Suggestions for students (to help you realize what you want to learn):

Aim to keep (extensive) ***notes*** *about what you've learned.* I am thinking about notes partly in class, but largely out of class as you study. Several aspects should be taken into account: general, often conceptual information about statistical applications, or more generally, research issues; and specific info about graphics, and R coding, such as the following:

a. to describe data, not just with words but using visualization and graphics; specifically, to use of statistically sound language to report what has been learned, such as from univariate plots (boxplots, stem and leaf, dotplots, density plots); also single vs. composite plots (cf. datadensity function in R [Hmisc package]), panel displays (pairs), pairs2 [TeachingDemos package], gpairs [YaleToolkit package], missing value information (use function help.search), scatterplots (done several ways) and various scatterplot matrices (e.g., splom, lattice); predictive relationships, from the simple to the complex, linear and non-linear. Also, loading, editing, and cleaning data.

b. comparing groups on outcomes (for 2, and more than 2, groups) or examining relationships among variables (two at a time, and more); be sure to reference the (research) question(s) that drive the comparisons, and the knowledge base that is most relevant. (The granova package should have particular value here.)

c. Prediction: linear and non-linear; one predictor or more; outliers, transformations of variables, roles of anomalies and re-expression in numerical and graphical applications. Standardization of variables as well as when, how (in R) to do, citing pros and cons for different purposes.

d. Be able to conduct formal inferences (especially w/ confidence intervals and contrasts) remembering that thought should often be given to various audiences when results are interpreted. Effect sizes should also be understood, as well as inferences (based on bootstrapping, which is another topic we should discuss).

e. R functions, packages and code; so for each of a. - d. items above; your notes should reflect how effectively to use R to generate graphics with the features you seek them to have. And remember to spell out where you need help to get to especially helpful displays; and what help worked best (manuals, books, students, instructor, etc.).

In nearly all real data applications you should ask: What are MY main questions? Think about alternative ways of expressing purposes, alternative ways to analyze and plot data, etc.? By writing your questions out, and comparing what and how you ask questions early in the course to those you ask later, you will get useful information about changes in yourself, that is, to be able to *self-asses* – a most important matter. You can seek answers through our wiki, your reading, etc. Start conversations both in and out of class when this seems likely to be helpful.

Ask any q's you deem relevant to getting the most out of the course, out of R, vis-à-vis both numerical and graphical data analyses, and more generally about statistical questions that seem close to what we do in class or to your own special interests outside of this class.

I will often (typically at the end of class) ask you to write a short paper (about 2 minutes) that asks you what the most important points were for the day, or week. Think ahead about what you will want to say!