

Graduate School of Biomedical Engineering

BIOM1010 – Week 3 Tutorial

Graphing and Data Presentation



Why do we graph data?

1. So we can figure out what it means. We can see a pattern when we graph the data (we hope!) and we make inferences about the cause of the pattern.
 - At this stage the graph can be quick and dirty. In fact, there are usually many graphs.
2. So we can convince other people that the data mean what we say it means. We hope to convince others that our inferences are correct.
 - Now the graph must be of high quality and drawn in a way that helps make our point. But honestly!

General considerations

The important thing is the data.

- Interesting graphs are at least bivariate if not multivariate. We are plotting one variable against another variable (or factor).
- We are showing that A and B are related. And maybe that differences in A cause differences in B.

Variables can be

- Quantitative (height, weight, concentration,...)
- Categorical (male vs. female, Drug vs. Control,...)

The vertical (y) axis usually represents a quantitative variable, an outcome or response, the dependent variable.

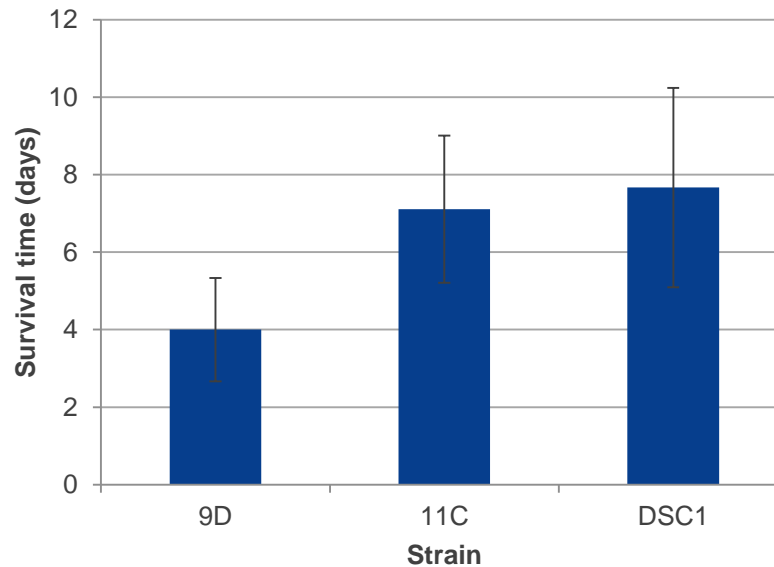
The horizontal (x) axis usually represents an independent variable or factor. It can be quantitative or categorical.

Types of graphs

Column (vertical bar) graph.

- Horizontal axis is categorical
- Vertical axis is quantitative
- Height of the bar indicates the mean (average) of a set of measurements.
- There may be “error bars” which give an idea of the spread of the data.
- Should state what the error bars represent, for example, the standard deviation (SD)

A column graph (Excel)



Categorical x-axis

Figure 1: Survival time of mice infected with 3 strains of typhoid organism. (mean \pm SD)

Types of graphs

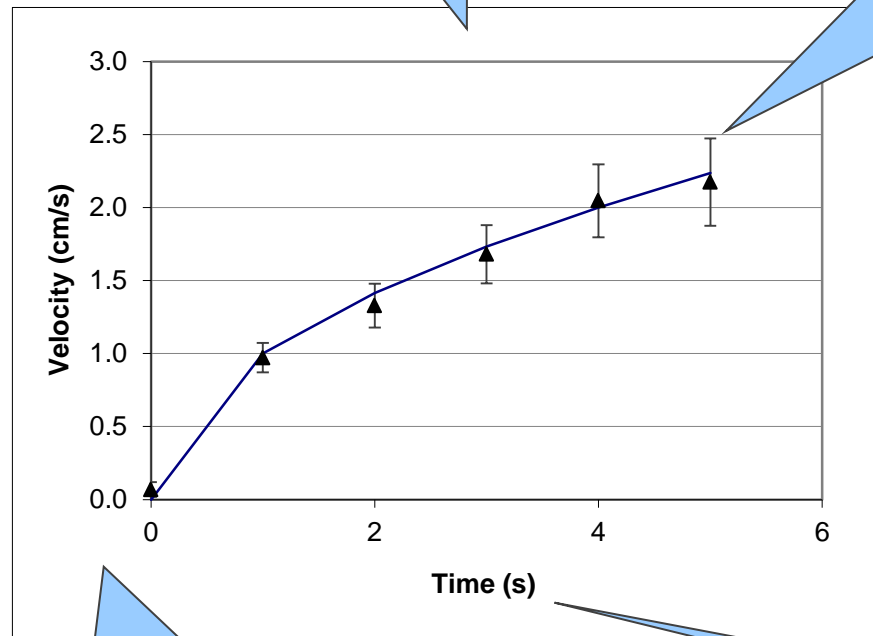
Scatter plot

- Horizontal (x) and vertical (y) axes are quantitative
- There may be multiple observations (y s) for each value of x .
- These might be shown as individual points or as a single point representing the average, often with error bars.
- There may be a line that represents a model of the data. Often this is a straight line (regression line).

A scatter plot (Excel)

Gridlines: none; or if necessary,
as with log axes, muted (grey).

Plot data as symbols
Plot a model as a line



Appropriate number of ticks and
appropriate number of digits in tick
labels

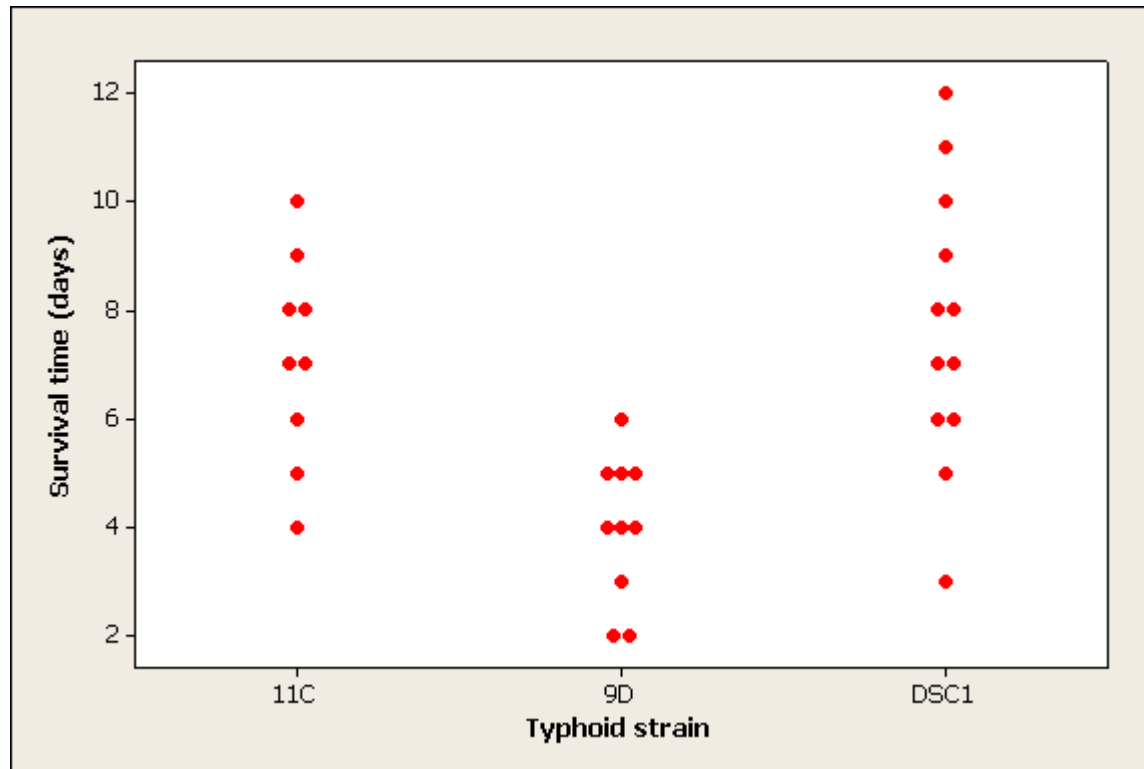
x-axis and y-axis labels
show the units.

Dotplot (show me the data!)

Sometimes it makes sense to show all of the actual data as points.

- When the number of observations is small
- When the distribution of the data is asymmetric (skewed) it is misleading to represent the data by the mean \pm error bar, which implies that the distribution is symmetric

Example (Minitab – a statistics program)



A Famous Example

From

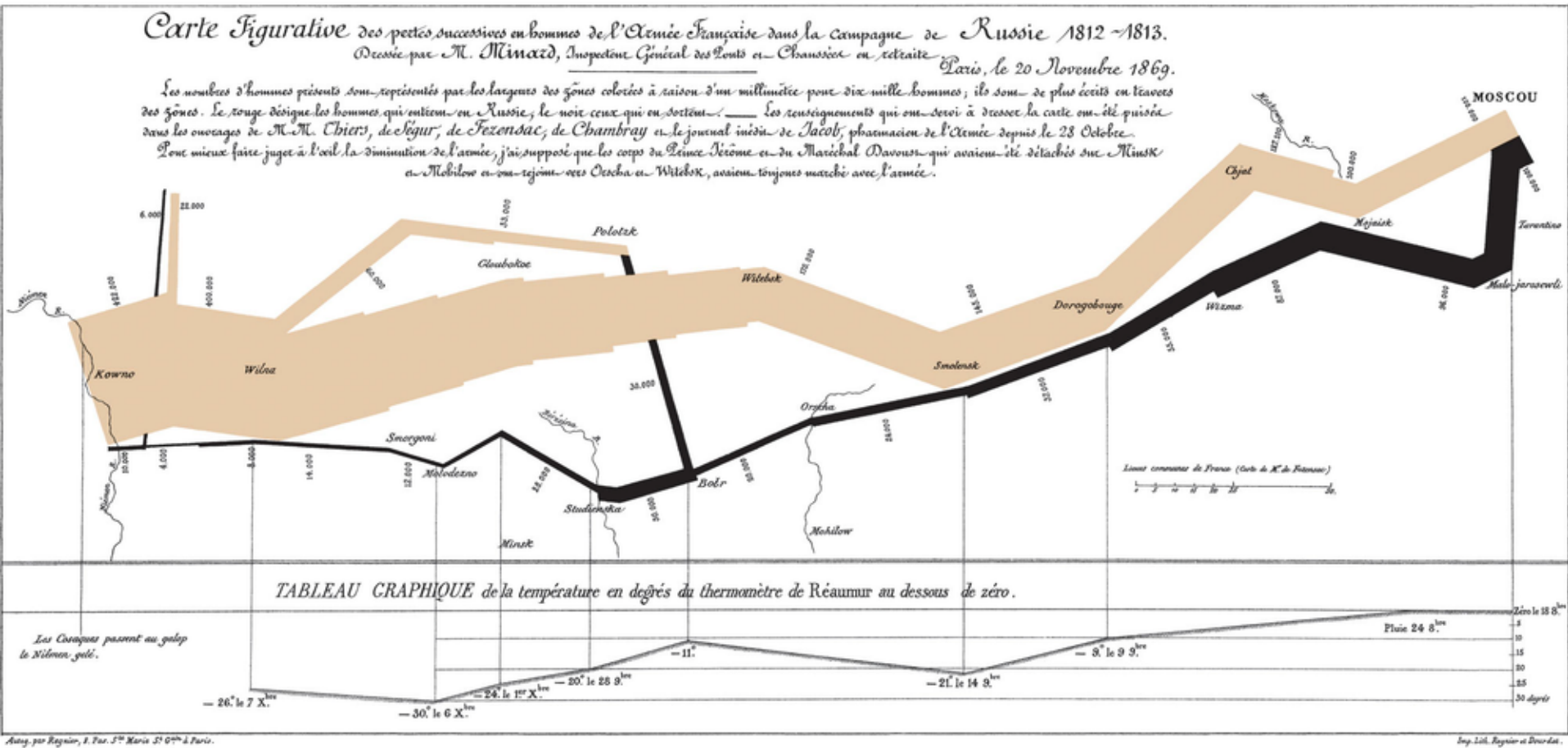
The visual display of quantitative information

Edward R Tufte

How many variables are represented in the graphic?

(You may want to zoom in to read the fine print.)

The French Army's Russian Tour 1812-13



Charles Joseph Minard's famous graph showing the decreasing size of the Grande Armée as it marches to Moscow (brown line, from left to right) and back (black line, from right to left) with the size of the army equal to the width of the line. Temperature is plotted on the lower graph for the return journey (Multiply Réaumur temperatures by 1¼ to get Celsius, e.g. -30 ° R = -37.5 ° C)