Analytical Graphing

*lets start with the best graph ever made*

"Probably the best statistical graphic ever drawn, this map by Charles Joseph Minard portrays the losses suffered by Napoleon's army in the Russian campaign of 1812. Beginning at the Polish-Russian border, the thick band shows the size of the army at each position. The path of Napoleon's retreat from Moscow in the bitterly cold winter is depicted by the dark lower band, which is tied to temperature and time scales."

"The graph illustrates an amazing point — how an army of 400,000 can dwindle to 10,000 without losing a single major battle."
Two uses of graphs in Journals for BIOE 108

- Pattern depiction
- Predictions of results from specific hypotheses
Elements of a 2 D graph

- **y – axis**: Dependent axis, Ordinate axis
- **x – axis**: Independent axis, Axis of abscissa

In general, the logic of the axes is that you are assuming or predicting that the value of y depends on the value of x.

Basics of analytical graph theory

- Graph types imply a basis of logic and are not always interchangeable.
- Even interchangeable graph types are not always equivalent (some are just non-informative).
- Be very clear about what you are trying to convey: models, stats or data structure.
- Graph construction (axes, scales etc) may obscure or make clear the points you are trying to make.
- Graph trickery is usually just that – and typically subtracts from the depiction.
Graph types imply a basis of logic and are not always interchangeable

• Summary Charts
• Density Charts
• Data plots: e.g. Scatterplots

Summary Charts

There are a series of general graphical displays useful for characterizing the relationship between independent variables (usually categorical) and summary statistics of dependent variables (usually continuous).
Examples of continuous and categorical variables

<table>
<thead>
<tr>
<th>Categorical</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (male, female)</td>
<td>Hormone level</td>
</tr>
<tr>
<td>Nationality (French, Italian)</td>
<td>Location (Latitude, Longitude)</td>
</tr>
<tr>
<td>Species (Human, Chimp)</td>
<td>Phylogenetic distance</td>
</tr>
<tr>
<td>Color (red, green, blue)</td>
<td>Color (wavelength)</td>
</tr>
<tr>
<td>Age Group (Young, Old)</td>
<td>Age (years, days)</td>
</tr>
<tr>
<td>Height Group (Short, Medium, Tall)</td>
<td>Height (cm, inches)</td>
</tr>
<tr>
<td>Weight Group (Thin, Obese)</td>
<td>Weight (grams, pounds)</td>
</tr>
<tr>
<td>Speed (Fast, Slow)</td>
<td>Speed (cm/sec)</td>
</tr>
<tr>
<td>Temperature (Cold, Warm)</td>
<td>Temperature (degrees C)</td>
</tr>
</tbody>
</table>

Summary Charts

There are a series of general graphical displays useful for characterizing the relationship between independent variables (usually categorical) and summary statistics of dependent variables (usually continuous).

An example would be a bar graph of the relationship between education level (categorical) and mean income (continuous). Some types of summary charts:
Summary Charts

1. Bar charts display bars (categorical x data)
2. Dot charts display dots (or other plot symbols) where the top of the bar would be (categorical x data)
3. Line charts displays a line where the dots or tops of bars would be and are very useful for conveying information where there are likely to be unexamined x values (continuous x data eg. Time).
4. Profile charts fills in the area under the line (continuous x data)
5. Pyramid charts draws pyramids instead of bars. (categorical x data). Try to stay away from these
6. Pie charts displays the proportion of counts or measures falling within the category (categorical x data) (Again these are graphs to stay away from)
Summary Charts

In the above chart types, the height or size of the display element usually represents one of the following summary statistics:

1. The count in that category.
2. The mean or median of the values in that category.
3. The percentage of cases in that category.
4. Other less common summary statistics such as sum, min, max standard deviation, variance, coefficient of variation.

Adding Strata

More complex graphs can compare the responses of dependent variables to multiple independent variables (For example the relationship between income, sex and education). The strata should be informative!!!
Which conveys the information most clearly – how about the comparisons of interest

Summary Charts
Line Graphs

- Line graphs can be used to depict responses to categorical variables (e.g., Gender). Usually these data are better represented in bar or dot graphs.
- Line graphs can also be used to convey a sense of the likely response to points that exist between measurements. For example, samples taken yearly.
Density Charts

- The density of a sample is the relative concentration of data points in intervals across the range of the distribution. A histogram is one way to display the density of a quantitative variable.
- For example, the size frequency distribution of the giant owl limpet (Lottia gigantea) along the west coast.
Raw Data Plots: e.g. Scatterplots,

- Scatterplots are probably the most common form of graphical display. **The key feature of scatterplots is that raw data are plotted** (in contrast to summary data as in summary charts). Regression lines with confidence bands or smoothers (e.g. linear, non-linear) can be added to help explain relationships among variables. An example is the relationship between mussel height, and length and mussel height and mussel mass.

- How to estimate length and mass of mussels?
Non-linear and linear smoothing

Basics of analytical graph theory

- Graph types imply a basis of logic and are not always interchangeable
- Even interchangeable graph types are not always equivalent (some are just non-informative)
- Be very clear about what you are trying to convey: models, stats or data structure
- Graph construction (axes, scales etc) may obscure or make clear the points you are trying to make
- Graph trickery is usually just that – and typically subtracts from the depiction
Which is best?

Graphing for patterns and hypotheses: example limpets and *Ulva*
General Hypotheses

• Grazing by limpets reduces *Ulva* Cover
General Hypotheses

- Grazing by limpets reduces *Ulva* Cover
- Freshwater runoff increases *Ulva* Cover and decreases Limpet numbers

General and Specific Hypotheses

- Grazing by limpets reduces *Ulva* Cover
General Hypotheses

- Grazing by limpets reduces *Ulva* Cover

*Specific hypothesis*

- Replicate plots cleared of limpets will show an increase in *Ulva* cover compared those where limpets are not removed
General Hypotheses

• Freshwater runoff increases *Ulva* Cover and decreases Limpet numbers

Specific hypothesis

– Replicate plots with limpets that are exposed experimentally to freshwater will show an increase in *Ulva* cover compared to replicate control plots where no freshwater is added
General Hypotheses

• Freshwater runoff is increases *Ulva* Cover and decreases Limpet numbers

  *Specific hypothesis*
  
  – Replicate plots with limpets that are exposed experimentally to freshwater will show a decrease in Limpet density compared to replicate control plots where no freshwater is added