Model Based Statistics in Biology

Chapter 2.4 Graphical and tabular display of data

ReCap (Ch 1)
Quantities (Ch2)
2.1 Five part definition
2.2 Types of measurement scale
2.3 Data collection, recording, and errorchecking
2.4 Graphical and tabular display of data
Critique of graphs and tables
2.5 Ratio Scale Units
Base
Standard Multiples
Commonly used units in biology

```
Not here last time?
     Course Outline
     Name on roster
Questionnaire results
```

```
Discussion of Cards Lab:
  Anybody come up with "wrong"
rule that works?
      In 1997 mutually exclusive
pairs introduced ("test" cards),
before going to multiple working
hypotheses ("crucial" cards).
   Ask for discussion of this,
comparison of "test" and
"crucial."
     In 1998 crucial cards only.
```

Recap Chapter 1
Quantitative reasoning: Example of scallops, which combined stats and models
Biological reasoning will take the lead.
Stats will be a tool, rather than defining the way we think about biological problems.
Models (equations) are ideas about the relation of quantities

Recap Chapter 2
Quantities: 5 components
- Name
- Symbol
- Procedural Statement (write out and review: Could I use this recipe?)
- Set of numbers collected into a vector
- Units on a defined measurement scale

Measurements made on four types of scale: nominal, ordinal, interval, ratio
Data collection, recording, and error checking

Today: Now that we covered the collection, recording, and error checking of measured quantities, we move to tabular and graphical display of these quantities.

Wrap-up
- Graphs and tables are an important part of quantitative work.
- Graphs and tables display fully defined quantities.
- Like prose, they are critiqued before presentation. Clear? Understandable?
- The purpose of tables differs from that of graphs.
- The design of tables and graphs depends on purpose.

Data Display: Tables & Graphs

Tables and Graphs differ in purpose.

Tables. Primary purpose is archival.
Another purpose is to provide summaries, as a matter of record.

Tables can also be used for communication of pattern, although this is usually less effective than graphs because digits do not convey magnitude as readily as lengths and areas on a chart.

An exception to this rule is a simple 2-column table
Lateral comparisons are easily made between two columns
This simple arrangement often requires less space

1	1
10	33
24	56
9	4
2	5

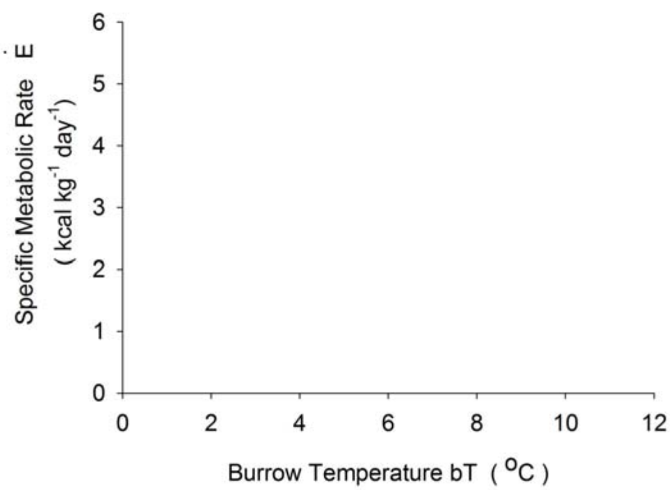
Graphs. Primary purpose is to display and communicate pattern.
(people look at graphs, rather than reading report).
Another purpose is to discover pattern.
(these graphs are not usually displayed).

Graphs more effective in communication, as rule, than tables.
People look at graphs in report, rather than reading report.

Graphs and tables display fully defined quantities Tables and graphs of a scaled quantity must contain three of the five components of a quantity: name, units, and numbers. A fourth component, the symbol, adds to the presentation by linking the graph or table to the text.

For graphs a convenient format that includes these components is to list the name of the quantity along an axis, then list the symbol and units. The following graph axes shows the format.

This format links the quantity in the graph to a procedural statement in the text.



Graph and tables display fully defined quantities.

The person collecting the data will not be present to explain material to reader, so enough information must be provided for the reader to understand the graph. Linking graphs and tables to a fully defined quantity (all 5 parts in text) achieves this.

Graphs and tables need to be self-explanatory

This means
- caption with adequate detail
- list units as column headings (tables) or on axes (graphs)
- use numbers, not exponents.

One of the common failings of graphical (and tabular) presentation of data is that the base of logarithms is not reported. If only the exponent is reported, without the base, then we have no way of knowing whether the number represented by the exponent 2 stands for 2^2 or 10^2. One solution to this common defect is to use the symbol to show exactly how the rescaling was done, as in the example of $\log_2(A/\text{cm}^2)$. The quantity A has been divided by its units, which rescales it to a unitless number. The logarithm of this number can then be taken (the operation of taking the logarithm of a unit such as cm^2 is not defined).

Design of Tables & Graphs   Design depends on purpose.

For archiving and summary, the primary criterion is complete documentation.

For discovering patterns the primary criterion is
- separating model from residual
- differentiate data from model
- use different colors, different line thickness to facilitate comparison across several variables

For communication the criteria are slightly different:
- self-explanatory
- clarity

Design of Tables

Tables are 2-dimensional structures. Tables should be constructed with this in mind. If 3-dimensional data is presented (e.g. temperature, behavior, food) then arrange a series a two-dimensional tables so as to have fewest possible tables.

Layout
- Comparisons work horizontally, so arrange data in columns
- Group columns and rows to facilitate comparison.
- Use several small tables instead of one big table.

Clarity
- Shorten numbers to readability.
- Use a clear title with memorable symbols.
- Add summary figures, deviations.

Critical review
- could someone else use this data ?
- enough information for some else to use ?

Design of Graphs (from Tufte 1983, 2001).

Graphs tell a story. Story is honest.
Think about model and residual in graphing X and Y
Avoid freak displays, unnecessary decoration, and other distractions.

Layout

Not too many curves in one panel.
Use stacked panels & small multiples (many small graphs, same x-axis).

Clarity

Distinguish true zeroes from missing values.
Arrange so as to bring out relation, especially multivariate relation.
Show deviation as well as story (model).

Bad practice:

inflated pictographs
zero should appear on vertical axis
this can be a problem for ratio data, though not interval data
Tufte: Chartjunk
Lie factor

Tufte, E

1983. The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press. (2001, 2nd edition, ISBN 0961392142).

1990. Envisioning Information. Cheshire, CT: Graphics Press. ISBN 0961392118.

1997. Visual Explanations: Images and Quantities, Evidence and Narrative. Cheshire, CT: Graphics Press. ISBN 0961392126.

2003. "PowerPoint is evil". In: Wired 11 (9). ISSN 1059-1028. .

2003. The Cognitive Style of PowerPoint. Cheshire, CT: Graphics Press. ISBN 0961392169.

2006. Beautiful Evidence. Cheshire, CT: Graphics Press. ISBN 0961392177.

Critique of graphs and tables

Table 5.2. Checklist for evaluating graphs and tables.

Graphs

Are the axis labels and titles (or captions) adequate?
Are units stated for x and y axes?
Have appropriate symbols or lines been used?
Does the graph have freak characteristics, such as uninformative decoration?
Does the graph convey a story (bring out relations between variables)?
Does the graph mislead the reader in any way?
Can the data in the graph be pulled off accurately as numbers?
Are there variables that could be added to help interpret the trends shown?

Tables

Are row labels, column labels, and captions adequate?
Can rows and columns be regrouped or rearranged to facilitate comparison?
Would the addition of statistics (sums, deviations, etc) help?

In looking at each graph or table ask yourself:

-- Is the relation of one variable to another quickly and easily grasped?

-- Can the information in the graph be translated into numbers?
(could you pull the numbers off the graph and redraw it yourself?)

When critiquing a graph, reflect on which problems are serious, and which are less so. State reasons or criteria for judging that some problems are serious, some are not.