A value beyond an Inner Fence but not beyond an Outer Fence	29 (this value is on the fence, but not beyond)
Far Out Value	A value beyond an Outer Fence	None in these data

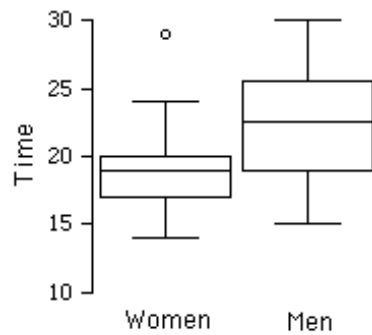
Continuing with the box plots, we put "whiskers" above and below each box to give additional information about the spread of data. Whiskers are vertical lines that end in a horizontal stroke. Whiskers are drawn from the upper and lower hinges to the upper and lower adjacent values (24 and 14 for the women's data).

Figure 2. The box plots with the whiskers drawn.



Although we don't draw whiskers all the way to outside or far out values, we still wish to represent them in our box plots. This is achieved by adding additional marks beyond the whiskers. Specifically, outside values are indicated by small "o's, and far out values are indicated by asterisks. In our data, there are no far out values, and just one outside value. This outside value of 29 is for the women and is shown in Figure 3.

Figure 3. The box plots with the outside value shown.



There is one more mark to include in box plots (although sometimes it is omitted). We indicate the mean score for a group by inserting a plus sign. Figure 4 shows the result of adding means to our box plots.

Figure 4. The completed box plots.

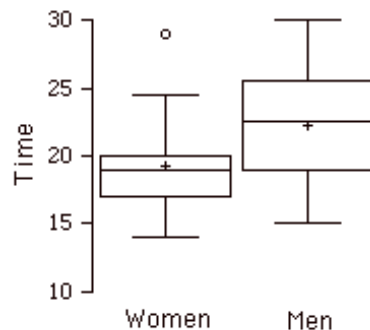
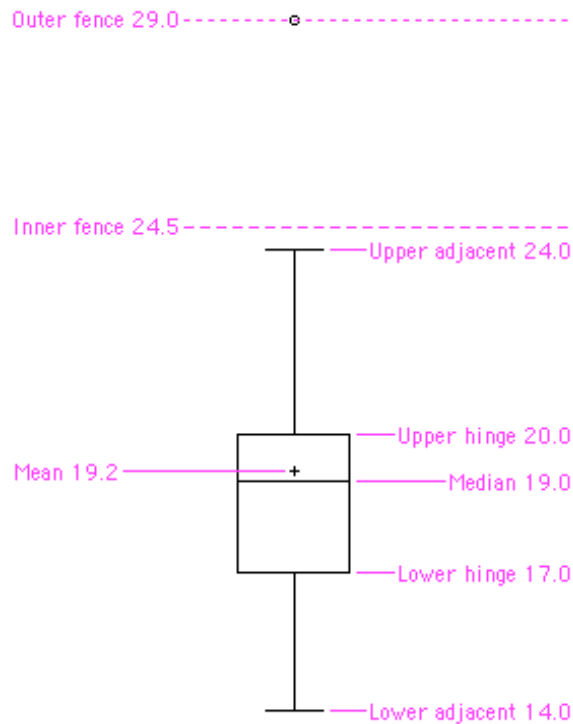


Figure 4 provides a revealing summary of the data. Since half the scores in a distribution are between the hinges (recall that the hinges are the 25<sup>th</sup> and 75<sup>th</sup> percentiles), we see that half the women's times are between 17 and 20 whereas half the men's times are between 19 and 25. We also see that women generally named the colors faster than the men did, although one woman was slower than almost all of the men. Figure 5 shows the box plots for the women's data with detailed labels.

Figure 5. The box plots for the

women's data.



Here are some other examples of box plots.

[Time to move the mouse over a target](#)

[Draft lottery](#)

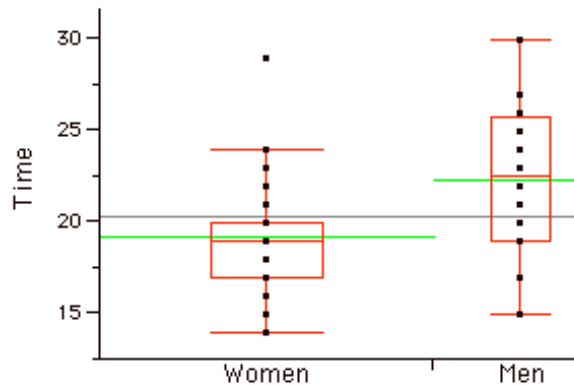
### Variations on box plots

Statistical analysis programs may offer options on how box plots are created. For example, the box plot in Figure 6 is constructed from our data but differs from the previous box plot in several ways.

1. It does not mark outliers.
2. The means are indicated by green lines rather than plus signs.
3. The mean of all scores is indicated by a gray line.
4. Individual scores are represented by dots. Since the scores have been rounded to the nearest second, any given dot might represent more than one score.
5. The box for the women is wider than the box for the men because the widths of the boxes are proportional to the number of subjects of each gender (31 women and 16 men).

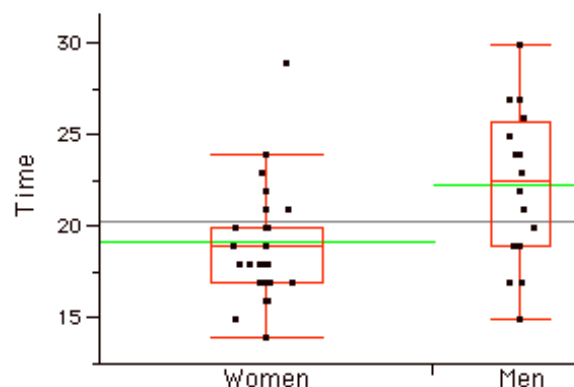
Figure 6. Box plots showing the

individual scores and the means.



Each dot in Figure 5 represents a group of subjects with the same score (rounded to the nearest second). An alternative graphing technique is to *jitter* the points. This means spreading out different dots at the same horizontal position, one dot for each subject. The exact horizontal position of a point is determined randomly (under the constraint that different dots don't overlap). Spreading out the dots allows you to see multiple occurrences of a given score. Figure 7 shows what jittering looks like.

Figure 7. Box plots with the individual scores jittered.



Different styles of box plots are best for different situations, and there are no firm rules for which to use. When exploring your data you should try several

ways of visualizing them. Which graph you include in your report should depend on how well different graphs reveal the aspects of the data you consider most important.