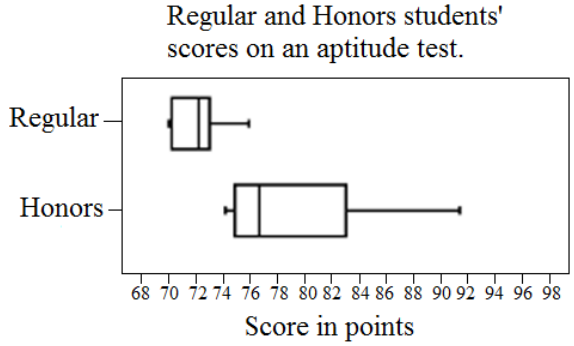
Describing Side-by-side Boxplots

When describing side-by-side boxplots do not use the words "unimodal" or "average". "Unimodal" is reserved for histogram description. "Average" is not the measure of center here, the median is.
Each of the bullets below represents one distinct comparison/contrast idea. The ideas should be fully described with values and units of measure for all boxplots involved.

- Outliers with values or "no outliers"
- IQRs
- Spreads/Ranges
  - You should state whether you are including the outliers or not.
- If one boxplot is "further along" or generally has higher values than another boxplot then percent "chunks" of individuals can be compared between the two boxplots (see the Regular vs. Honors students example below).
  - The percent "chunks" must be in increments of 25%. In other words, 25%, 50%, 75% or 100%.
  - This idea can be used for TWO different comparisons: one comparison at the high end and another comparison at the low end of the boxplots.
- Five-number summaries
- Symmetry/Skewness if applicable (see below)

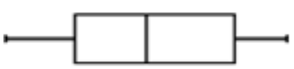
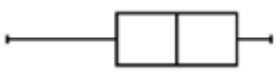


The side-by-side boxplot to the left shows us that

1. 50% of the Honors students scored higher than all of the Regular students.
2. 75% of the Regular students scored less than any Honors student.

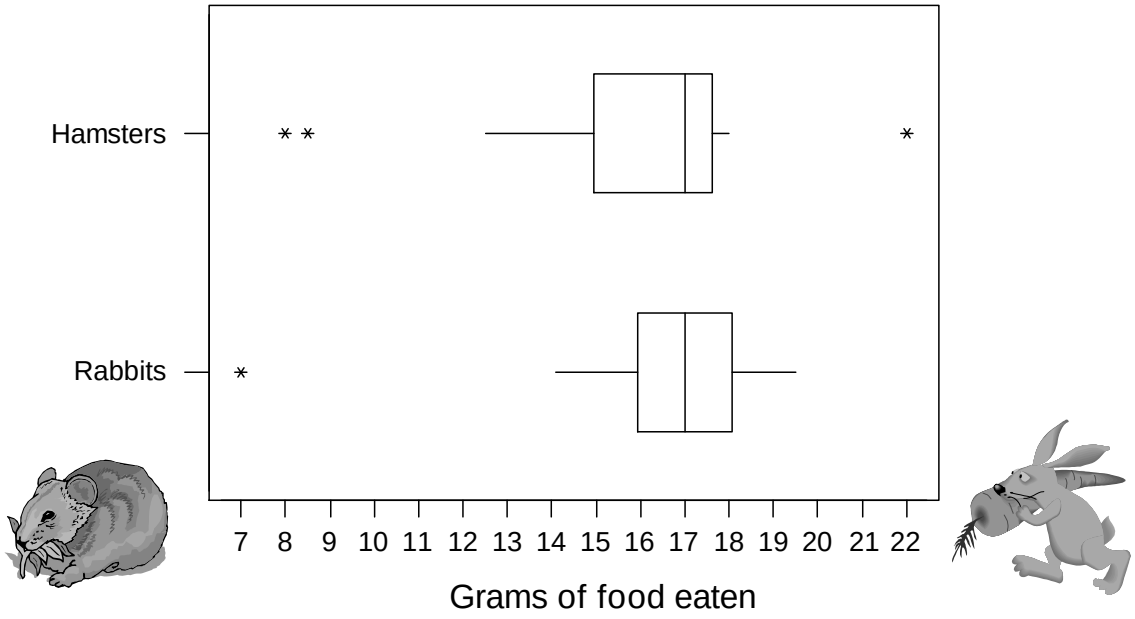
Symmetry/Skewness of Boxplots

It is not always possible to determine symmetry or skewness of data from a boxplot (it is better to view data through a histogram). However, there are some *special cases* of boxplots where this can be seen. The rule is this: if the longer half of the box is followed by the longer whisker then the data is skewed in that direction. If the symmetry/skewness is not discernable from the boxplot then you should not comment on it.

 Symmetric	 Skewed Left	 Skewed Right
 Roughly Symmetric	 Slightly Skewed Left	 Symmetry/Skewness cannot be determined from boxplot alone

Example: *Hamsters and Rabbits*

The manager of a pet store records the amount of food eaten for each hamster and rabbit in the store. The information he obtains is represented in the side-by-side boxplot below. (Note that the pets in the store are the individuals, the type of pet (hamster, rabbit, etc.) is a categorical variable and the "grams of food eaten" is a quantitative variable).



NOTE: Each bullet below represents <u>one</u> distinct comparison/contrast.

- Hamsters have three outliers: two at the low end (8 grams and 8.5 grams) and one at the high end (22 grams). Rabbits have one outlier at the low end (7 grams).
- The Interquartile Ranges (IQRs) of both groups are about the same: hamsters IQR is $17.5-15=2.5$ grams, rabbits IQR is $18-16=2$ grams.
- The spread for hamsters is from 8 grams to 22 grams and the spread for rabbits is from 7 grams to 19.5 grams. The range for hamsters is $22-8=14$ grams which is just a little more than for rabbits $19.5-7=12.5$ grams. If we don't include the outliers we see that the range is the same for rabbits ($19.5-14=5.5$ grams) as it is for hamsters ($18-12.5=5.5$ grams).
- If we exclude the hamster that ate 22 grams, then we can say that 25% of the rabbits ate more than *any* hamster (Q_3 for rabbits is the same as maxX for hamsters, again excluding the high outlier for hamsters).
- The median grams of food eaten by rabbits is the same for hamsters (17 grams), however both Q_1 and Q_3 for rabbits are higher than Q_1 and Q_3 for hamsters, respectively. Also, the rabbits' minX and maxX are higher than the minX and maxX for hamsters, respectively (ignoring outliers, otherwise the reverse is true!).
- The distribution for hamsters is definitely skewed left, whereas the distribution for rabbits is fairly symmetrical.