

CS 5630/6630

Scientific Visualization

Elementary Plotting Techniques II

Motivation

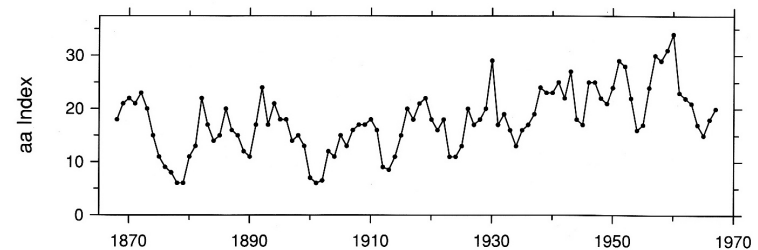
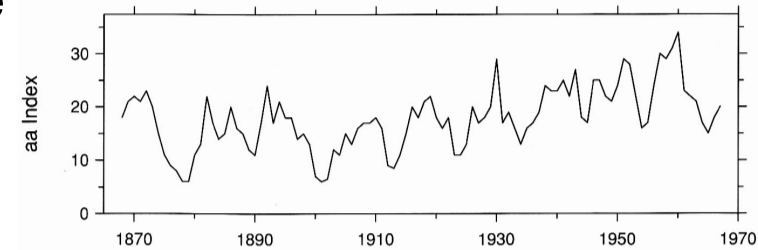
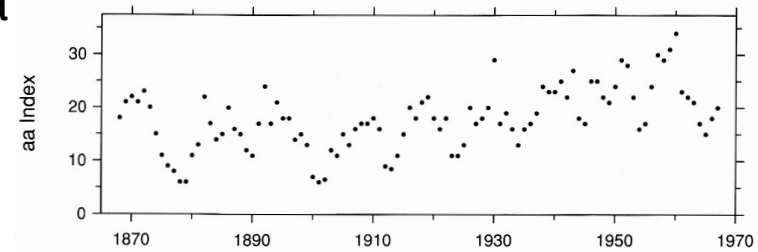
- Given a certain type of data, what plotting technique should I use?
- What plotting techniques should be avoided?
- How do I encode additional information in my plot?

Summary

- Basic Plotting
 - Connected Symbol Plots
 - Dot Plots
 - Scatter Plots
 - Histograms
 - Others
- Advanced Plotting
 - Multimodal Data
 - Higher Dimensional Data
 - Correlation
 - Uncertainty and Variation

Basic Plotting

- Connected Symbol Plots
 - Used for graphing a time series or other 1D data
 - Symbols, connections, or connected symbols can be used
 - Symbols: High frequency data (spikey) where only the low frequency trend is important
 - Connections: Low frequency data (smooth) where points do not add additional information
 - Connected Symbols: In between data where the points can show concentrations of data and the connections can show the trend of the data



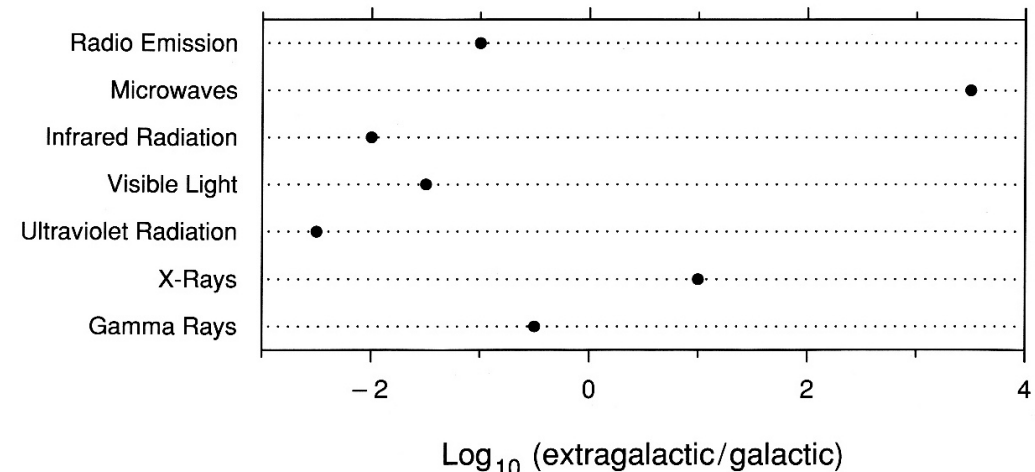
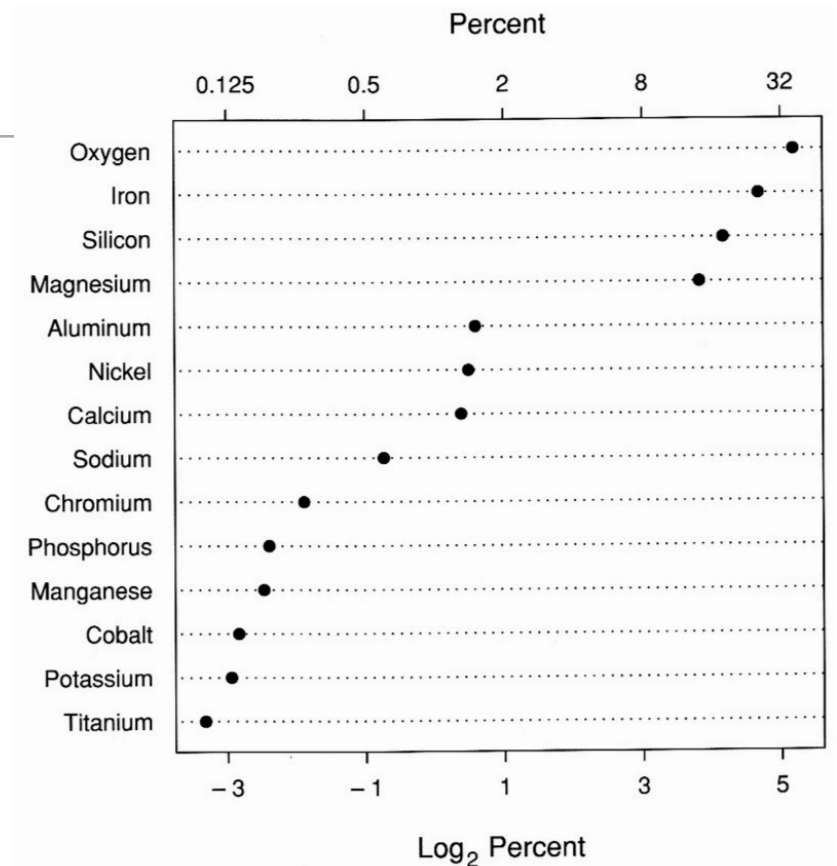
Basic Plotting

- Connected Symbol Plots

Mauna Loa VisTrails Demo

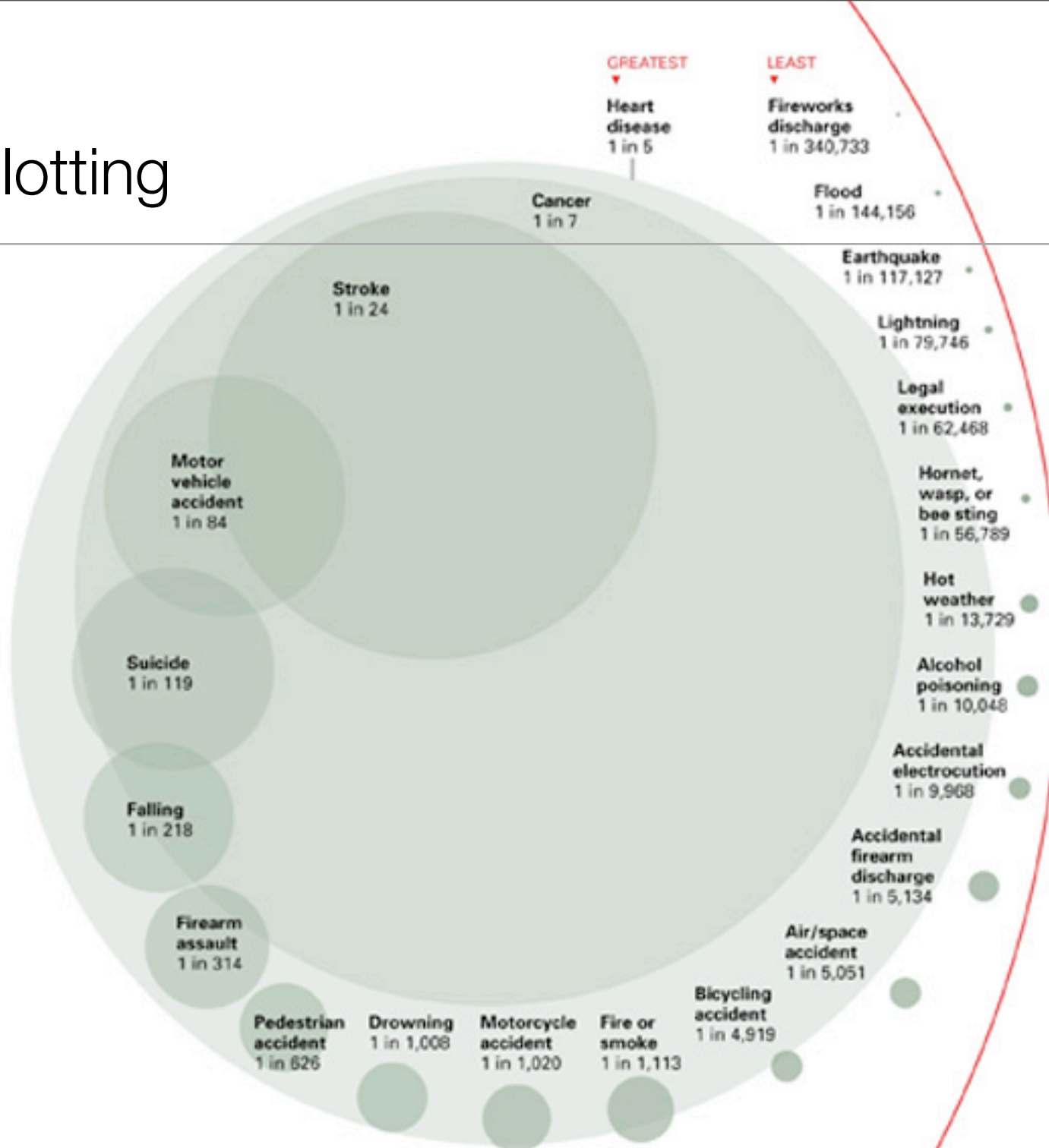
Basic Plotting

- Dot Plots
 - Used for quantitative, labeled data
 - Similar to the more familiar bar charts and pie charts
 - Order the plot in one of two ways:
 - Data: Sort from highest to lowest going from top to bottom
 - Label: Sort by label if it has an inherent order



Basic Plotting

- Dot Plots



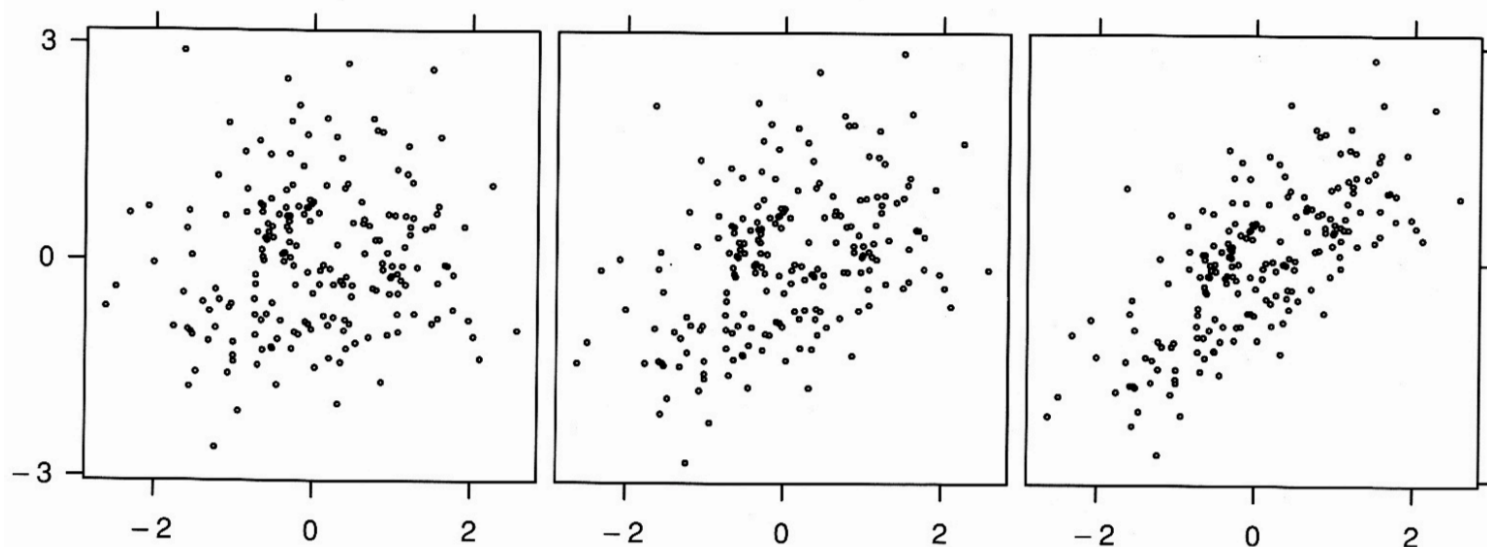
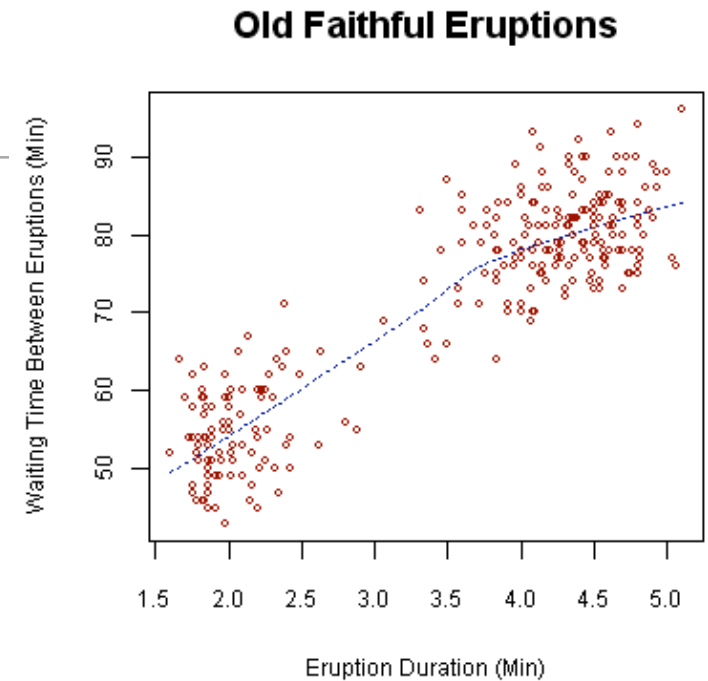
Basic Plotting

- Dot Plots

Death Rate VisTrails Demo

Basic Plotting

- Scatter Plots
 - Used to show how one variable is affected by another (correlation) in 2D data
 - Works well for lots of data samples
 - High vs Low vs. No Correlation



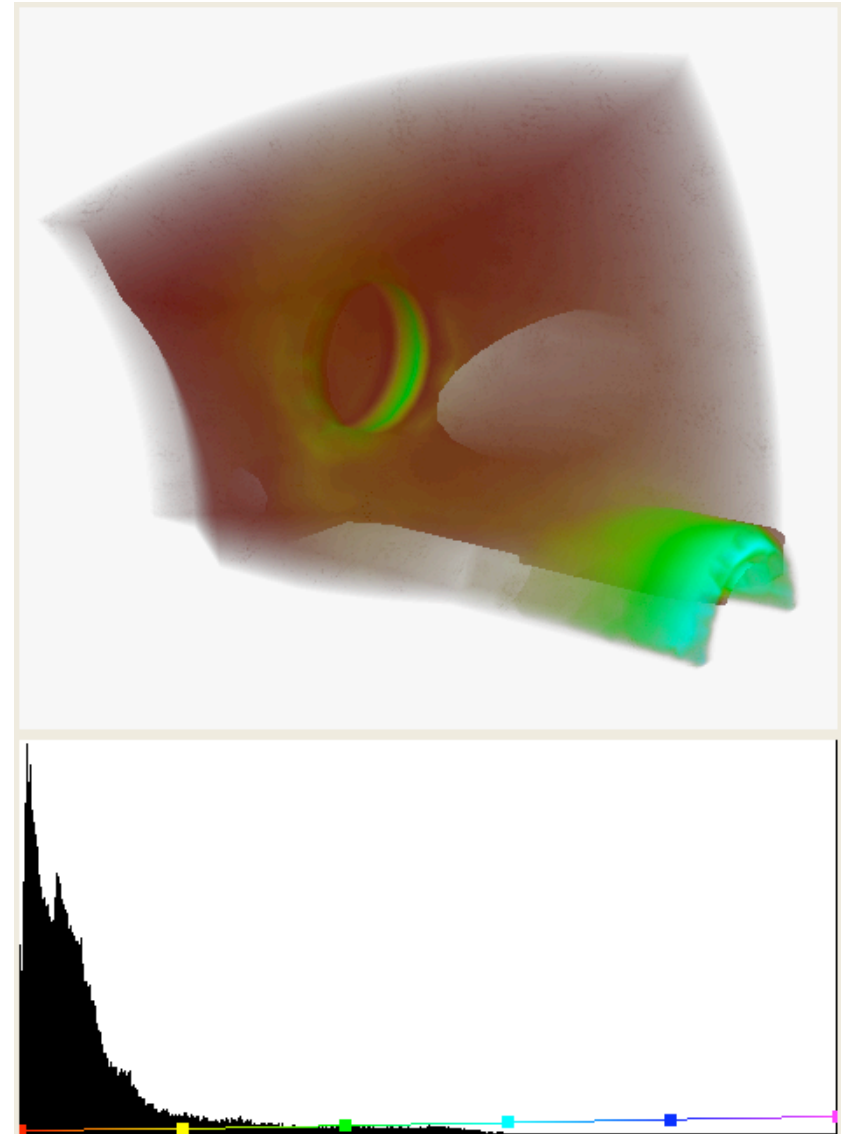
Basic Plotting

- Scatter Plots

Mammal Scaling VisTrail Demo

Basic Plotting

- Histograms
 - Used for analyzing distributions in large amounts of quantitative data
 - Horizontal axis is a set of bins (intervals) of the data
 - Vertical axis is the number of entries in the bin
 - Reduces the amount of data, interval selection is important



Basic Plotting

- Histograms

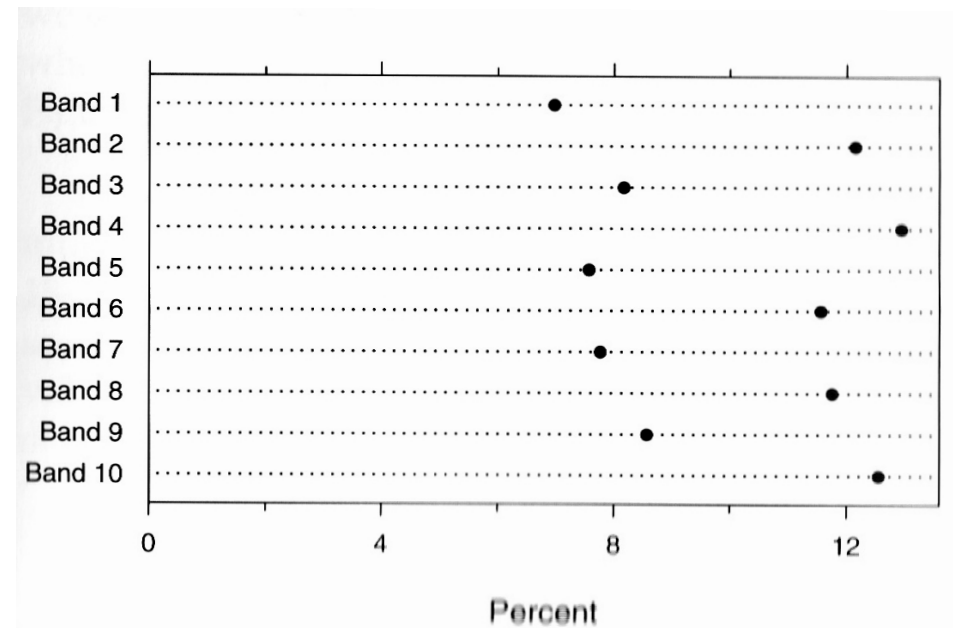
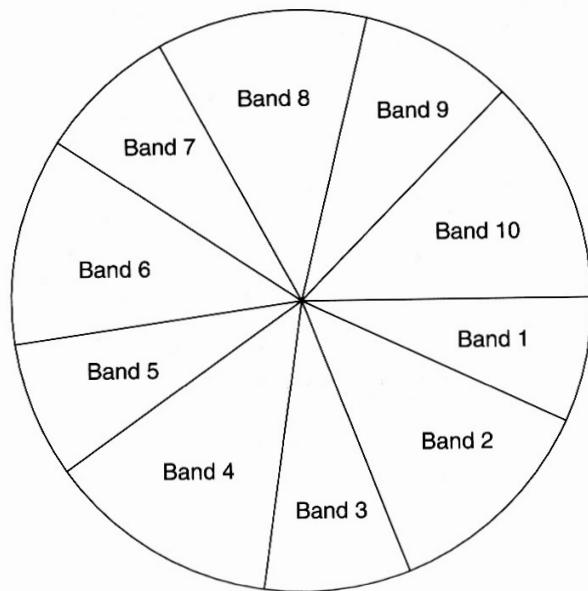
Terminator VisTrail Demo

Basic Plotting

- Others

- Pie Charts

- Don't use for scientific data, use a dot plot instead
 - Poor pattern perception: judging area is difficult!

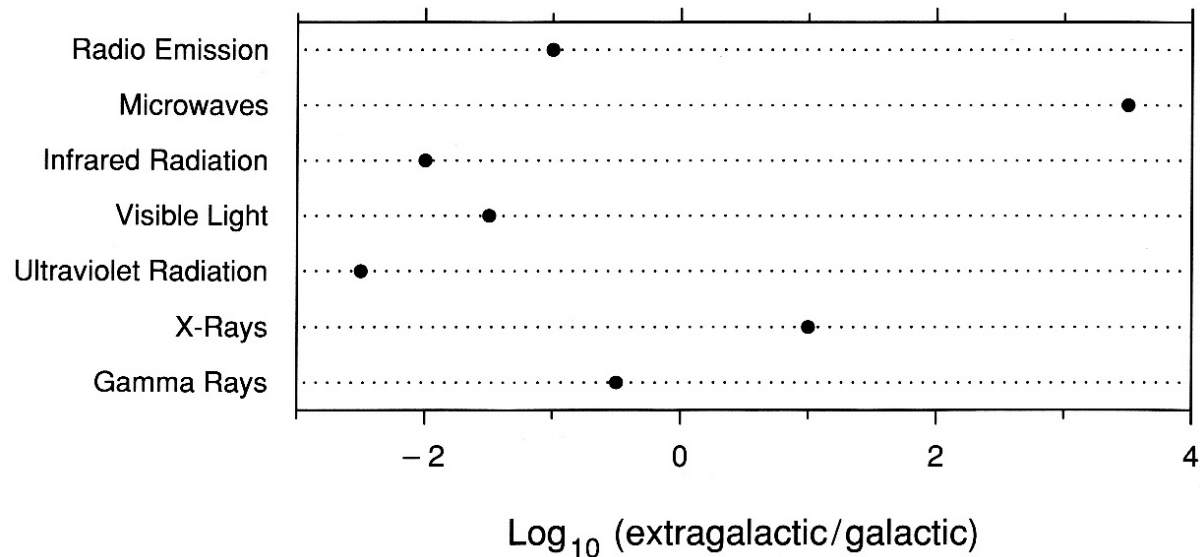


Basic Plotting

- Others

- Bar Charts

- Don't use for scientific data, use a dot plot instead
 - How do you show data that does not have a zero baseline?



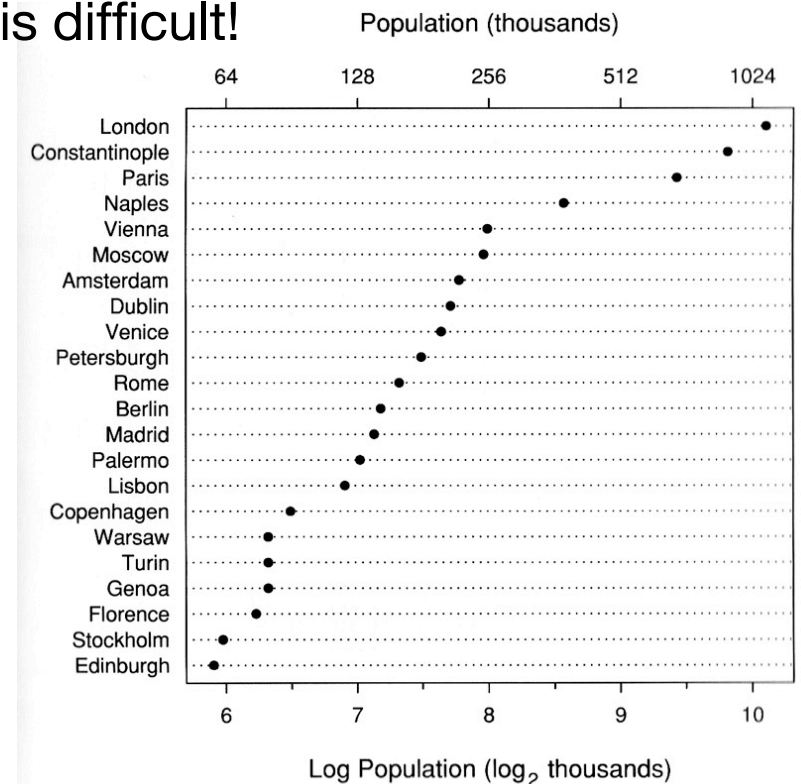
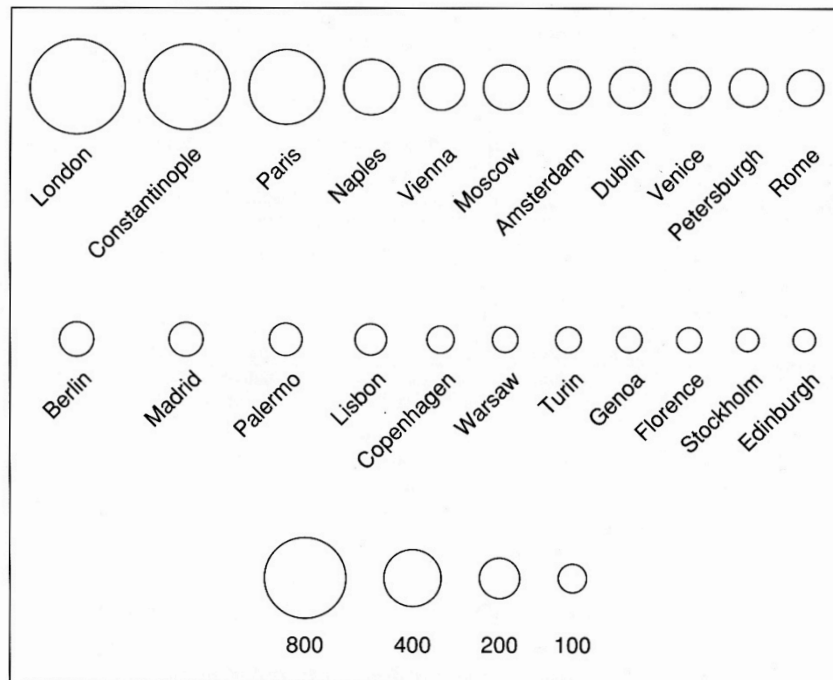
Basic Plotting

- Others

- Area Charts

- Don't use for scientific data, use a dot plot instead

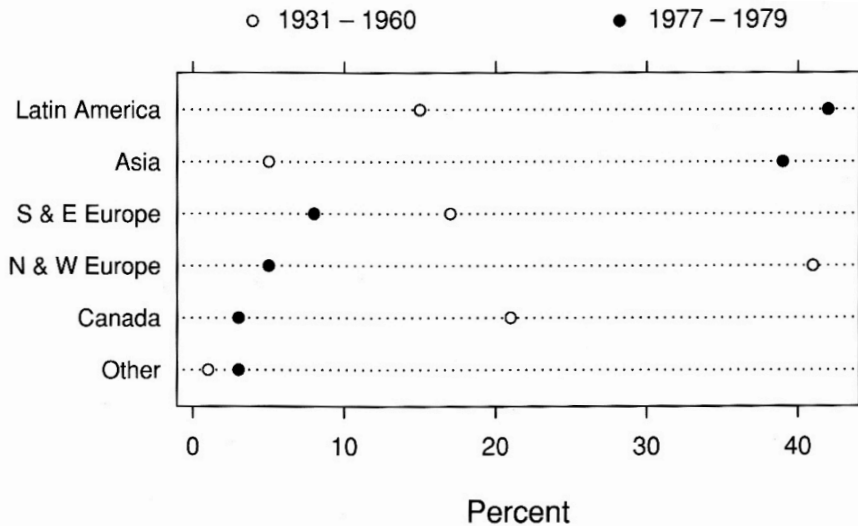
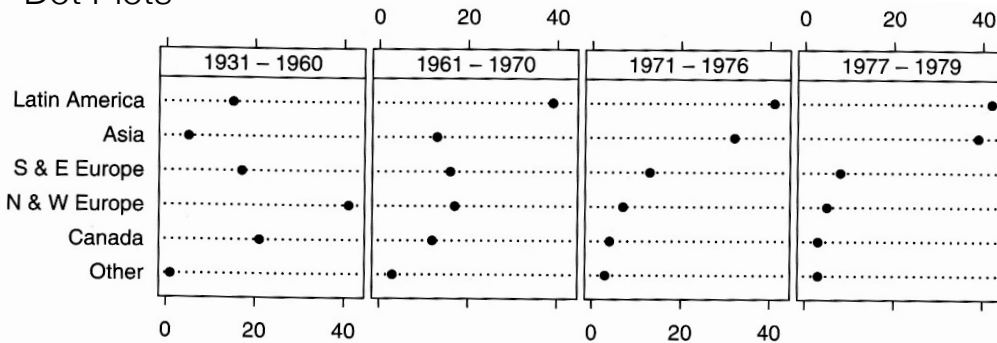
- Poor pattern perception: judging area is difficult!



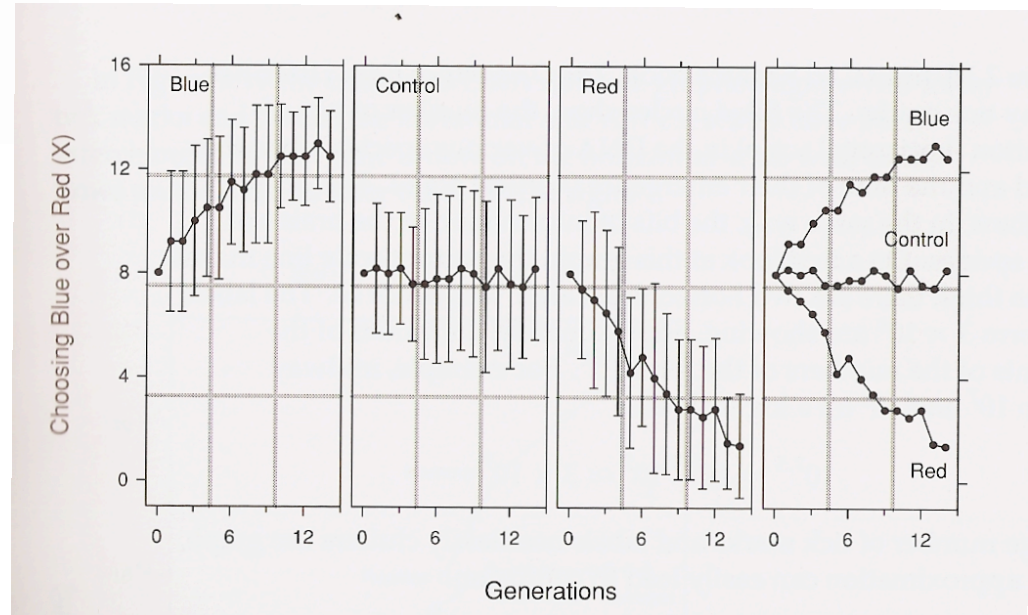
Advanced Plotting

- Multimodal Data
 - Juxtaposed vs. Superposed

Dot Plots

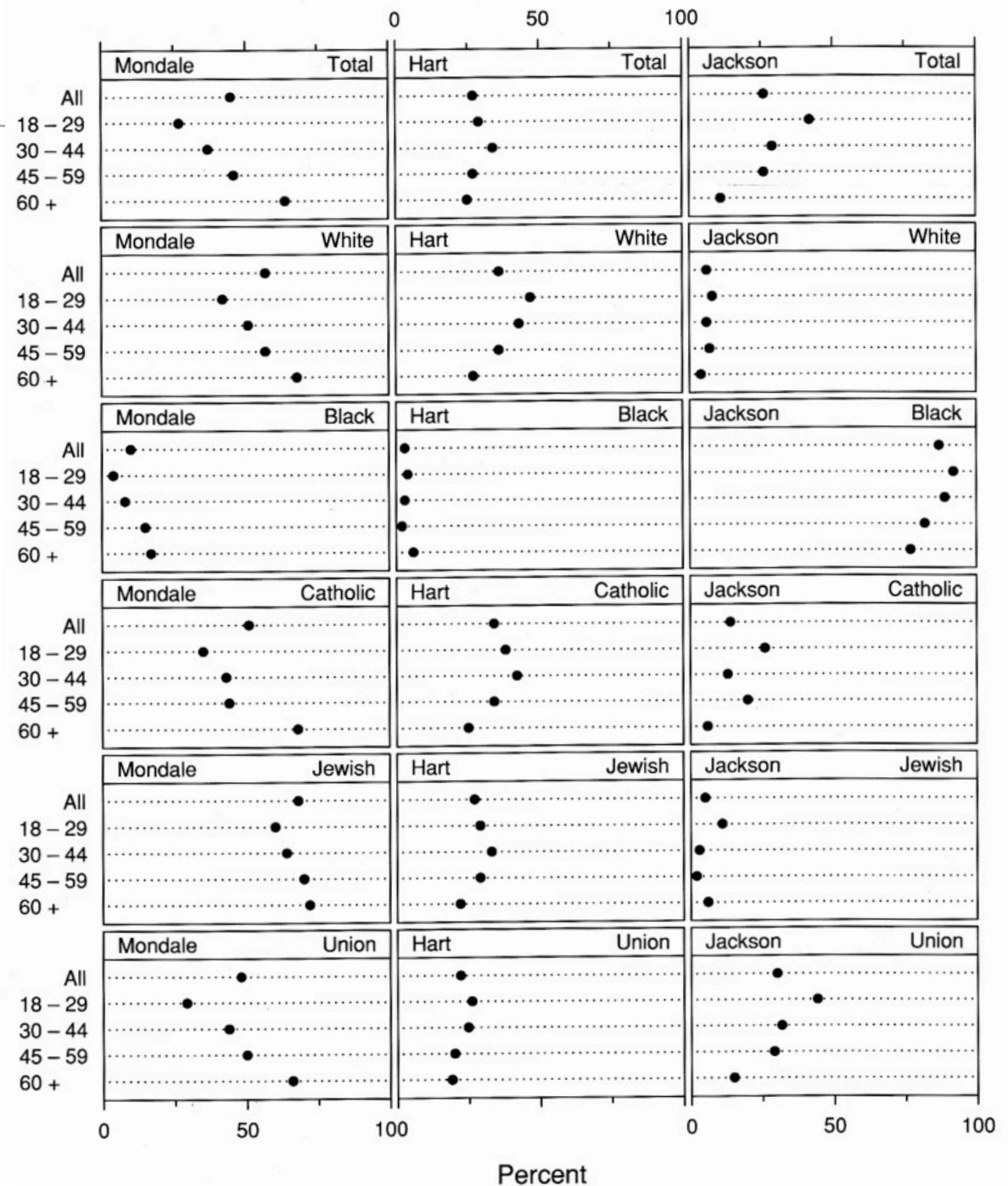


Symbol and Connection Plots



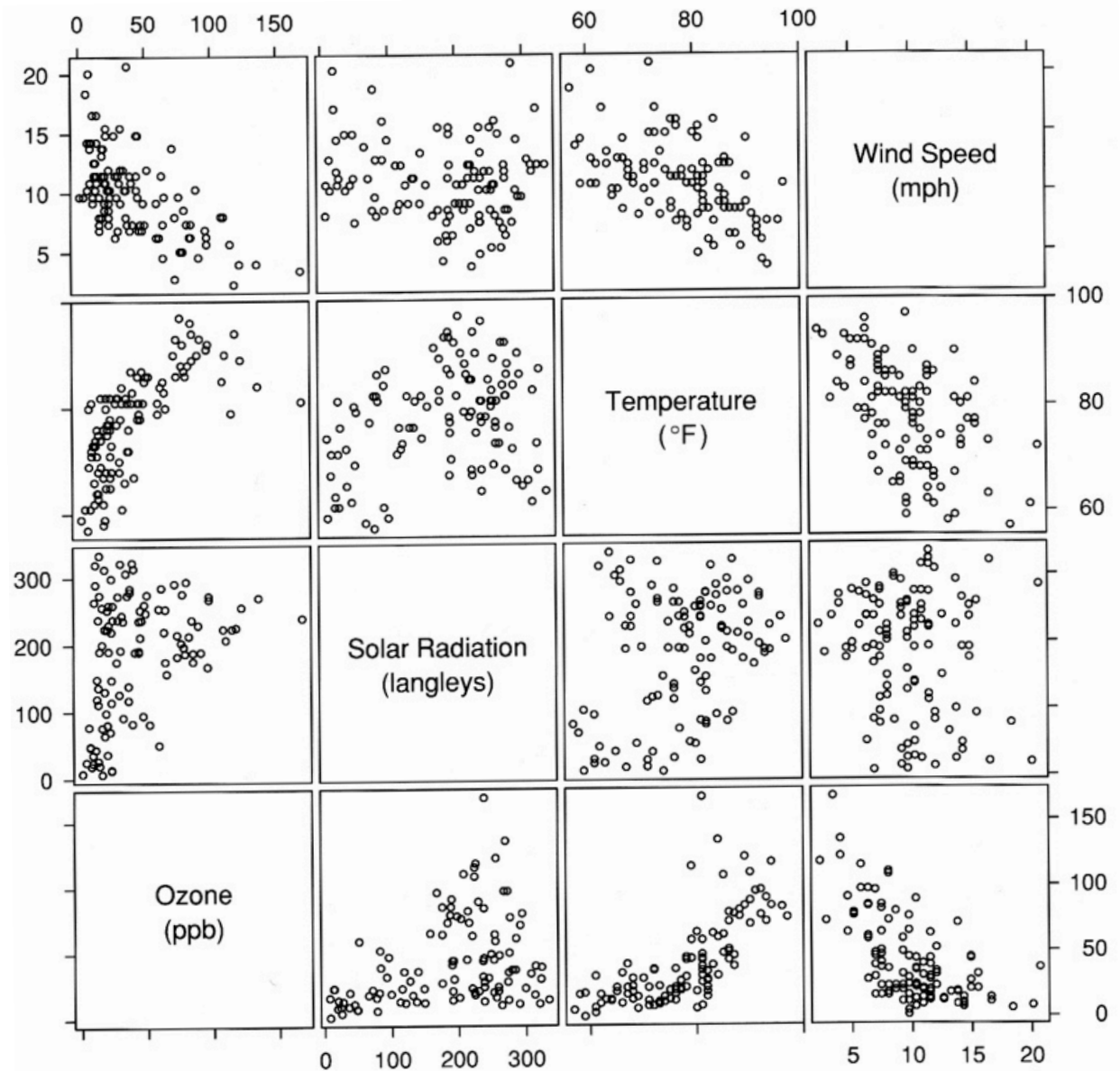
Advanced Plotting

- Higher Dimensional Data
 - Multiway Dot Plots



Advanced Plotting

- Higher Dimensional Data
 - Scatterplot Matrices



Advanced Plotting

- Correlation

- Linear Regression using least squares

- Find the regression line: $y = a_0 + a_1x$

- Where the summed squares of the vertical distances: $\Delta = \sum_0^n (y_i - f(x_i))^2$

- And the best parameter set for the fit is achieved when the sum of the squares of the distance Δ is minimal for the approximation:

$$\frac{\delta \Delta}{\delta a_i} = 0$$

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_0 \\ 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Advanced Plotting

- Correlation
 - Linear Regression using least squares

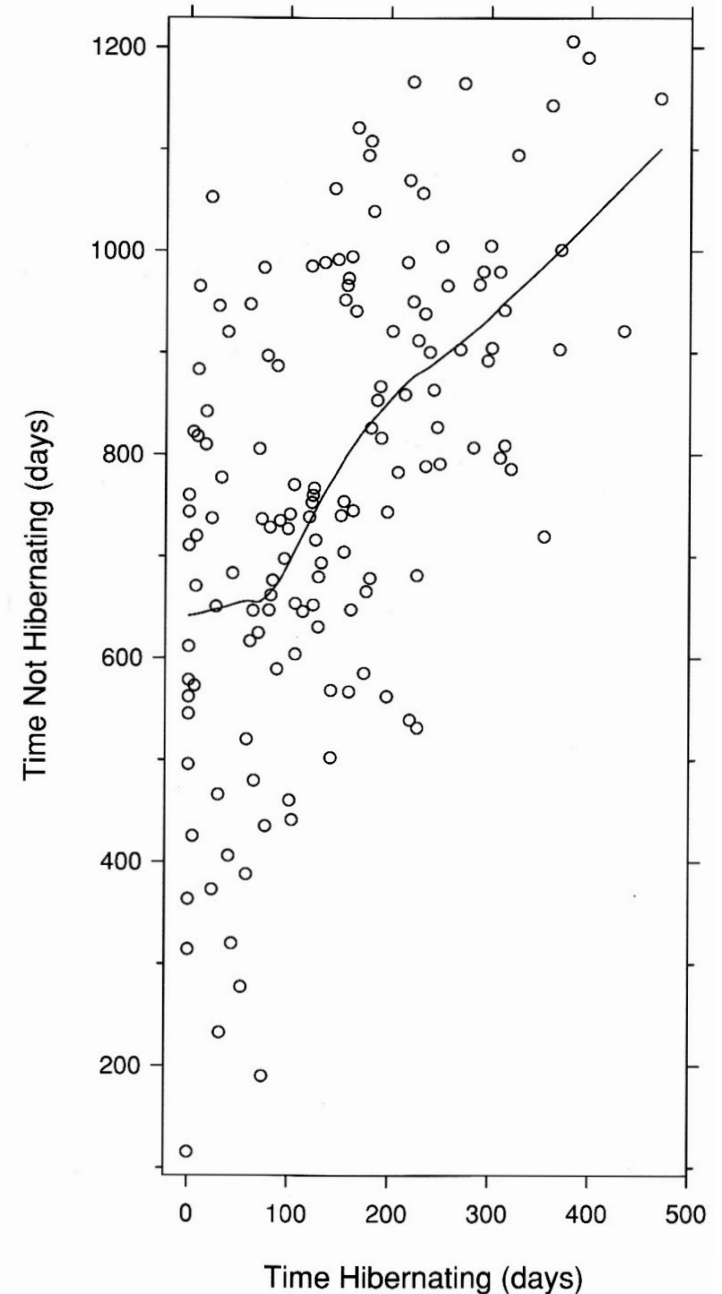
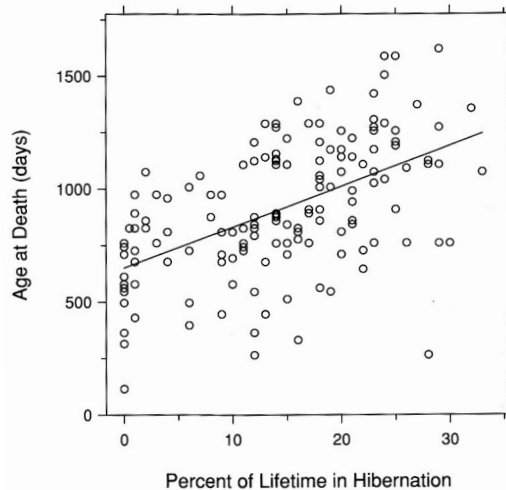
Correlation VisTrails Demo

Advanced Plotting

- Correlation
 - Locally Weighted Regression (Loess) using weighted least squares
 - Same as least squares but now with a weight:

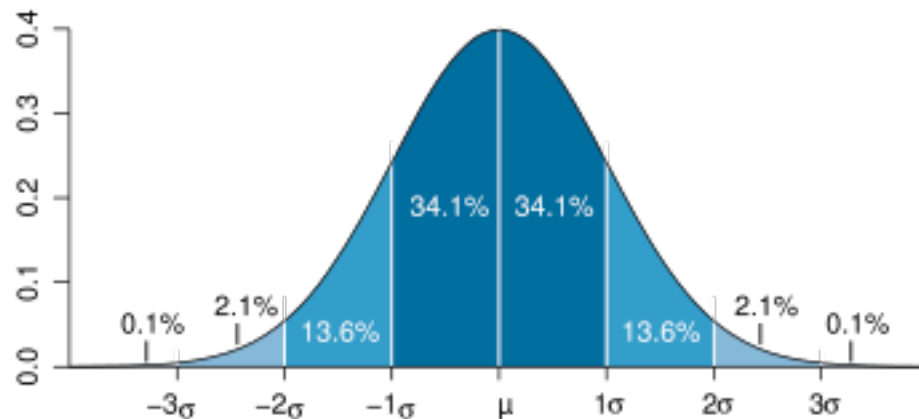
$$w_i = \frac{1}{n} / \sigma_i^2$$

$$\Delta = \sum_0 w_i (y_i - f(x_i))^2$$



Advanced Plotting

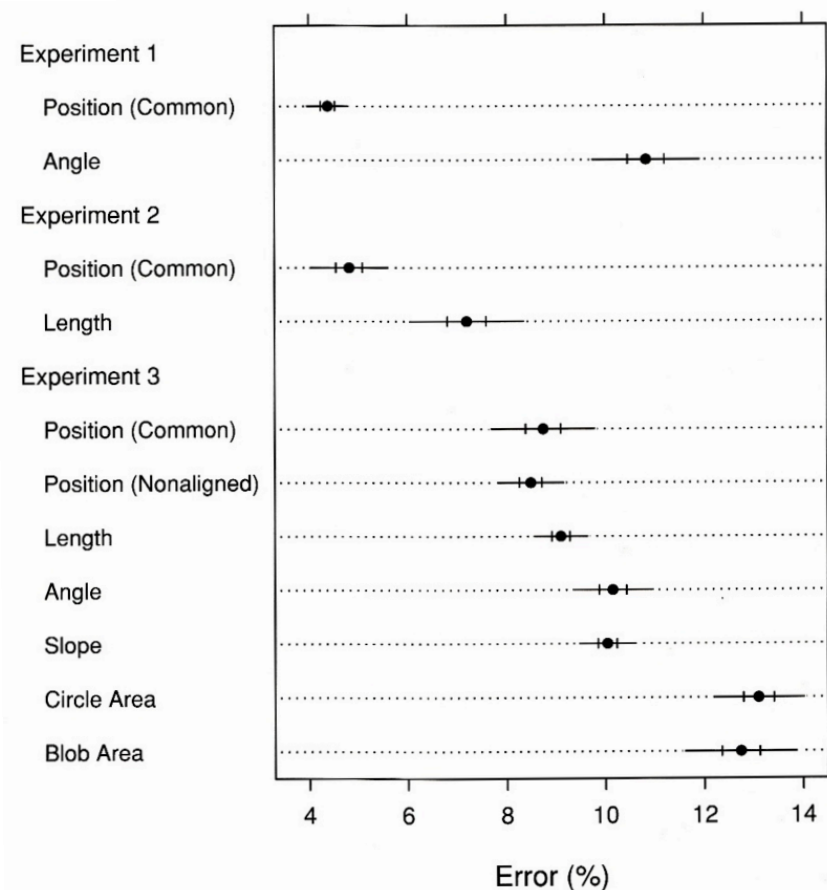
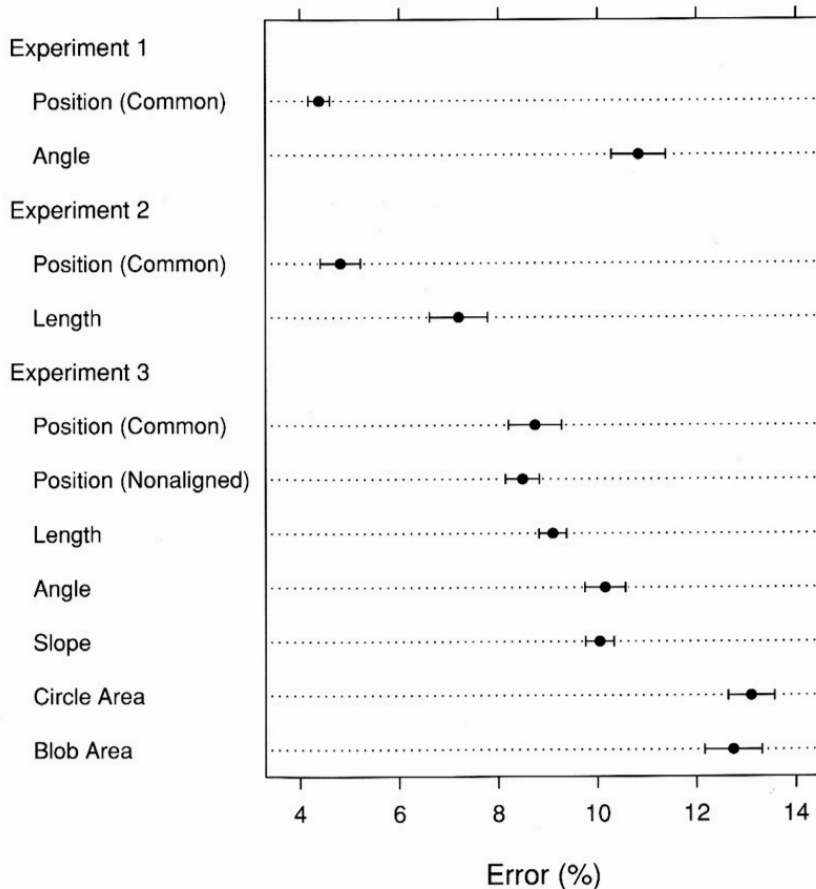
- Uncertainty and Variation
 - Two types of variation
 - Empirical Distribution: The variation captured in the experiment
 - Sample to Sample Variation: The variation that occurs in statistics from a sparse sampling and a denser sampling
 - Represent mean, standard deviation, and confidence intervals for a normal distribution



Advanced Plotting

- Uncertainty and Variation

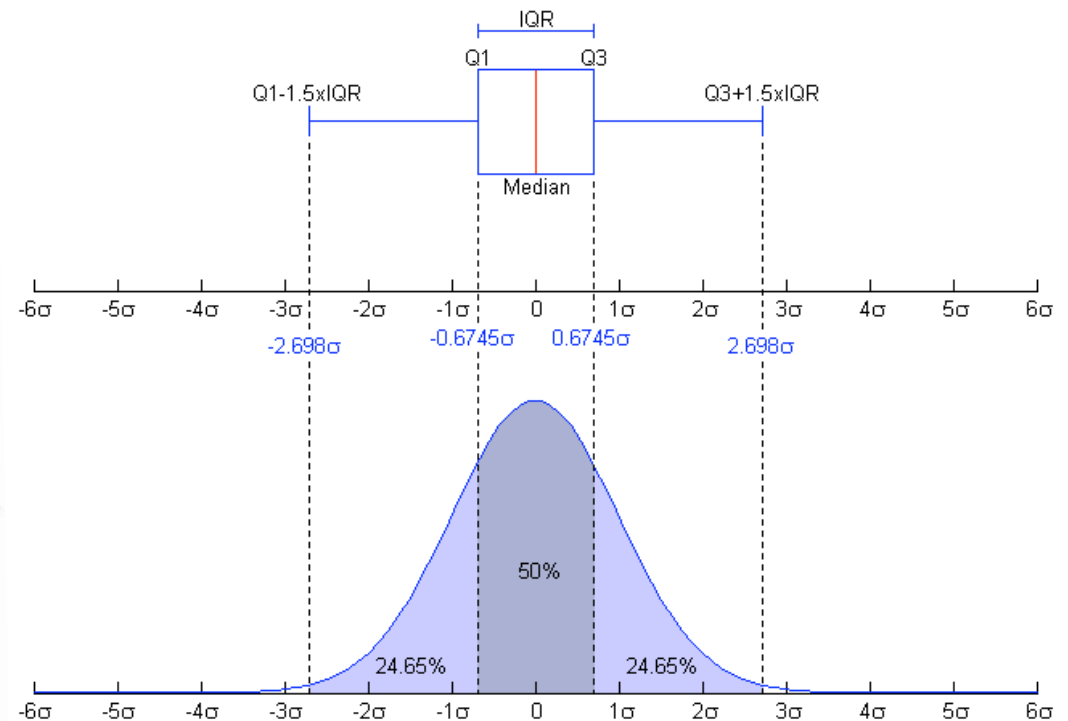
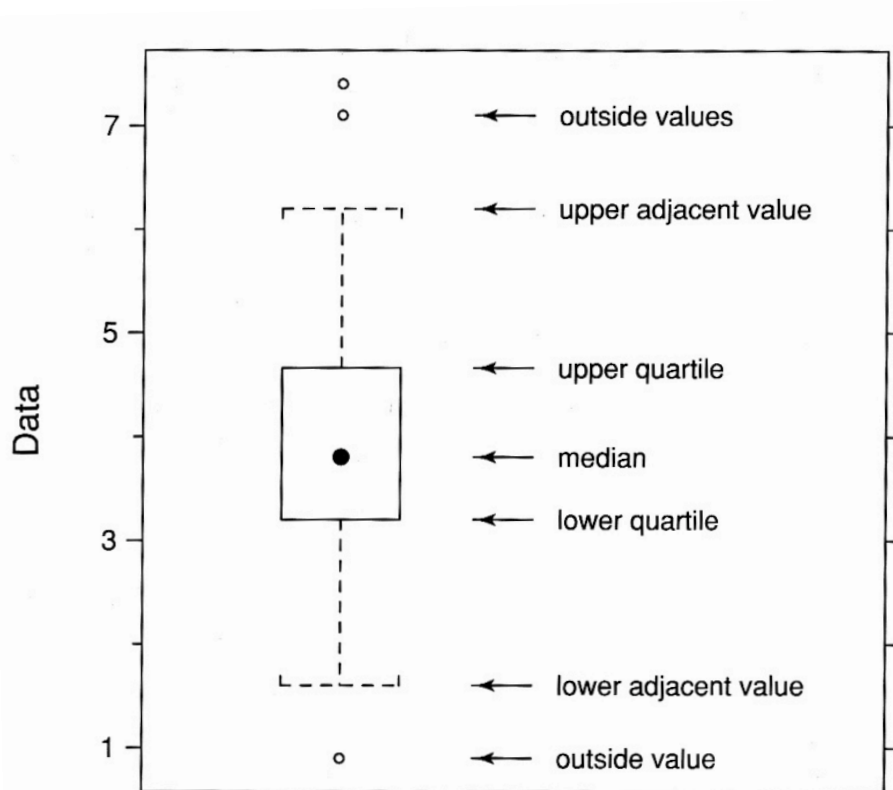
- Error Bars: Mean and one standard deviation or mean, 50%, and 95% confidence intervals



Advanced Plotting

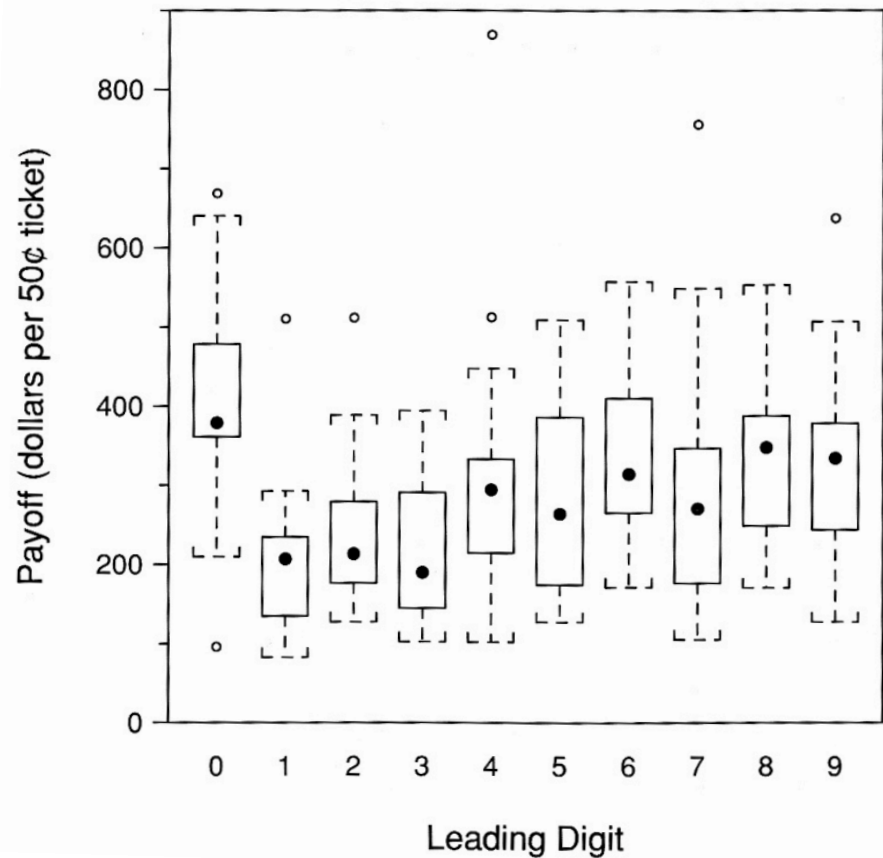
- Uncertainty and Variation

- Box Plots (Tukey Bars): First quartile, second quartile (mean), third quartile, adjacents (first-1.5r, third+1.5r), and outside.



Advanced Plotting

- Uncertainty and Variation
 - Box Plots (Tukey Bars)



Advanced Plotting

- Uncertainty and Variation
 - Box Plots (Tukey Bars)

Grades VisTrails Demo