

Plot type specific considerations

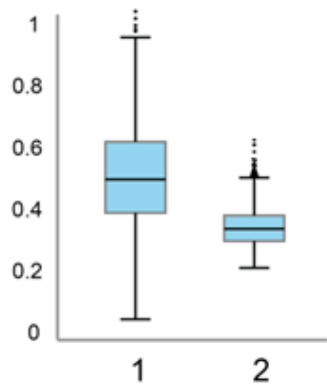
v2018-02

Anne Segonds-Pichon
Simon Andrews
Phil Ewels

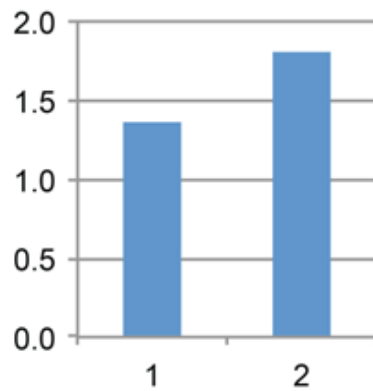
anne.segonds-pichon@babraham.ac.uk
simon.andrews@babraham.ac.uk
phil.ewels@scilifelab.se

Types of plot

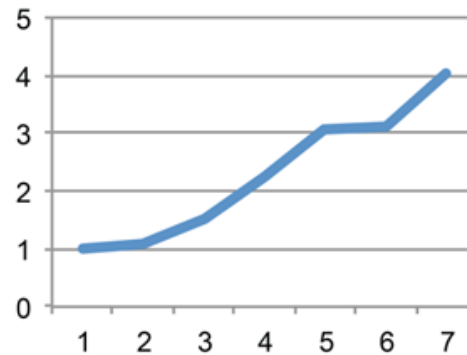
Things you can illustrate



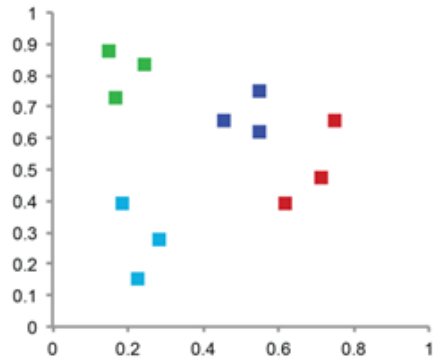
Distribution



Comparison



Relationship



Composition

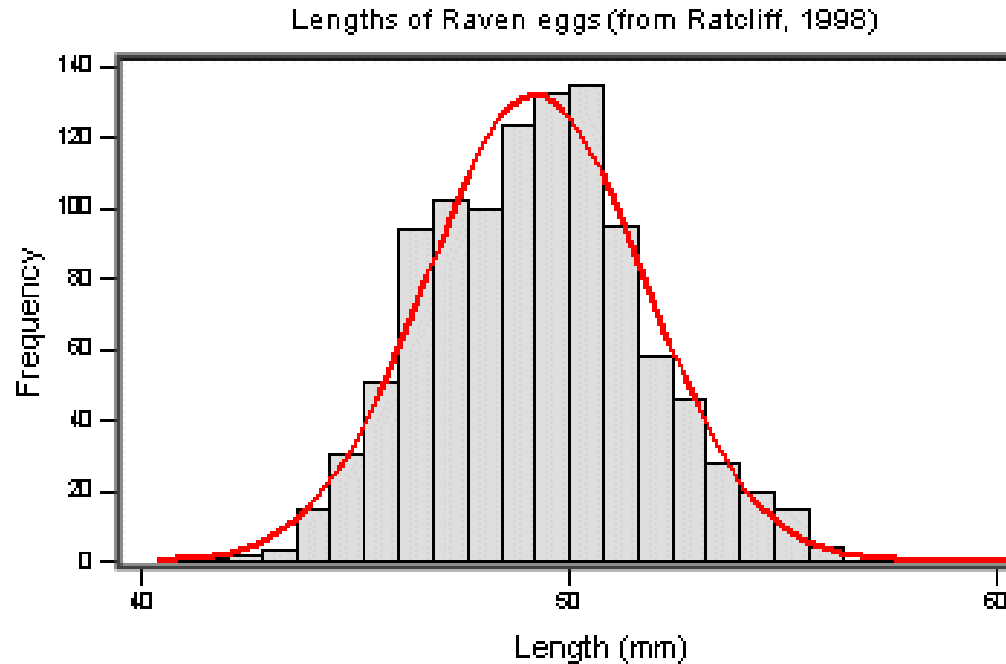
Plot Properties

- Exploration, Presentation or both?
- Effectiveness
- Scalability

- Options
- Potential Problems

Distributions

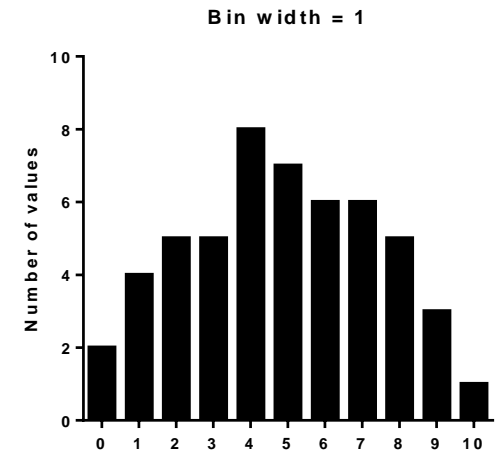
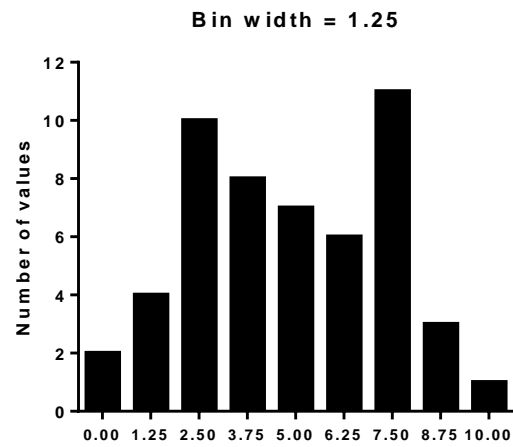
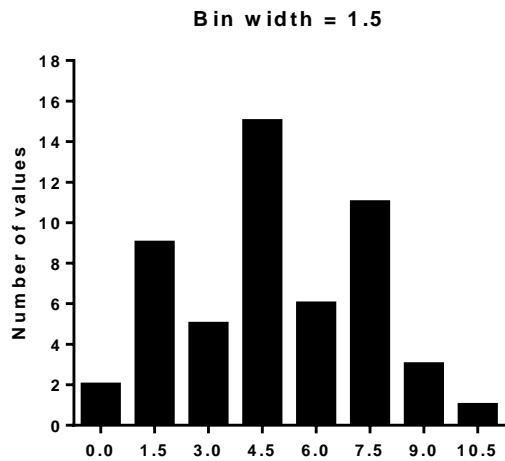
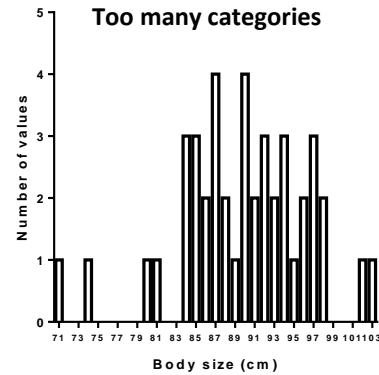
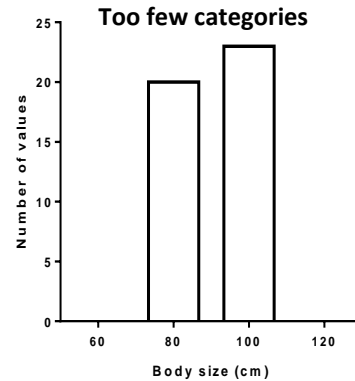
Histograms / Density Plots



- Exploration or Presentation Both
- Effectiveness Good
- Scalability Poor

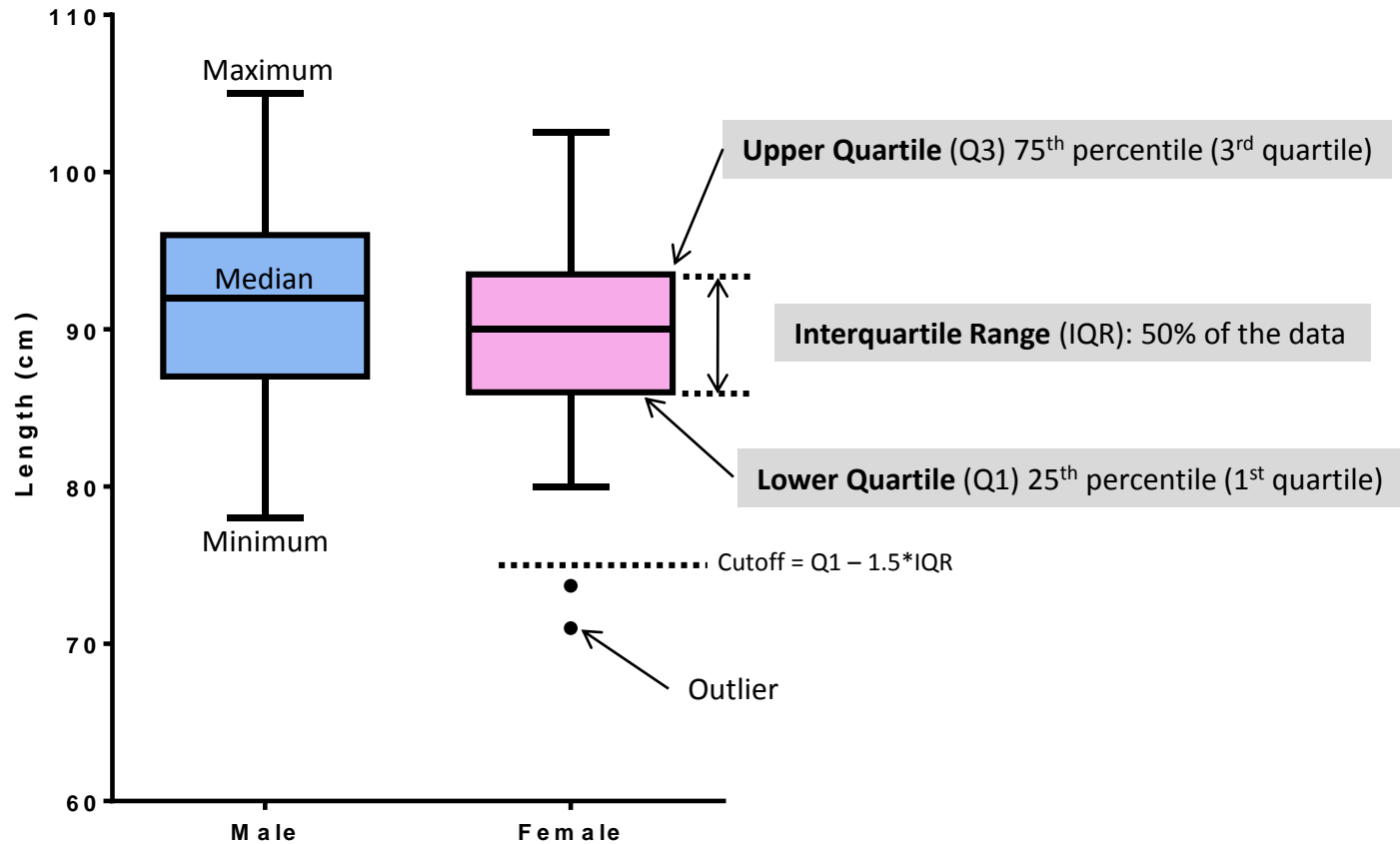
Histogram Options / Problems

- Bin Size



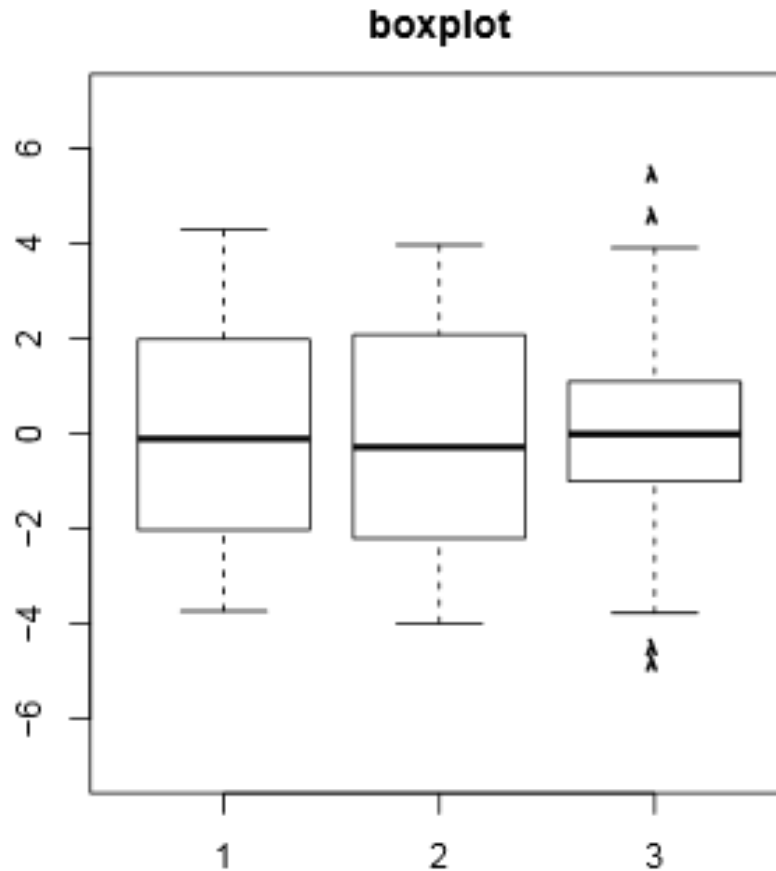
- Discrete Data

Box Plots



- Exploration or Presentation Presentation
- Effectiveness Good
- Scalability Good

BoxPlot Problems



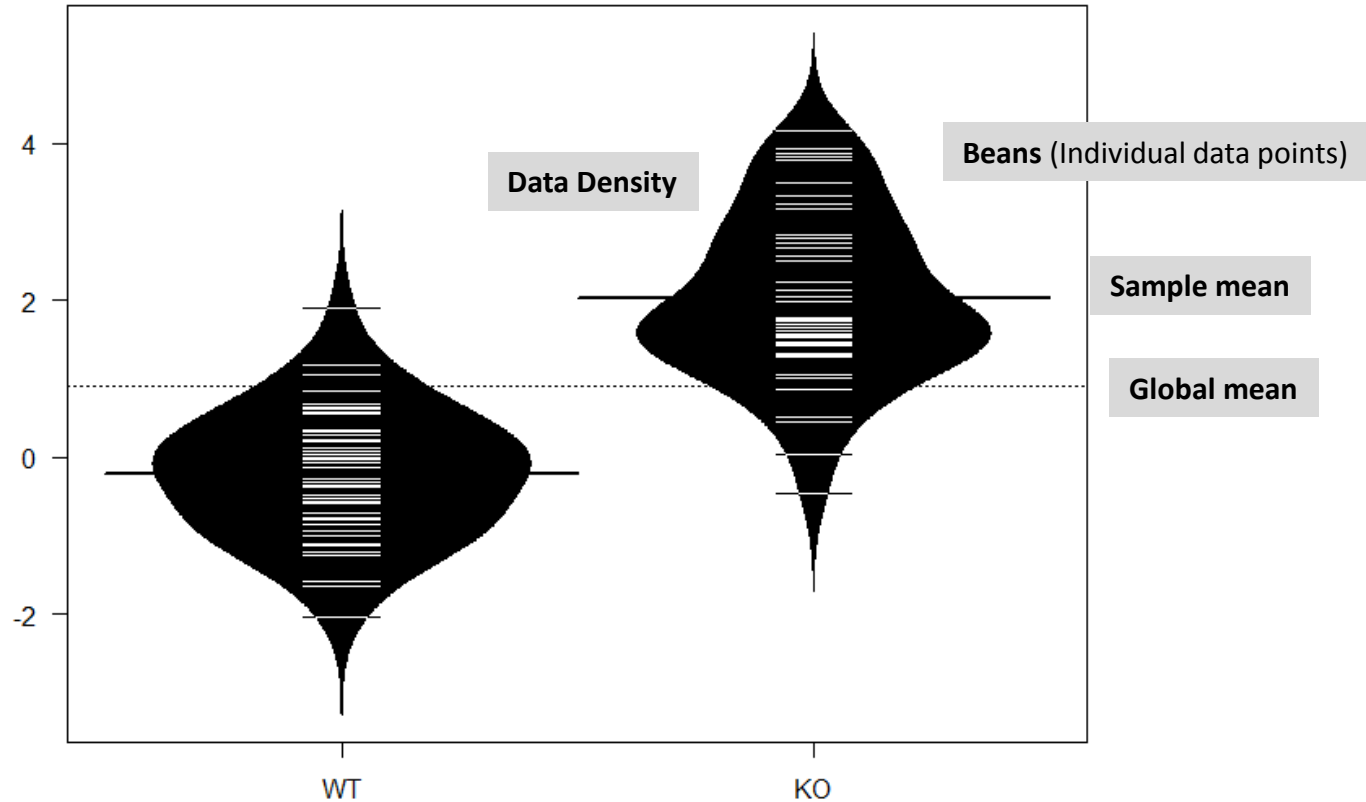
Assumes a large, normally distributed dataset

Misleading plots from small or non-normal datasets

In most cases there are better alternatives

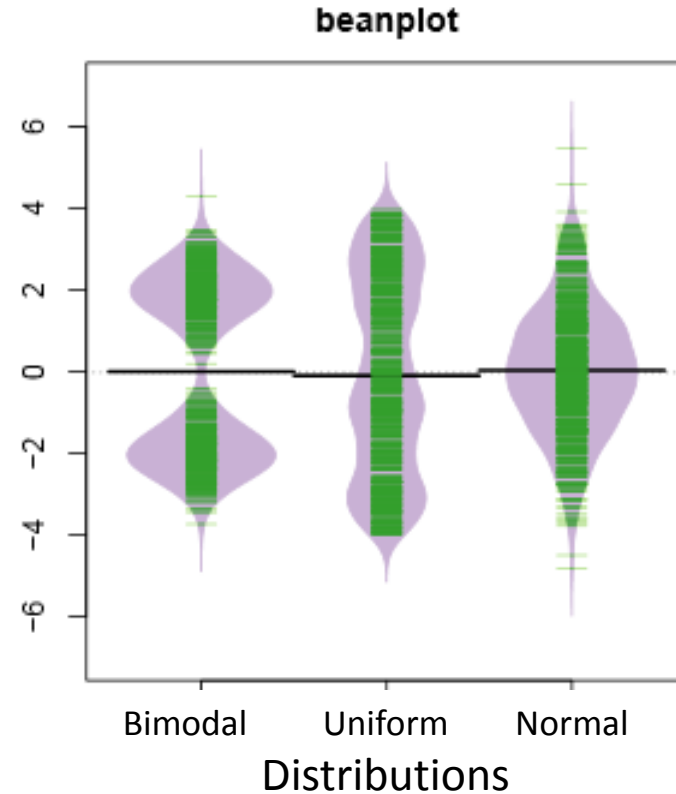
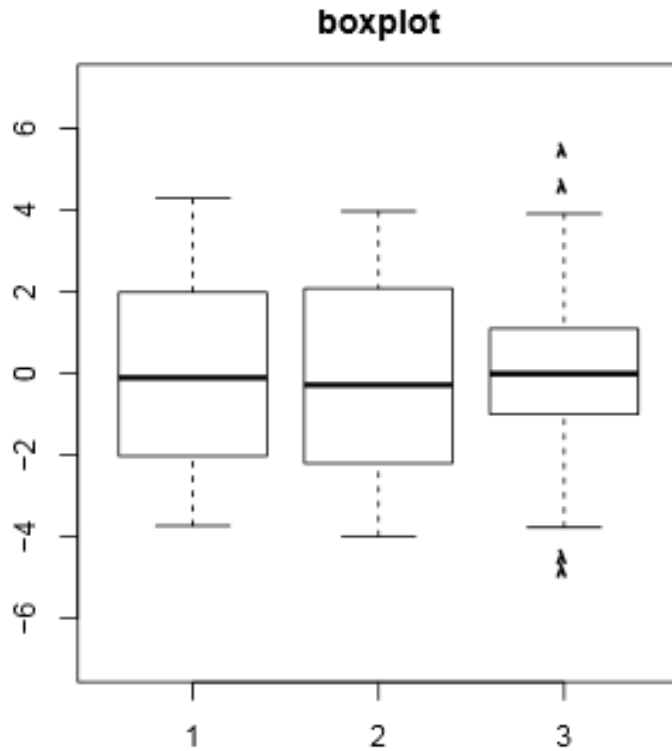
Bean Plots

A beanplot



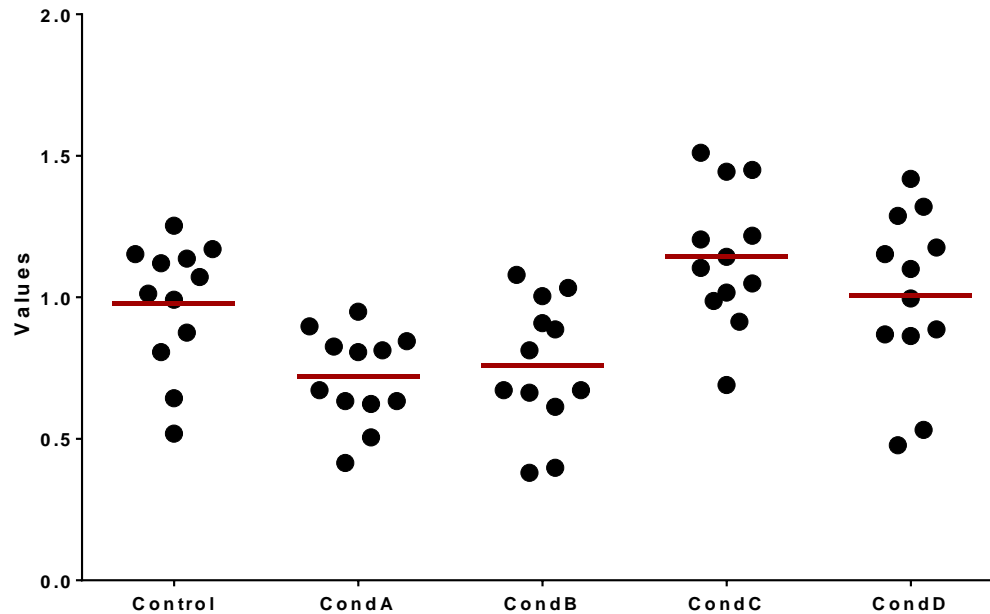
- Exploration or Presentation Both
- Effectiveness Good
- Scalability Good / Intermediate

BoxPlot vs Beanplot



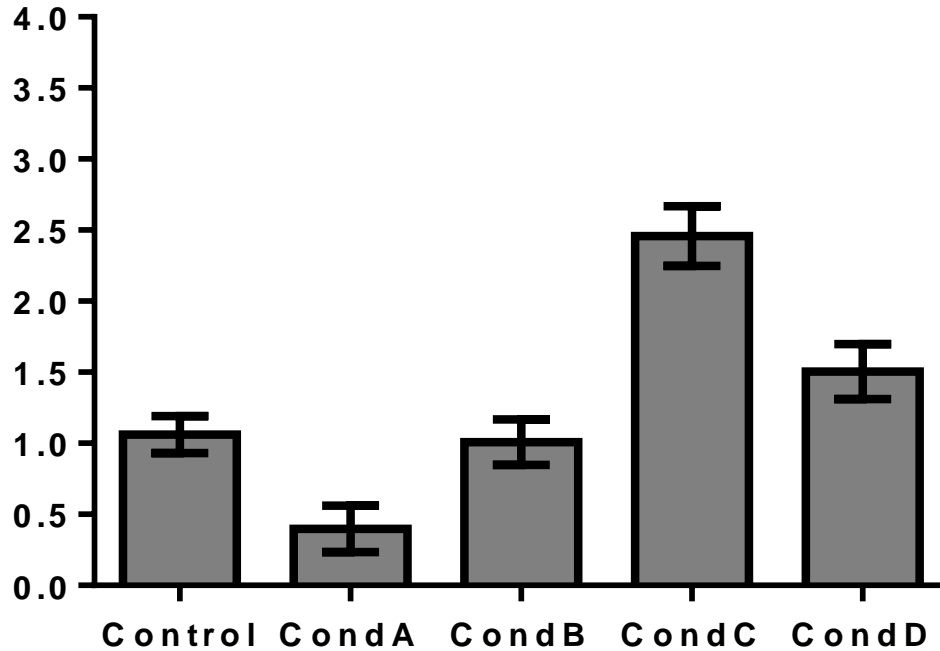
Comparisons

Stripcharts



- Exploration or Presentation Both
- Effectiveness Good
- Scalability Poor

Barplot

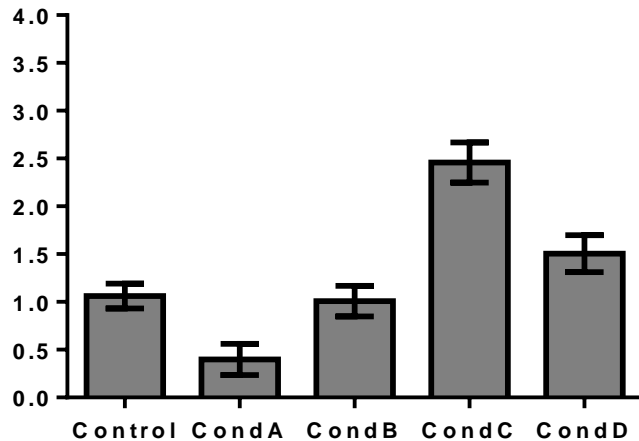


- Exploration or Presentation Presentation
- Effectiveness Good
- Scalability Good

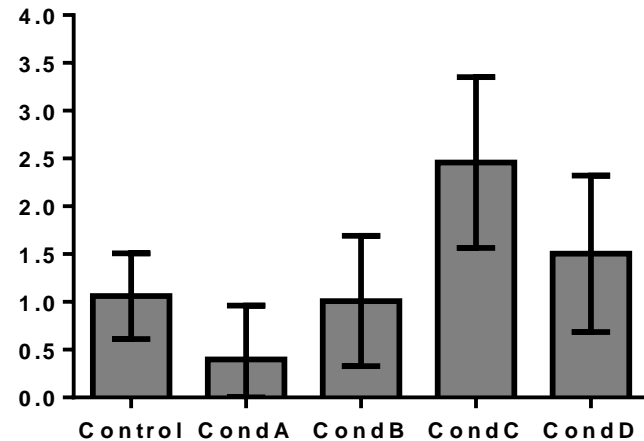
Barplot Options

Selection of suitable confidence measures

Standard error

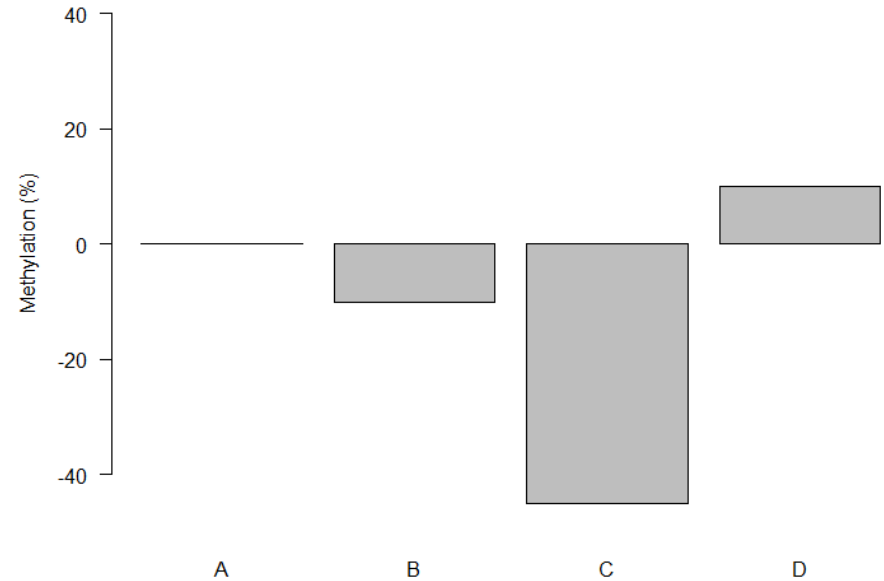
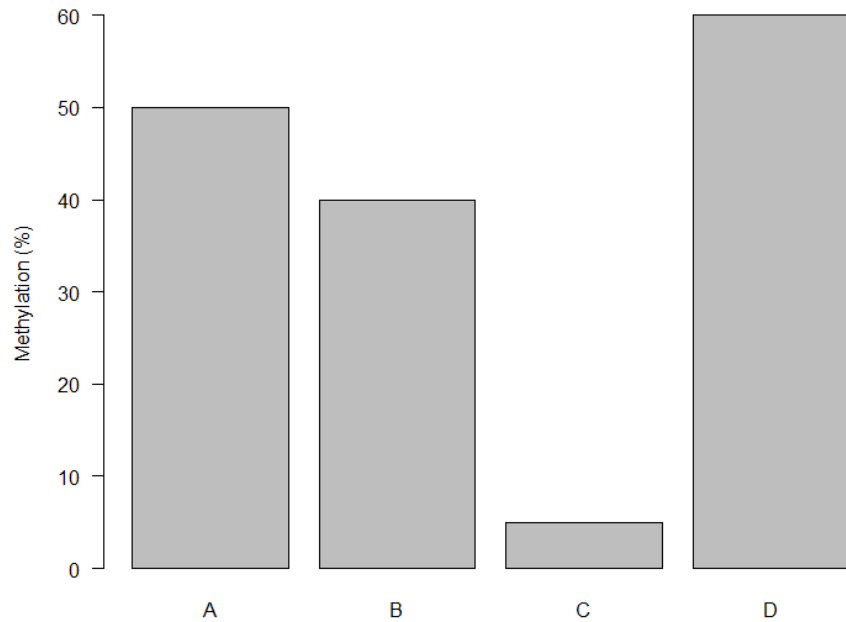


Standard deviation



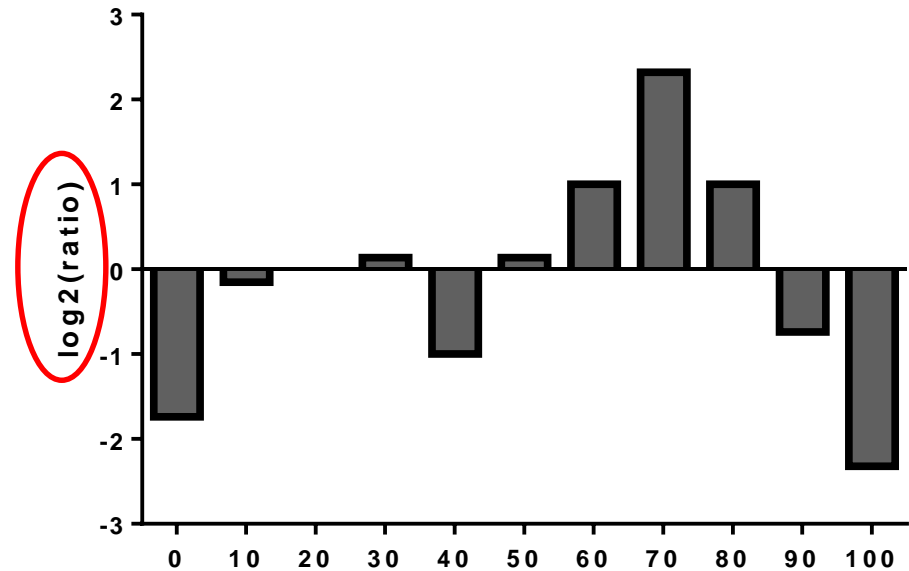
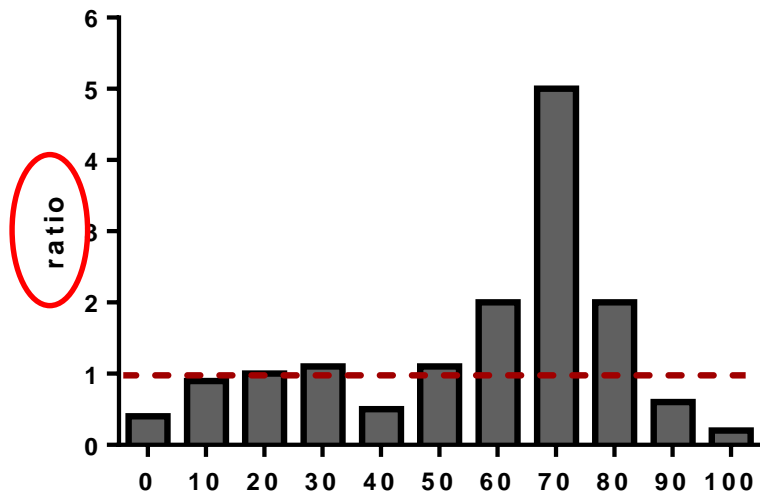
Boxplot Problems

Setting a suitable baseline

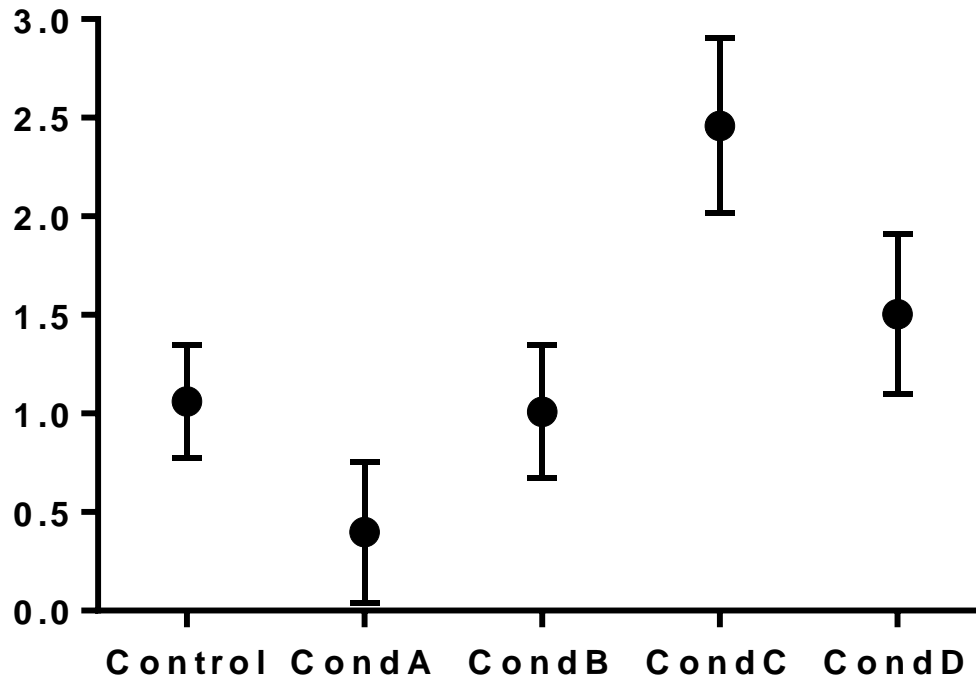


Boxplot Options / Problems

Dealing with ratio data



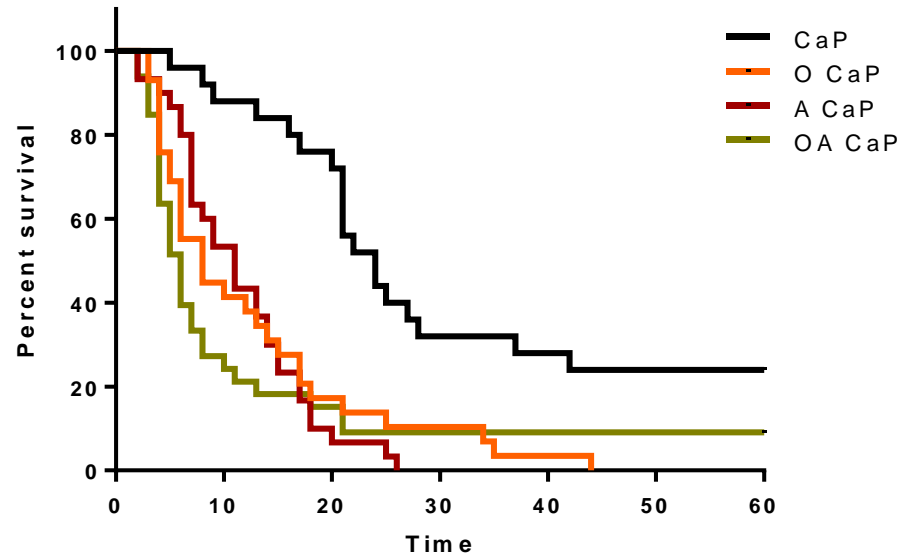
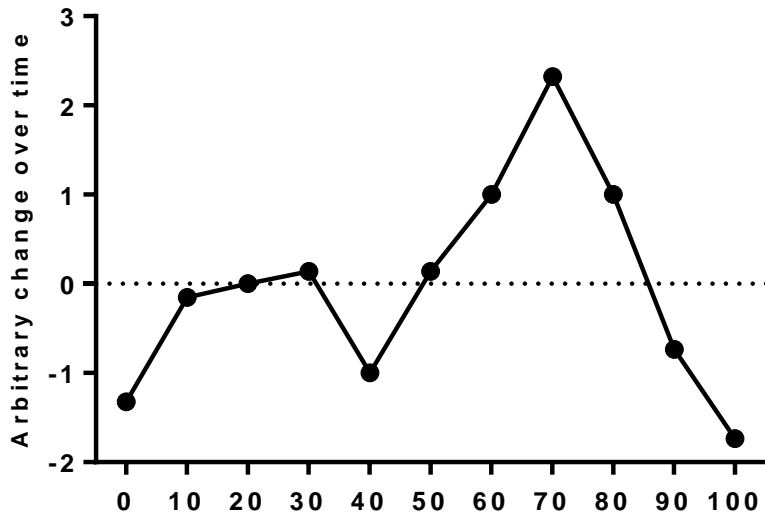
Confidence Interval Plots



- Exploration or Presentation Presentation
- Effectiveness Good
- Scalability Good

Relationships

Line Graphs

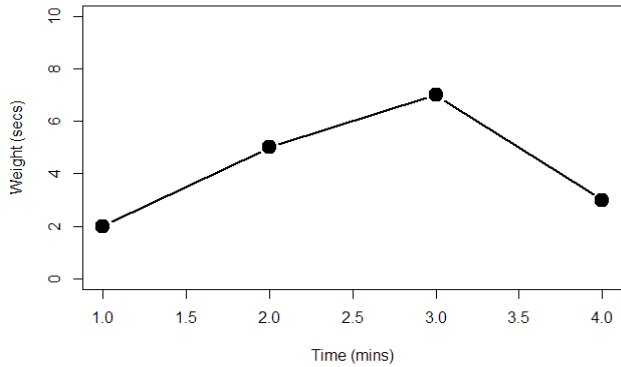


- Exploration or Presentation Both
- Effectiveness Good
- Scalability Poor

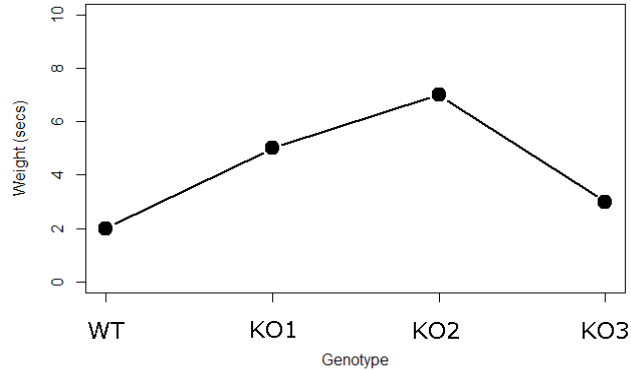
Line Graph Problems

Discrete Data

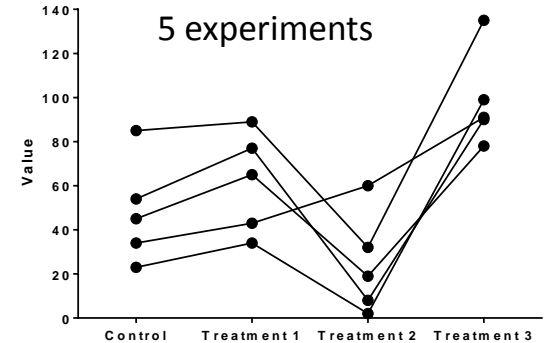
Good line graph



Bad line graph

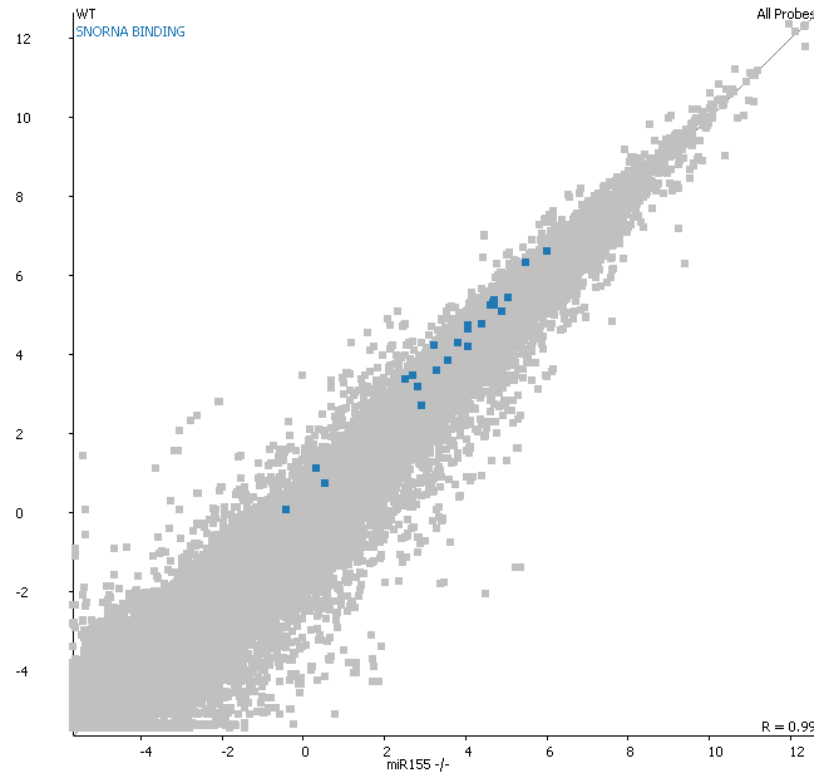


5 experiments



Implies interpolation
Can be useful for exploration
Shouldn't use for presentation

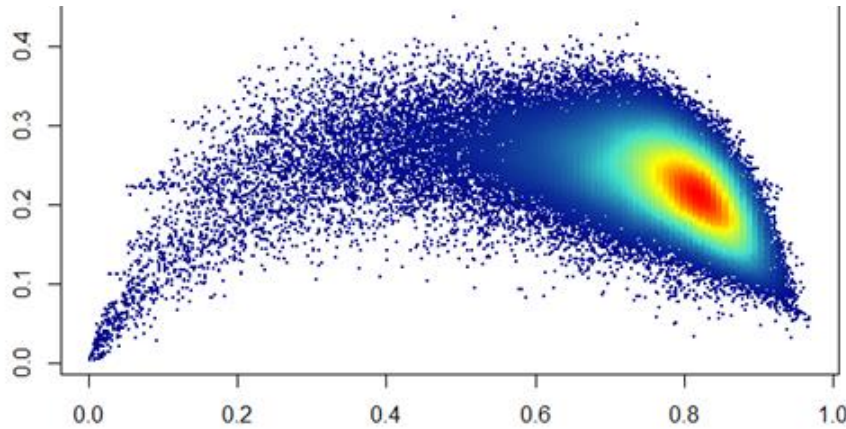
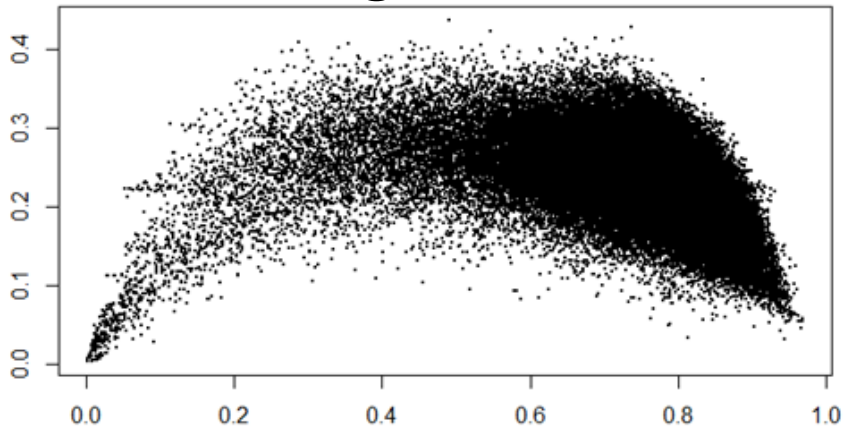
Scatterplots



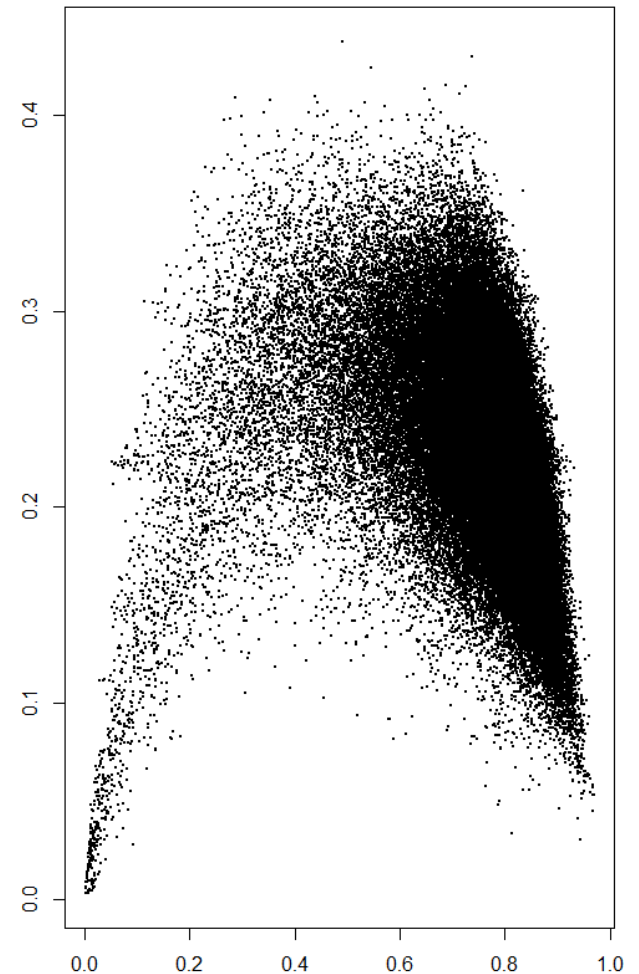
- Exploration or Presentation Both
- Effectiveness Good
- Scalability Intermediate

Scatterplot Options / Problems

Large Data

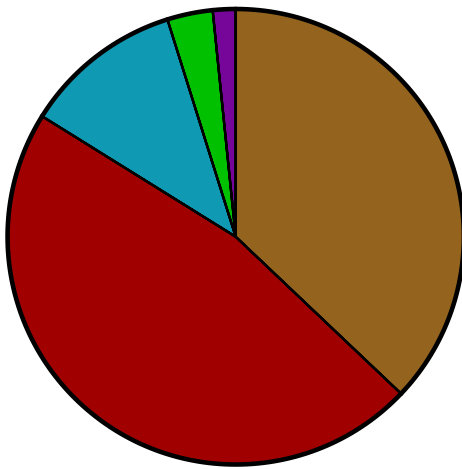


Equality of Axes

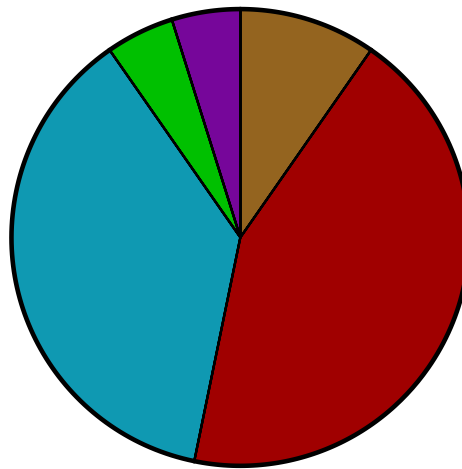
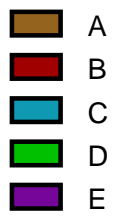


Composition

Pie Charts



Total=62

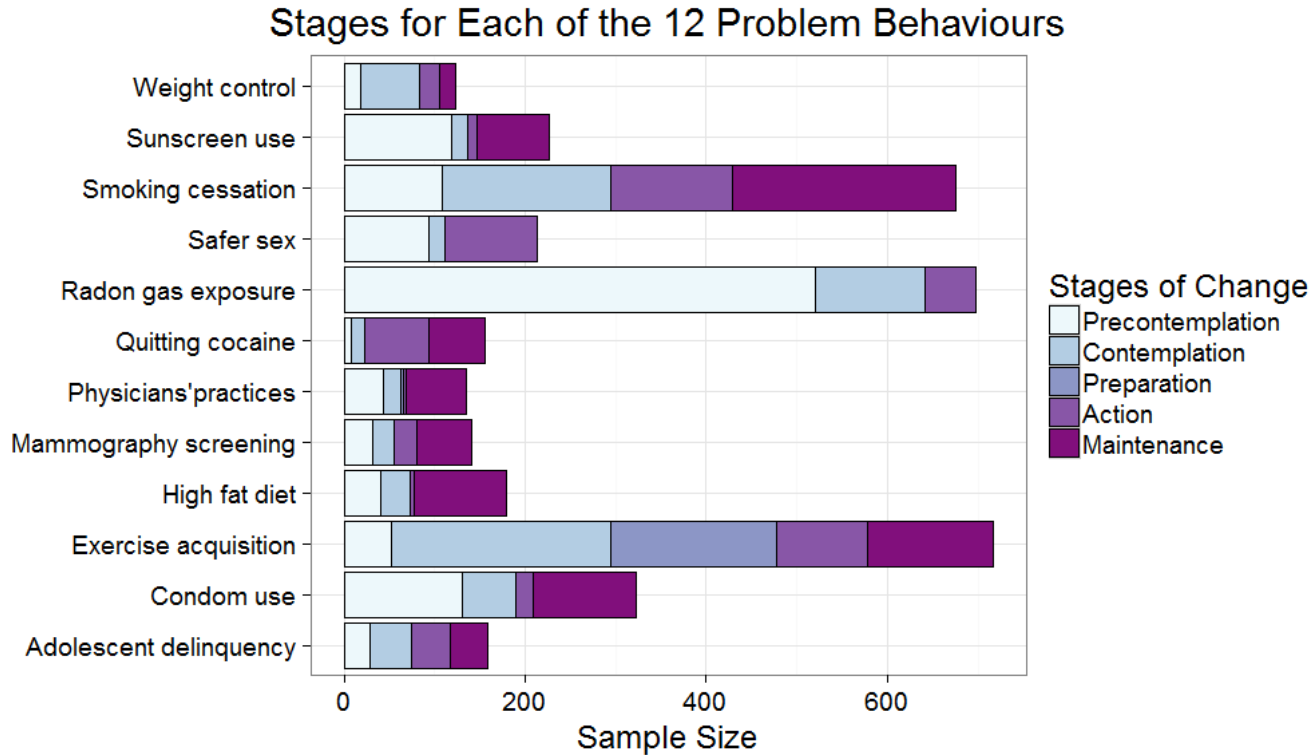


Total=62



- Exploration or Presentation Both
- Effectiveness Intermediate
- Scalability Poor

Stacked Bar Charts



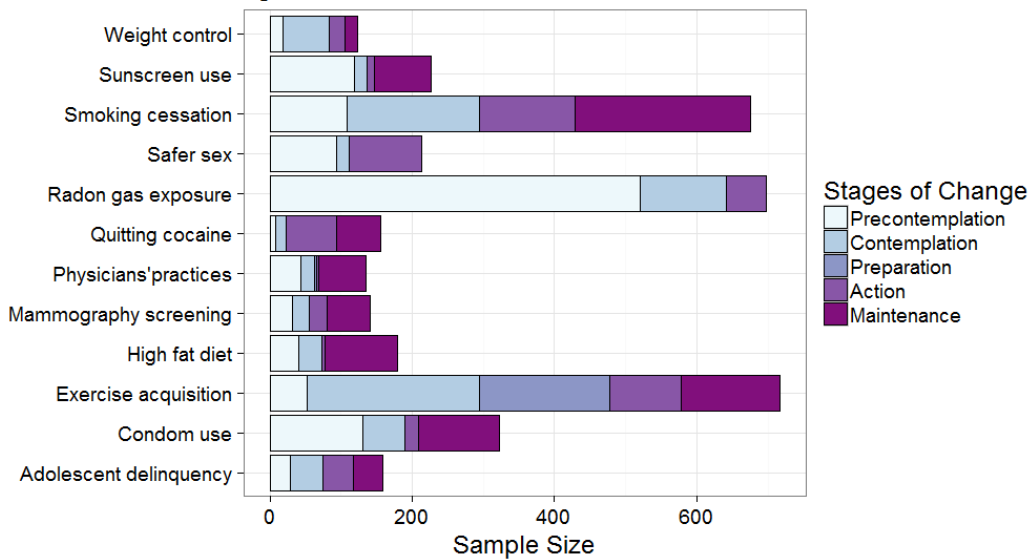
- Exploration or Presentation
- Effectiveness
- Scalability

Both
 Good / Intermediate
 Intermediate

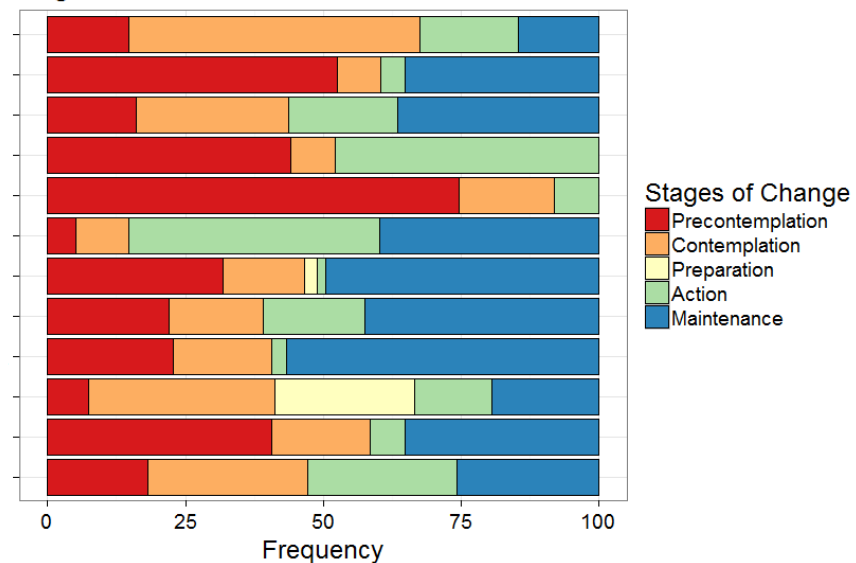
Stacked Bar Chart Options

Scaling and Ordering

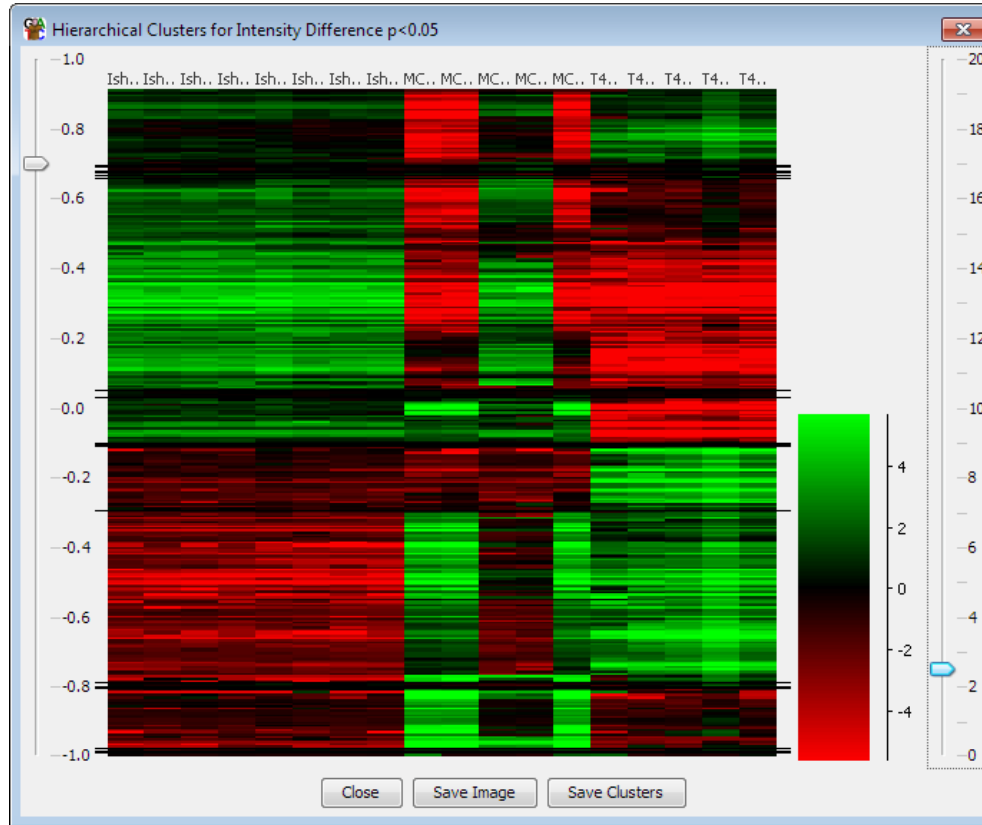
Stages for Each of the 12 Problem Behaviours



Stages for Each of the 12 Problem Behaviours



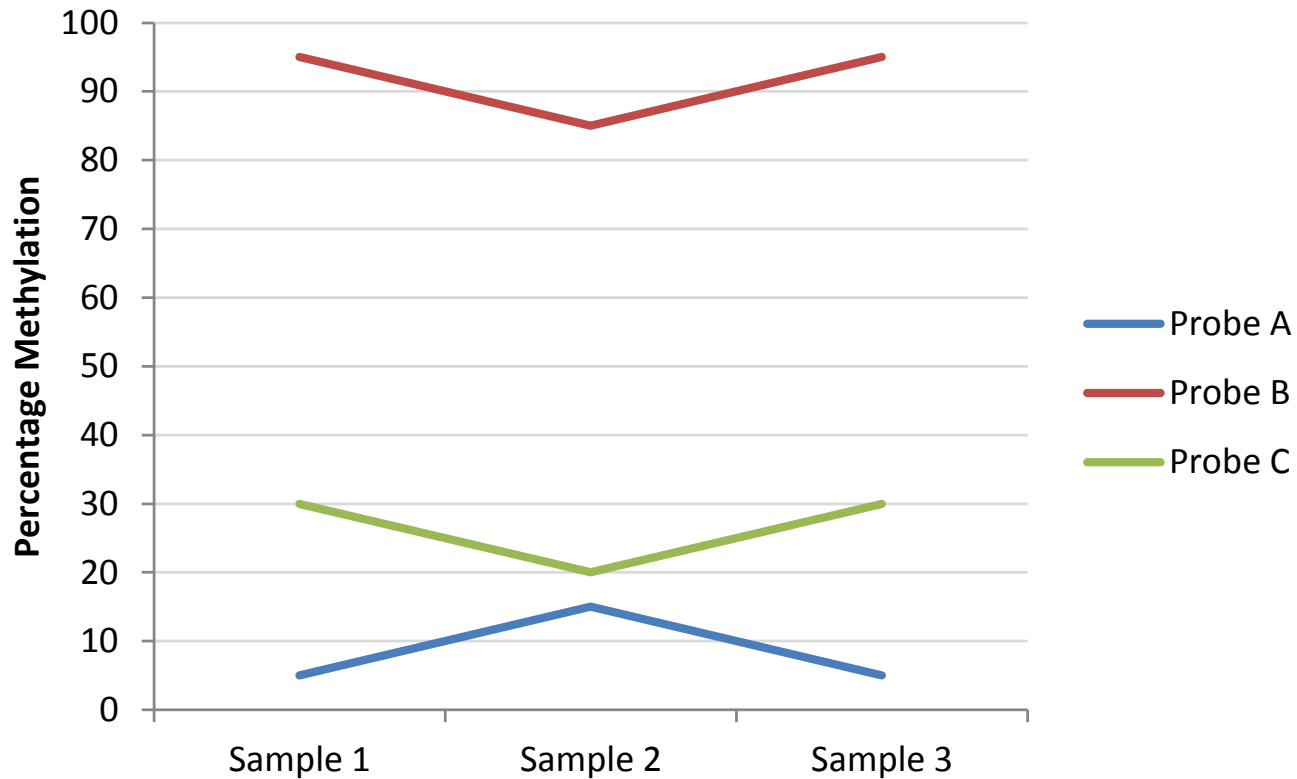
Heatmaps



- Exploration or Presentation Both
- Effectiveness Poor
- Scalability Excellent

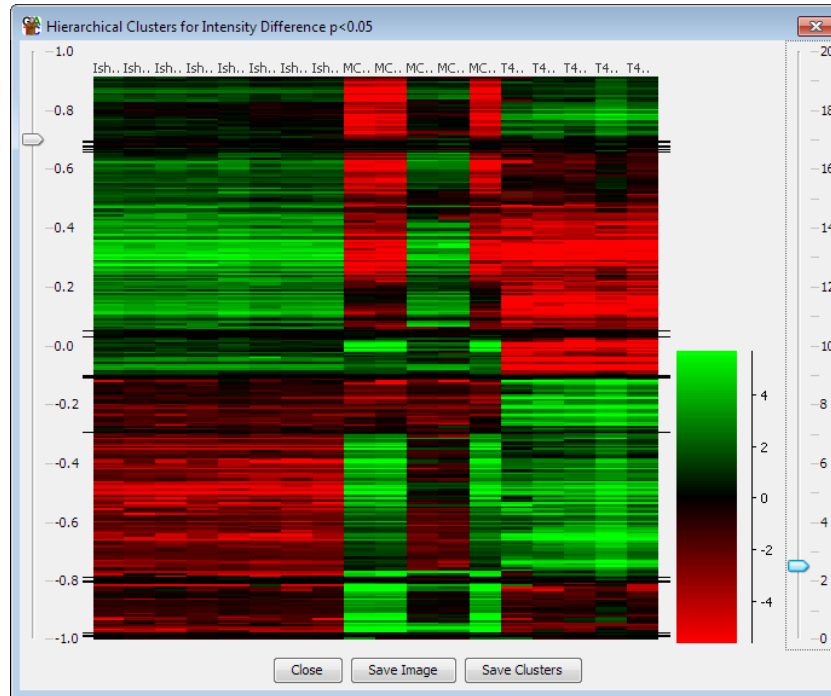
HeatMap Options

Clustering



HeatMap Options

Colours



Turns quantitative differences into categorical

Ethics of data representation

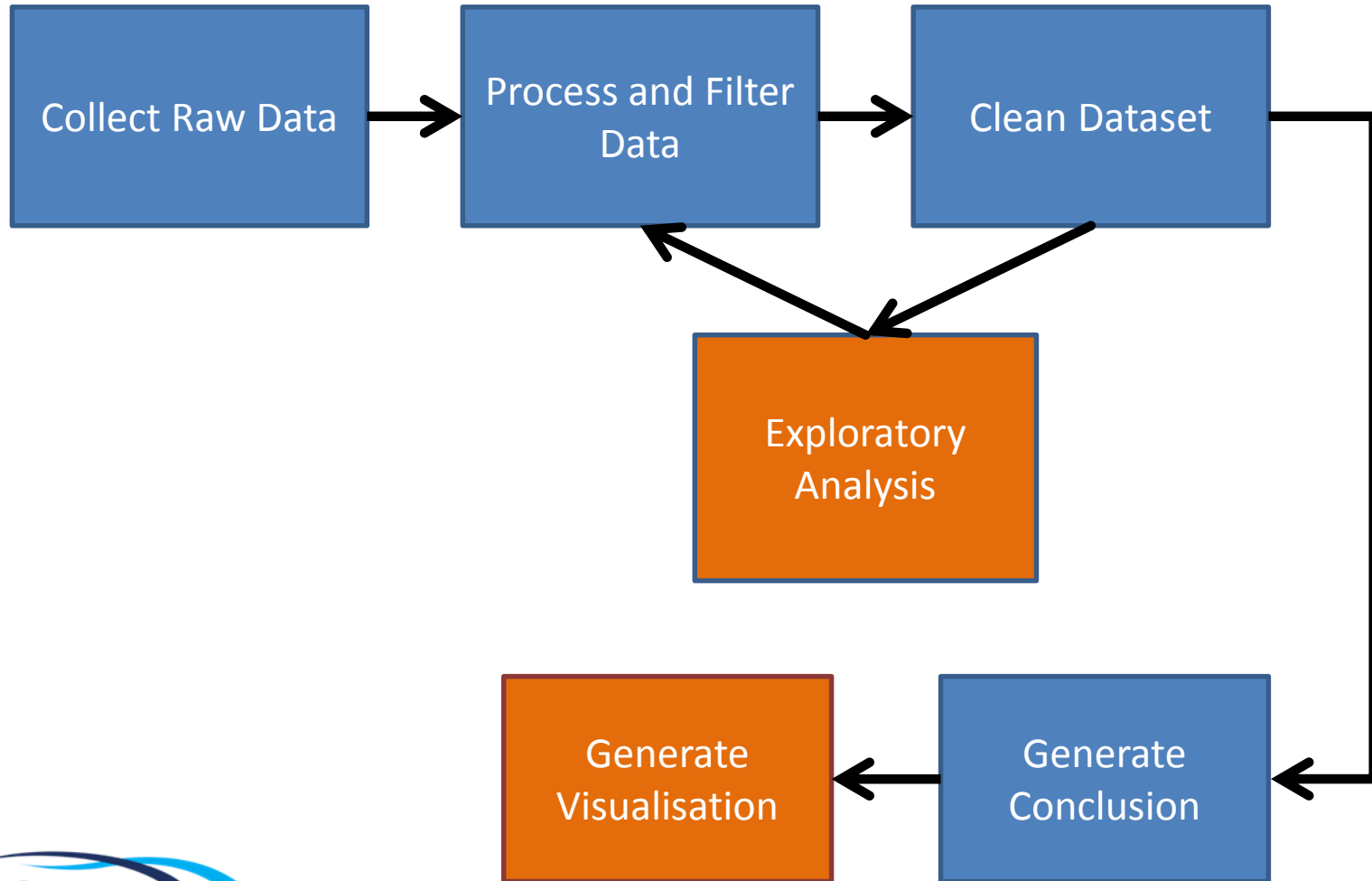
v2018-02

Simon Andrews, Anne Segonds-Pichon

simon.andrews@babraham.ac.uk

anne.segonds-pichon@babraham.ac.uk

Data Visualisation Process



What is Ethics

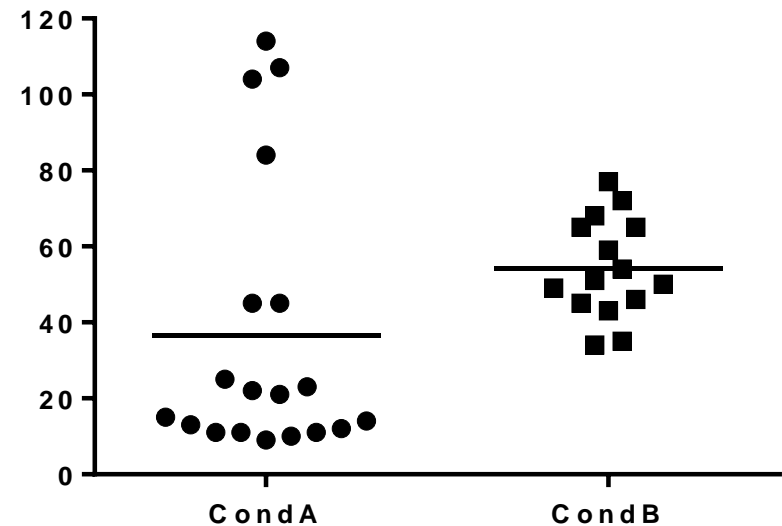
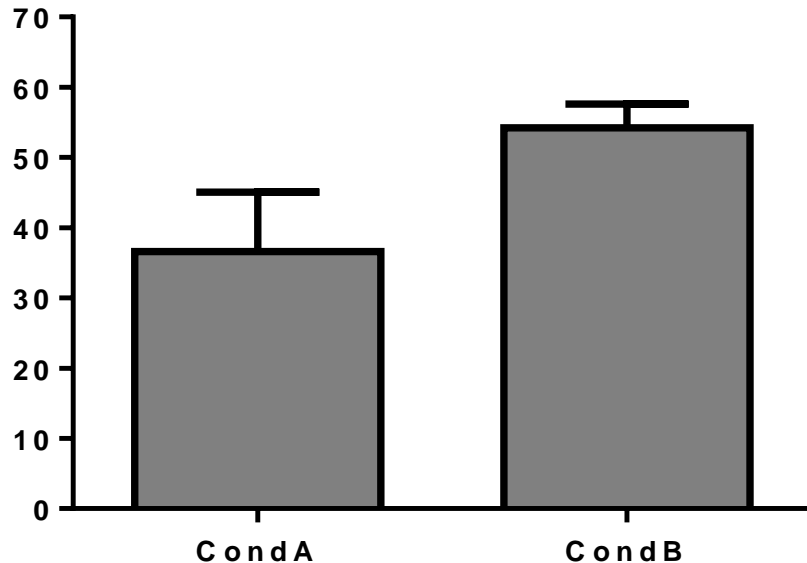
when it comes to data visualisation?

- The figure/graph/image should show what is actually happening and not what you want to happen.
- Different ways of being unethical:
 - not exploring/getting to know the data well enough,
 - misusing your chosen graphical representation.
 - deliberately showing the data in a misleading manner,
 - choosing the ‘most representative’ image/experiment.

Not exploring/getting to know the data well enough

Example 1

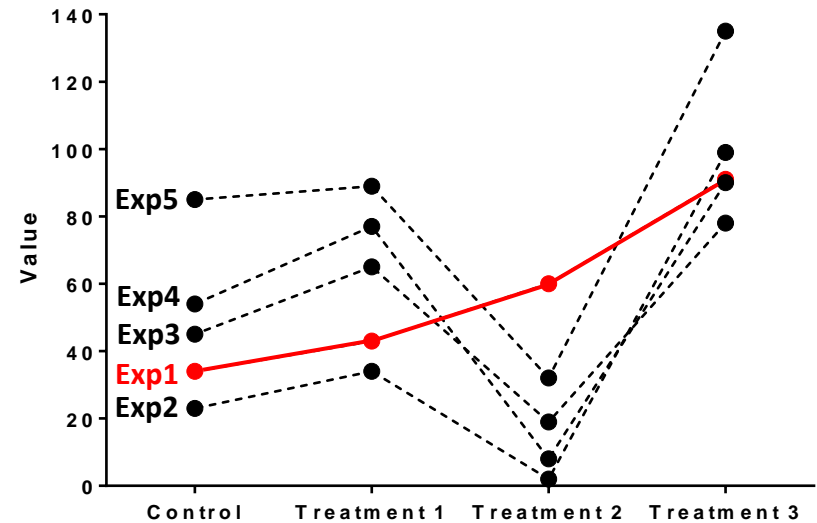
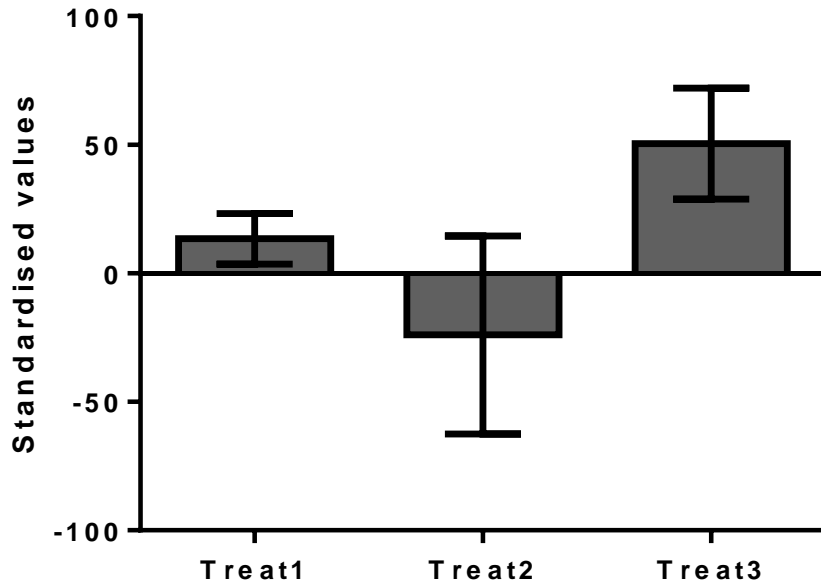
- One experiment: change in the variable of interest between CondA to CondB.
 - ❖ Data plotted as a **bar chart**.



Not exploring/getting to know the data well enough

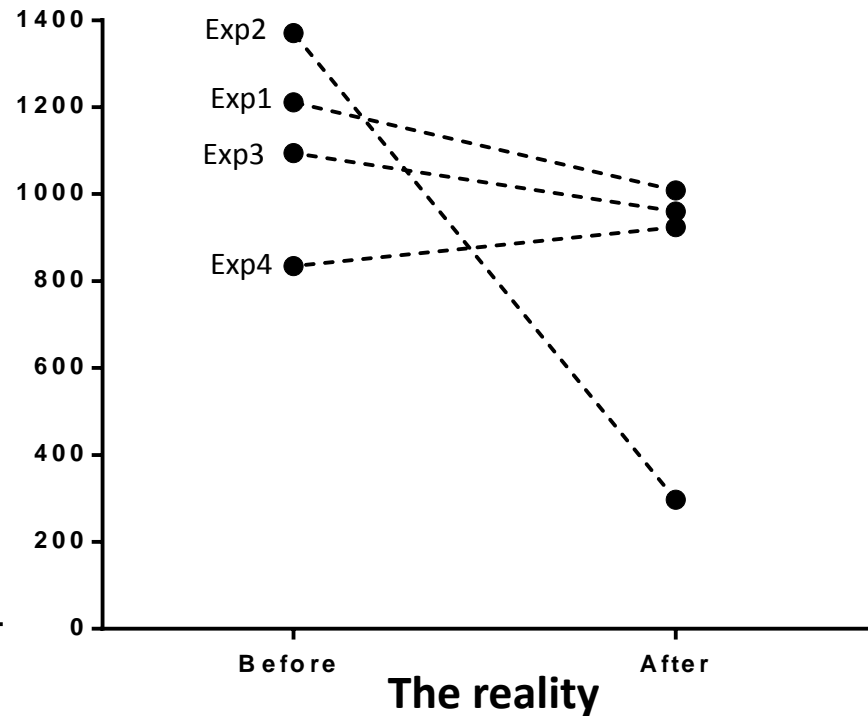
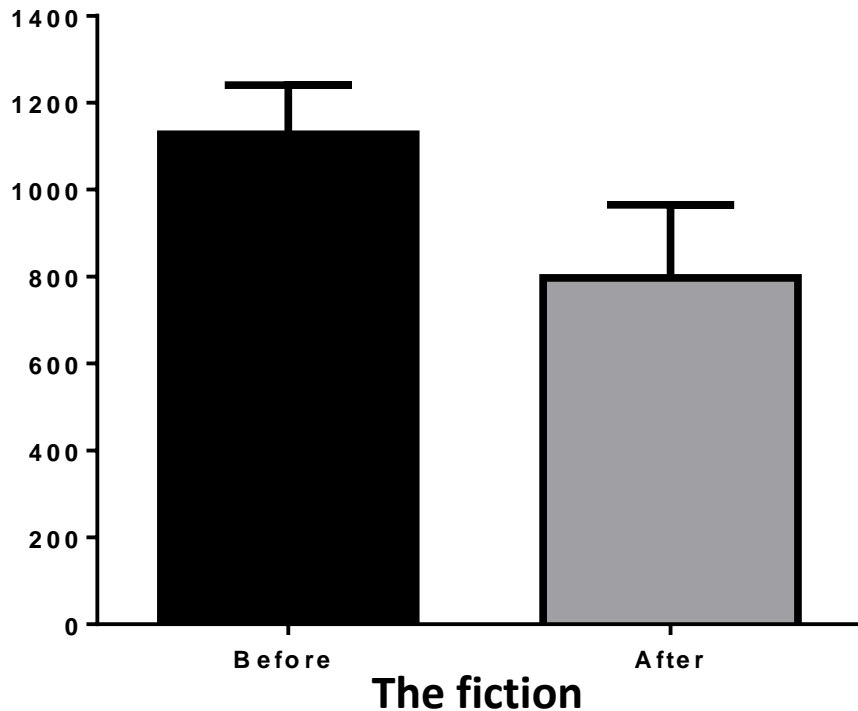
Example 2

- Five experiments: change in the variable of interest between 3 treatments and a control.
 - ❖ Data plotted as a **bar chart**.



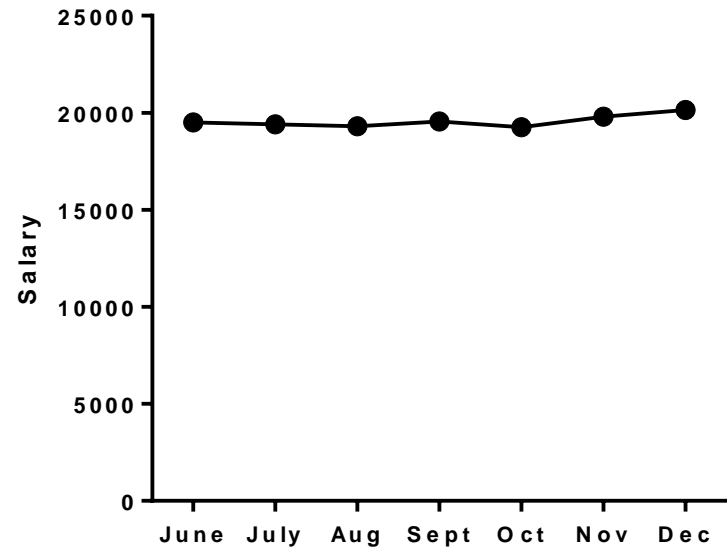
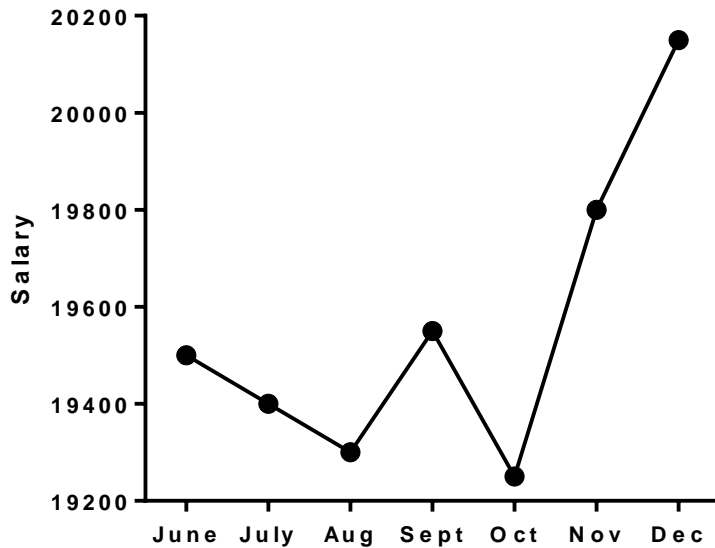
Choosing the wrong graph to present the data

- Four experiments: Before-After treatment effect on a variable of interest.
- Hypothesis: Applying a treatment will decrease the levels of the variable of interest.
 - ❖ Data plotted as a **bar chart**.



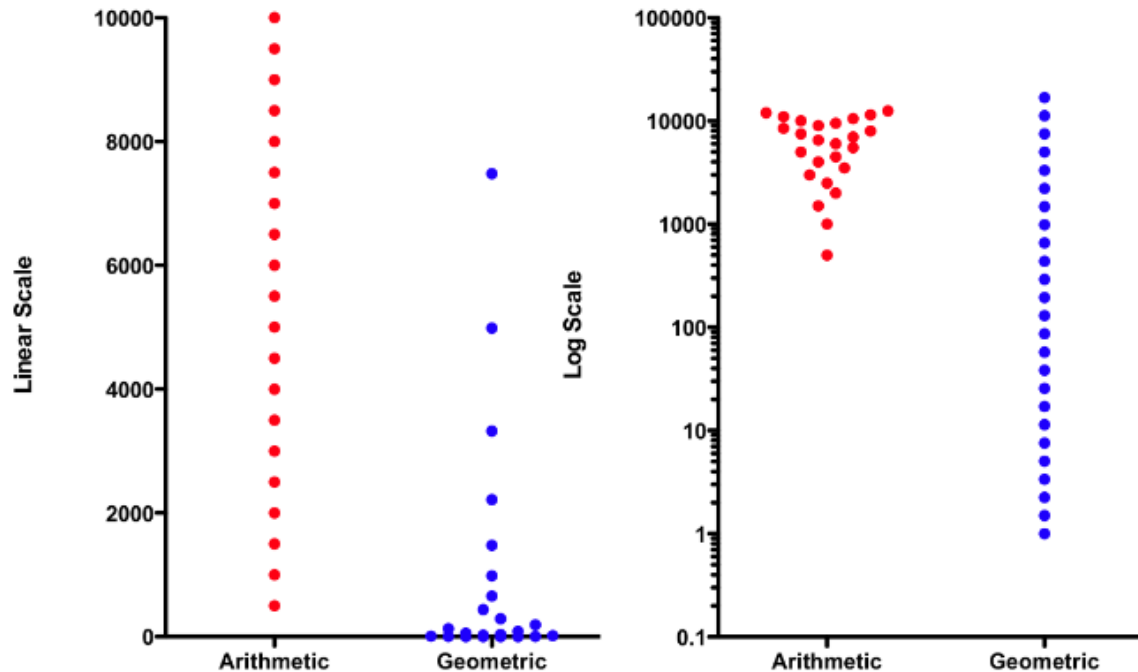
Choosing the wrong axis/scale

- Example: increase in salary in the last term.



Choosing the y-axis/scale

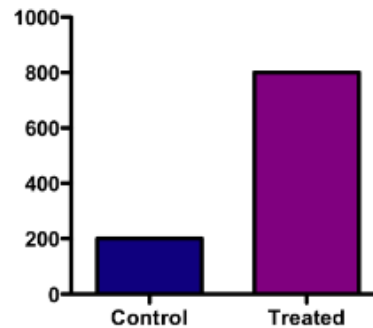
- Be careful with Linear vs. logarithmic scale.



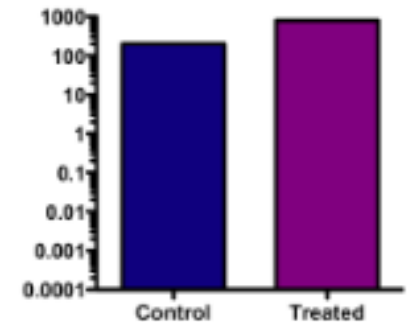
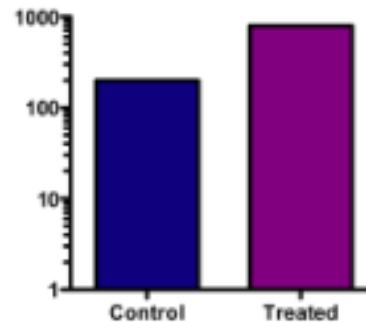
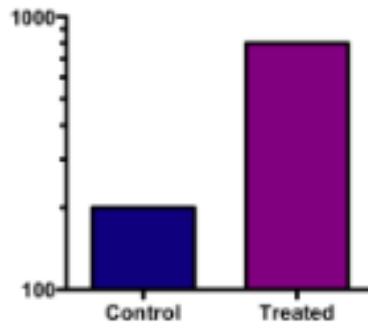
Choosing the y-axis/scale

- For **cheating**, a bar graph using a **log axis** is a **great tool**, as it lets you either exaggerate differences between groups or minimize them.

Linear scale

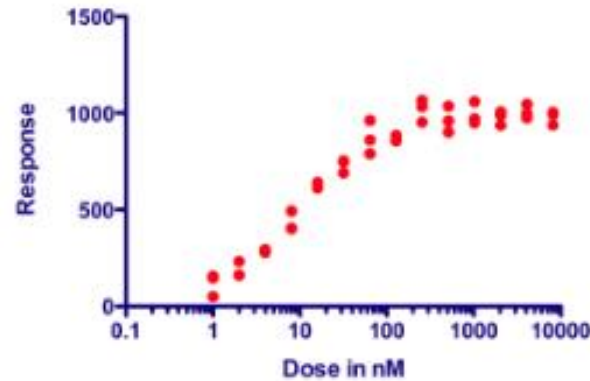
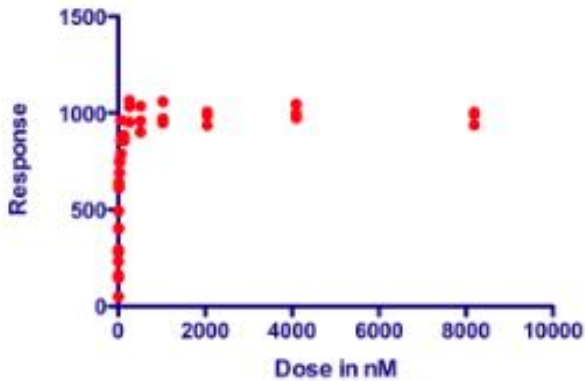


Logarithmic scale

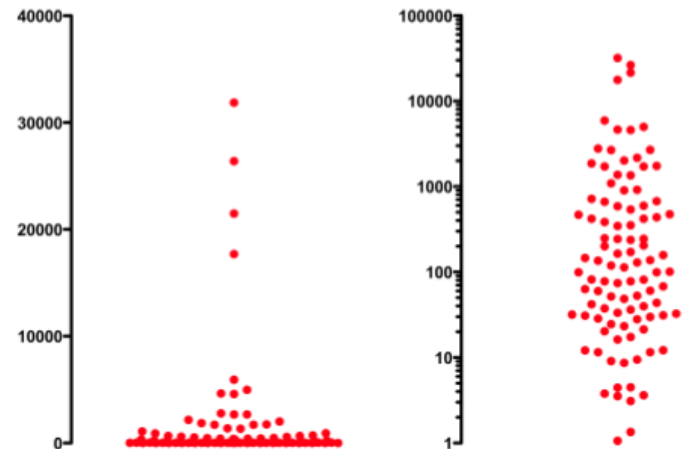


Choosing the y-axis/scale

- **Logarithmic axis** should be used for:



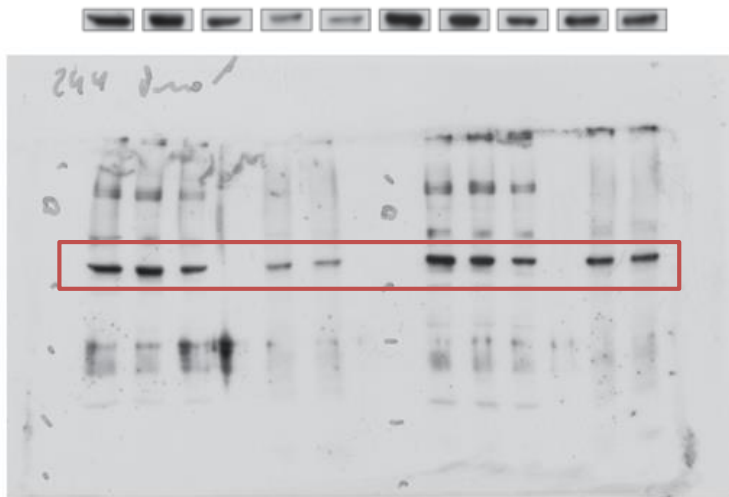
Logarithmically spaced values



Simply Cheating:

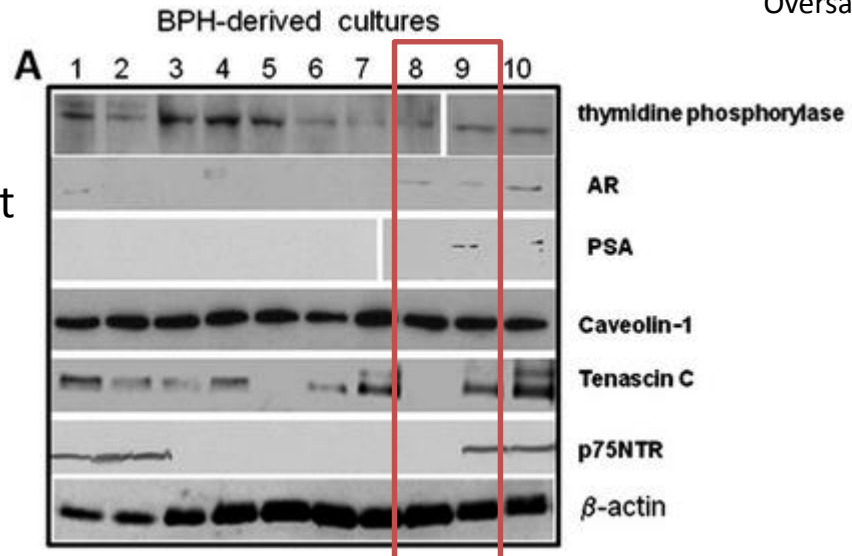
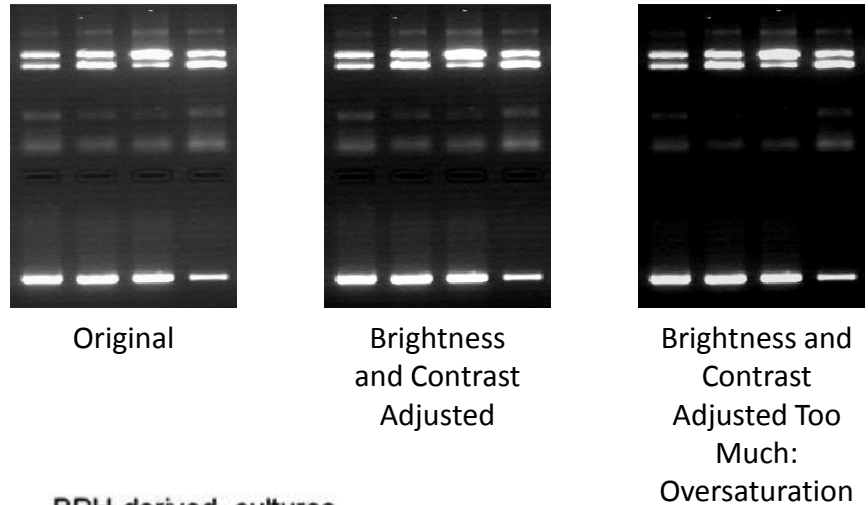
Manipulating images: Western blot

- Presenting bands out of context



- 'Rebuilding' a Western blot from several cuts

- 'Playing' too much with contrast



Is my plot ethical?

Would a reader come to a different conclusion if they could see the details of the data which were omitted from the plot?