Stat 998, Fall 2012 (Larget)
General Guidelines for Analyzing Data
(based on 1973 notes of Brian L. Joiner)

- **How were the data gathered?**
  From a designed experiment? Or from an observational study?
  Was randomization used? How?
  What is/are the experimental unit(s)?
  Were measurements made in quick succession or under a limited set of conditions?
  Are there hidden confounding conditions?

- **Why were the data gathered?**
  What sort of results were expected?
  Are there primary and secondary goals?
  Any interest in finding patterns and results not anticipated in the original plan?

- **Plot the data.**
  Try several ways: versus time, treatment or blocking factors, location, technician, etc.
  Use interaction plots, superimposed plots, matrix plots, etc.
  Does there seem to be support for the hypotheses of interest?
  Do unexpected patterns arise leading to new or revised hypotheses?
  Do suspicious patterns (trends, unequal variance) question assumptions?

- **Fit the data as needed.**
  Start with simple models, perhaps with subsets of the data.
  Later consider more complicated models if necessary.
  Guide analysis by the needs of the client and the way the data were gathered.
  Simpler models that serve a client's needs are better than complicated ones.
  (Sometimes a t-test or confidence interval is all that is needed.)

- **Check the assumptions of each model you try.**
  How reasonable are assumptions based on what the experimenter knows?
  How reasonable are assumptions based on how the experiment was conducted?
  Check residuals for each model you fit.
  Plot residuals in various ways: versus predicted values, time order, other variables.
  Look for outliers and autocorrelation.
  Do the residuals suggest the need for a transformation or addition of new variables?
  Should a weighted analysis be used?
  Modify models as needed and recheck.
  (You may need to repeat this step several times.)