Chapter 5 Lecture Notes

Understanding and Comparing Distributions

Plotting the 5-Number Summary

Recall that the 5-number summary consists of

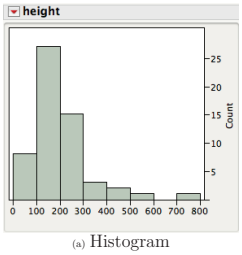
Min. Q_1 Median Q_3 Max.

How to draw an (outlier) box plot:

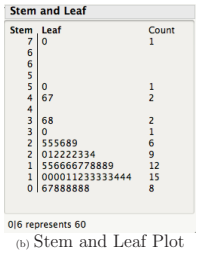
1. Draw a plot line long enough to contain all of the data.
2. Mark the median and quartiles (Q_1 and Q_3).
3. Make a box from Q_1 to Q_3
4. Draw a Line (Whisker) extending up from the box to the *smaller* of either
   (a) The Maximum Value or
   (b) $Q_3 + 1.5\times$ IQR
5. Draw a Line (Whisker) extending down from the box to the *larger* of either
   (a) The Minimum Value or
   (b) $Q_1 - 1.5\times$ IQR
6. Data value farther than $1.5\times$ IQR from the box are denoted as dots and considered possible outliers

Example: Heights (ft) of Bungy Jumps in the USA

Minumum	Q_1	Median	Q_3	Maximum
60	115	160	234	700



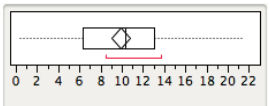
(a) Histogram



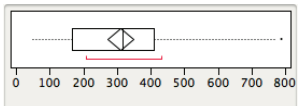
(b) Stem and Leaf Plot

Interpreting Boxplots

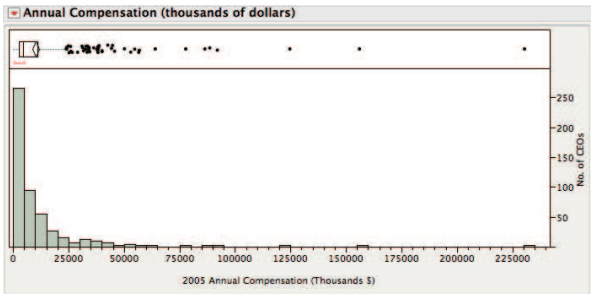
- Symmetry



(c) Symmetric



(d) Skewed



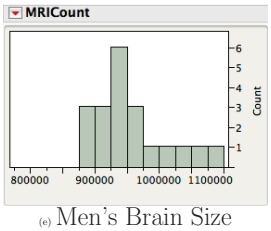
- The distribution seems to have outliers — but could it just have a long tail on the right?
- *Outlierhood* is a judgement that we must make and defend if necessary.
- What do we do with outliers?

Comparing Groups -

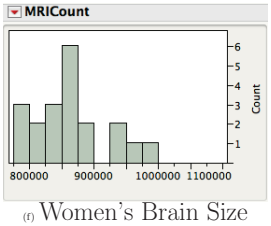
- Shape
- Center
- Spread
- Outliers

Comparing Groups - with Histograms

Willerman et al. (1991) collected a sample of 40 right-handed Anglo introductory psychology students at a large southwestern university. Subjects from Willerman et al. (1991) took four subtests (Vocabulary, Similarities, Block Design, and Picture Completion) of the Wechsler (1981) Adult Intelligence Scale-Revised. The researchers used Magnetic Resonance Imaging (MRI) to determine the brain size of the subjects. Information about gender and body size (height and weight) are also included.



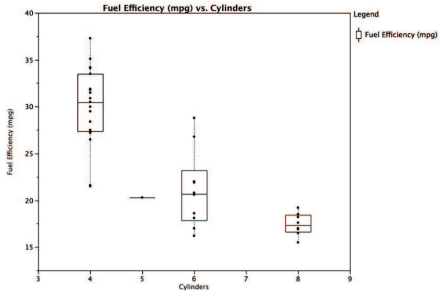
(e) Men's Brain Size



(f) Women's Brain Size

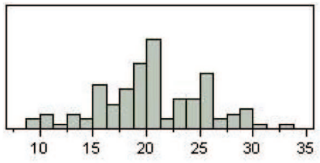
Comparing Groups - with Side-by-Side Boxplots

Fuel Efficiency (mpg) vs. Cylinders

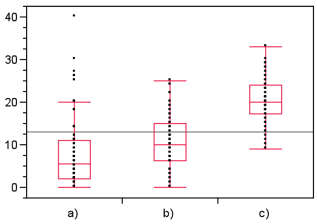


Is there an association between Fuel Efficiency and Cylinders?

Which size engine has the least variability in its fuel economy?



(g) Histogram



(h) Which is the matching Boxplot?