ESS 595B Scientific Writing and Graphics

2009 November 25

Discussion of:

Chapter 3, Cleveland, W.S. (1994). *The elements of graphing data,* Hobart Press.

Reporter: Ryan Eastman

Each section shown in red represents a sub-chapter discussing a new type of plot or figure in chapter 3 of Cleveland's *The elements of graphing data*. Listed below are points made by discussion leaders or by members of the class while discussing examples of the plot types.

Our goals are: 1) to decide the appropriate type of figure to use for a given type of data and a given goal in expressing that data, and 2) to discuss and determine the best format for the figures.

Log Plots

- Provides improved resolution and distribution of some data sets.
- It is important to choose the base so tick mark values can be easily distinguished.
- Steve suggests using the actual values versus log values on the axis labels.

Residuals

- Allows the reader to see patterns in deviation from a best fit line or curve.
- Be careful using this, because it can ask the reader to do a lot of interpretation.

Distributions

- Scatter plots, histograms, quantiles, box plots, Q-Q
- Choose your distribution figure type based on what you would like to emphasize. (e.g. to compare data, to see a trend in data, to read exact values, etc...)

Dot Plots

- Would a table be more useful?
    - It depends on whether you want to show a distribution of values, or give exact values.
    - Perhaps use a hybrid showing numbers distributed on a plot, provided the figure is not too cluttered.
- A histogram will not replace a dot plot if a log scale is involved.

- Multiway dot plots:
    - Choose these based on what aspects of data you would like to show.
    - Often a bar graph or series of bar graphs may work better, though a dot plot may appear less cluttered.

Symbols and Curve Types

- Avoid overlap by: Moving points slightly, using sunflowers, jittering (adding noise).
    - Move data only a small amount compared to your uncertainty.
- Superpose symbols and curves using different types or colors to show overlapped data better.
- Compensate for less prominent colors with thicker lines and symbols when using this method.

Reference Grids

- These are used to enhance comparison between multiple panels plotting different versions of the same variables.
- Be reasonable with the number of panels shown, too many is not a good thing.

Loess (locally weighted regression)

- Be careful when choosing alpha so that the Loess curve fit to the residuals is flat.
    - The class asks how necessary is it to get this curve perfect if the nature of the relationship is shown either way?
    - Is it even necessary to show this curve, given that a pattern may already be clearly shown in the plot?
- If this curve is non-linear, perhaps consider a log plot.

Time Series Plots

- Choose line & dot, line, dot, or bars based on desired properties you wish to highlight:
    - To emphasize long period variation when exact values aren't needed to be shown, use a line plot.
    - To show exact values or shorter periods of variation, use dots or bars.

- Many types of time series plots:
    - Cycle plots
    - Step function plots
    - Stack plots (broken time series instead of a long one)
    - Scatter plot matrices
        - Doesn't explicitly show one time series, but compares multiple plots to each other.
        - Be careful with how many plots shown, more than 4x4 or 5x5 becomes challenging for the reader.

Showing Variation in Data

- Whisker plots are preferred to show not only scatter in the y data, but also the relationship of y with x.
    - Whisker plots are referred to straight error bars, since different distributions can create the same sized error bars.
    - This brings up the question: is it necessary to assume a normal distribution around the mean?

Graphical Perception

Cleveland, W.S. and R. McGill. 1985. Graphical perception graphical methods for analyzing scientific data. *Science* 229, 828-833.

- Tasks:
    - Judge & compare scales.
    - See slope and angles and how they vary over time or space.
    - Compare lengths, magnitude, and position along a common scale.
- Plot Analysis notes:
    - When showing a plot of a variable versus day, such as day 1, day 2, day 3, etc... label the space between ticks instead of the ticks, so time of day can also be inferred by the reader.
    - Be careful making a figure too pretty and cluttering up the window.
        - This prettiness may draw people to your figure, though, so be smart about it.
        - Find a balance to make the reading experience interesting, but not overly challenging.