Today:

Feedback from presentations
Review for Final Exam
(See review Oct 15 for course material up to that time)

Feedback from Presentations

Generally of very high quality!  Good Q&A afterward too.

Presentation tips:

Face the audience!
Slides: Keep words to a minimum (hard to read and listen simultaneously)
Objectives clear at start
Start with an example, details later
Use graphical explanations where possible (visualization!)
Recap main points at end.

Ch 5 & 6

Ch 5 Hypervariate Data

Role of Variables
Matrix Plots and Visual Linking (Brushing)
Four Variable Coplots ( 1 dept + 3 indept)
        Problem with equal count algorithm (p 133 and p 278)
        Looking for interactions in columns, rows and along diagonal of plot matrix
        Use of 3 coplots for same data
        Verbalizing the visualization
Cropping – simple and complex – needed?
Coplots of hypervariate surfaces – how different from coplots of hypervariate data
s-l plots of residuals – purpose and consequences
q-q plots of residuals – purpose and consequences
Simplification using Indexes ( e.g. Iris Elongation p 300)

Ch 6 Multiway Data

Categorical Variables and Counts
Transformations of Counts (positive measurements are often skewed right)
Making use of unspecified characteristics of a graph (like order of labels, order of panels)
Median ranking for ordering – robustness
Modeling Counts – additive and multiplicative fits – use for residual analysis
Spatial (or geographic) location conditioned on frequency ( Fig 6.11 p 319)

Sections 6.3 can be omitted.  6.4 relates to barley data in book intro.

Assignment Lessons:

Coplot (Galaxy)
Possible to do it but not as good a level plot for visualizing rotation info (which was of interest in this context)

Ozone
Traditional analysis can find interactions if there are enough degrees of freedom , but it is hard to describe what the interactions are from the fitted linear model.  (Visualizations better for this.)
Stepwise methods not too useful in this context – better to include all variables in the descriptive model, at least for a first look.
Note difference in description of relationships and explaining the relationships – may need different approaches for these goals.

Statcan
Data is free to researchers.  Can be quite detailed and quite interesting.  Graphical description is the key to summarizing these relationships.  Note much of "official statistics" data is not sample data, but population data. And lots of time series. Nevertheless, variability questions arise even with population data.

Bootstrap
A very useful general purpose technique to use when the situation is complex or when the population is completely unknown.

, Statcan, Bootstrap