

NOTES FOR THE WEEK OF OCT 9 TO OCT 16
(Reading assignments and Exercises to hand in are at the bottom of the page)

Last week you learned how to collect good data. The main theme of the material this week is how to describe a set of data once you have it. There are two basic ways to describe data – *pictures* and *numerical* summaries. There are also two basic types of data – *categorical* (male, female; smoker, non-smoker, etc) and *quantitative* (height, income, etc.). Both types of data can be described with both methods - *pictures* and *numerical summaries*.

Describing *categorical data* is easy – you just **count how many individuals fall into each category** (male, female) or combination of categories (male smoker, male non-smoker, female smoker, female non-smoker). *Bar graphs* and *pie charts* can be used to **visually summarize** these counts.

Describing *quantitative data* is more complicated, because there are several features that may be of interest. They generally fall into four categories – center (**median, mean**), spread (**range, interquartile range, standard deviation**), shape (**bell-shaped, skewed, bimodal**) and outliers (**unusually small or large values**). *Pictures* give information about all of these features, especially *histograms* and *stem and leaf diagrams*. *Boxplots* provide information about center, spread and outliers, and can give some indication of skewness.

Outliers are of particular interest because they may indicate something unusual and interesting about the situation. Make sure you read and understand how to define outliers (p. 27), their influence on the mean and median (p. 40), how to identify them (p. 43) and how to handle them (Section 2.6). Not dealing correctly with outliers is a common misuse of statistics.

Shape is also important. In particular, it is important to identify bell-shaped datasets because many statistical methods encountered later in the course require that a dataset be approximately bell-shaped. Make sure you read and understand Section 2.7.

Reading and Study Assignment for this week:

Book Chapter or CyberStats Unit	Focus on:
Chapter 2: Turning Data Into Information	All, but especially topics covered by assigned exercises, and the material above in these notes.
Unit A5: <i>Describing Data Graphically</i>	Basics 1 and Basics 2
Unit A6: <i>Describing Data Numerically</i>	Uses 3 and 4; Self-assess test; Interactivities listed below

Interactivities to play with:

Unit A5: Basics 2

Unit A6: Basics 2; Basics Practice 2 (Question 9), Basics 3, Uses 4

Exercises to hand in (Due Oct 16):

Chapter 2: 5, 13, 43bdef, 49, 61, 75, 76, 84ab, 91, 96, 112