

Statistical Reasoning Box and Whisker Plots

Name: _____

Date: _____

Class: _____

Box and Whisker Plots

Five-number summary

A summary of information to approximately describe a distribution. It identifies the following:

- **Maximum and Minimum** — give bounds for the data set — all values fall between
- **Median** — a measure of center, middle it
- **Q_1 and Q_3** — measure of variability: $IQR = Q_3 - Q_1$ width of box

Can use median and quartiles together to get an indication of shape:

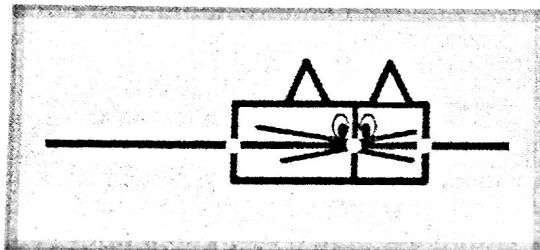
- may be skewed if median is much closer to one of the quartiles or if whiskers are not the same length (approximately)

Should not be used alone as indicator of shape of distribution—not enough detail—hides some features such as bimodality. Best graphs to find shape of distribution:

histogram, stem plot, dot plot

To draw "full" or "modified" boxplot (generally used):

1. Identify 5 number summary (min Q_1 med Q_3 max)
2. Calculate the IQR
3. Calculate the "thresholds." The upper and lower limits where data is considered outliers:
 - a. $Q_3 + 1.5(IQR)$ and $Q_1 - 1.5(IQR)$
 - b. Don't draw thresholds on boxplot—they are not data values. Only use them to identify outliers.
4. Identify which numbers are outliers
5. Construct a Box and whisker plot
 - a. Draw a scale, or number line, that accommodates all data values. Include numbers and units. The plot may be vertical or horizontal.
 - b. Draw rectangular box with ends at quartiles and a line through box at median.
 - c. Draw two "whiskers" from corresponding ends of box to most extreme data value that is not an outlier—inside thresholds. Put dots or other marks for each outlier value.



class was
 ate the number of stu
 before taking a major exam.
 follows:

are as follows:

Student | Hours of

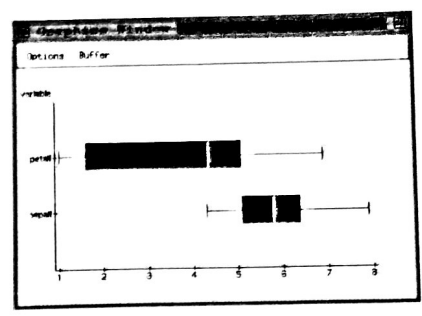
Stu

Why 1.5 x IQR?

John Tukey, inventor of boxplots, answered that one was not enough and two was too much.

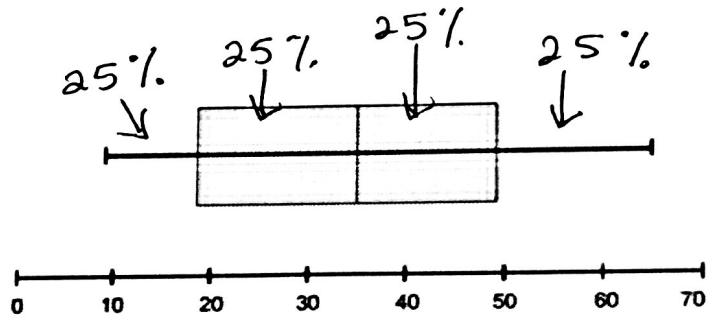
Multiple boxplots

- useful to compare several boxplots of related data on same scale
- Small differences may be meaningless, especially with small data sets.



Percentiles

- Q_1 and Q_3 are the 25th and 75th percentiles respectively.
- Q_1 —25th percentile means 25% are below Q_1
- Q_2 —median—50th percentile
- Q_3 —75th percentile means 75% are below Q_3



Anatomy of a Typical Box-and-whisker

