

Plot type specific considerations

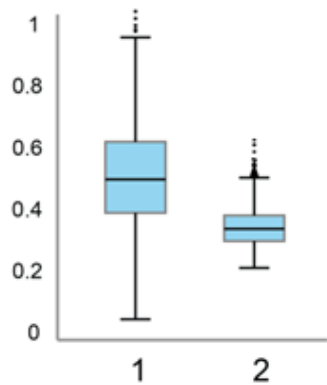
V2017-06

Anne Segonds-Pichon
Simon Andrews
Phil Ewels

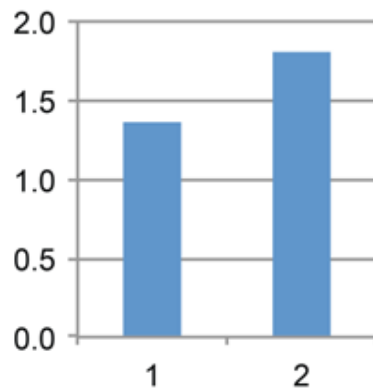
anne.segonds-pichon@babraham.ac.uk
simon.andrews@babraham.ac.uk
phil.ewels@scilifelab.se

Types of plot

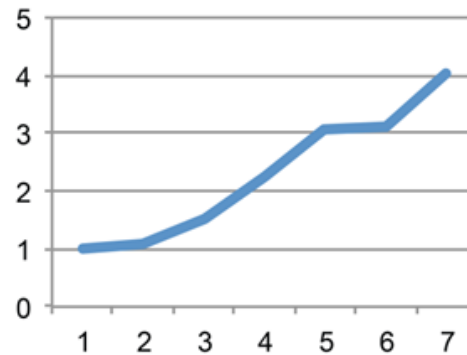
Things you can illustrate



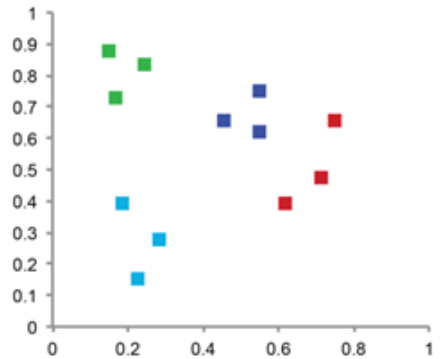
Distribution



Comparison



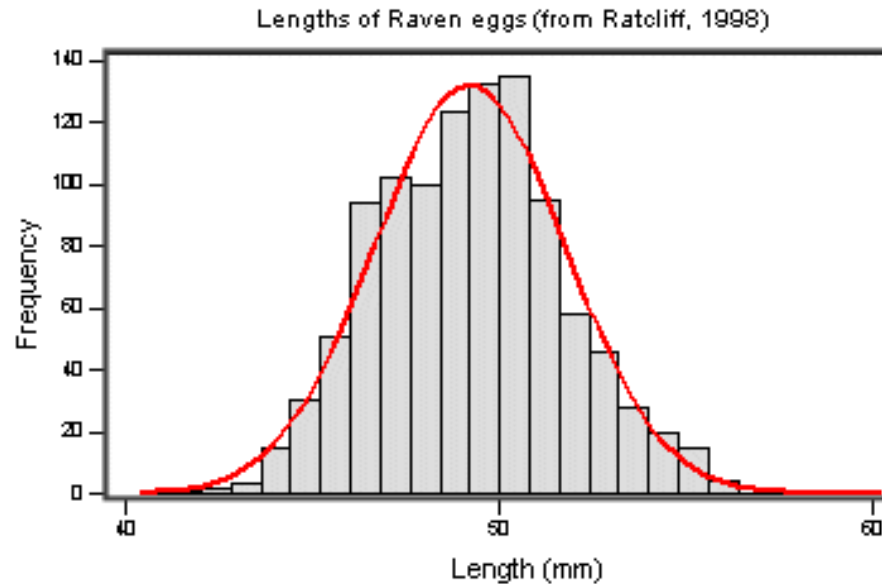
Relationship



Composition

Plot types – Distribution/Exploration

Histograms

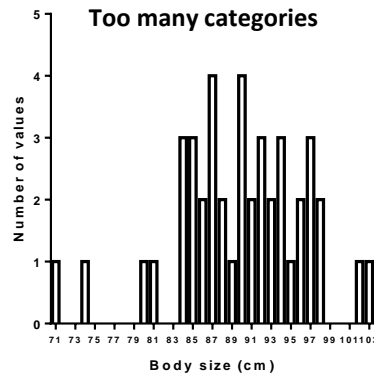
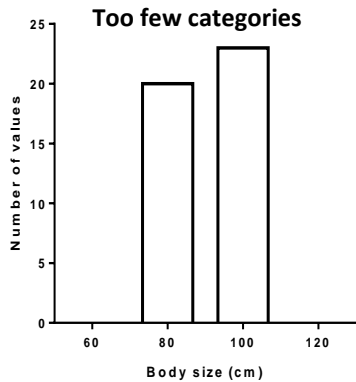


- **Very good for** exploring data. Better on big dataset.
- Rules: Number of intervals $\approx \sqrt{N}$ and Interval width $\approx \text{Range} \div \sqrt{N}$
- **Histograms are great but** careful with the resolution (= number of bins) as it affects the shape of the distribution.

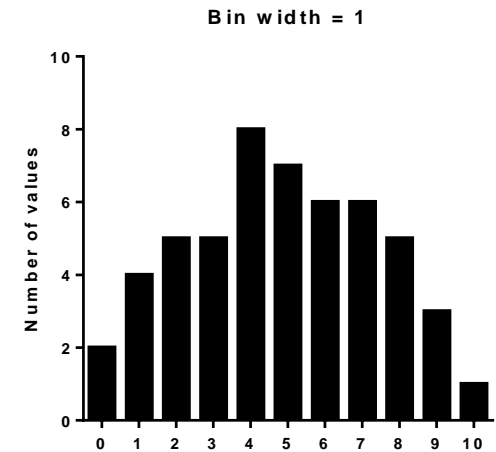
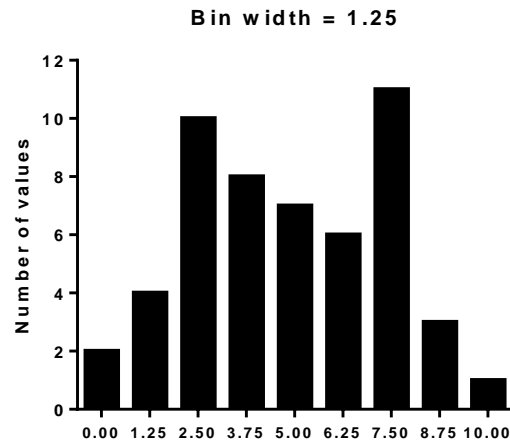
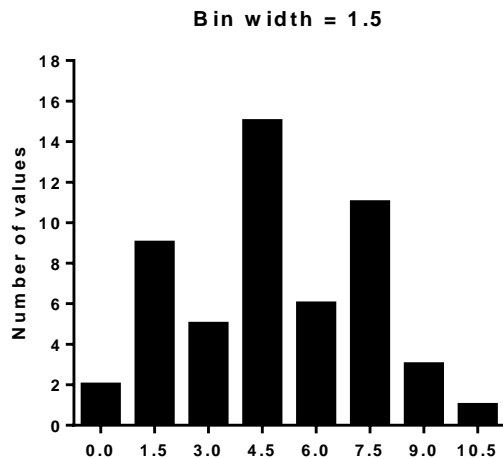
Plot types – Distribution/Exploration

Histograms

- Be careful with the resolution ...



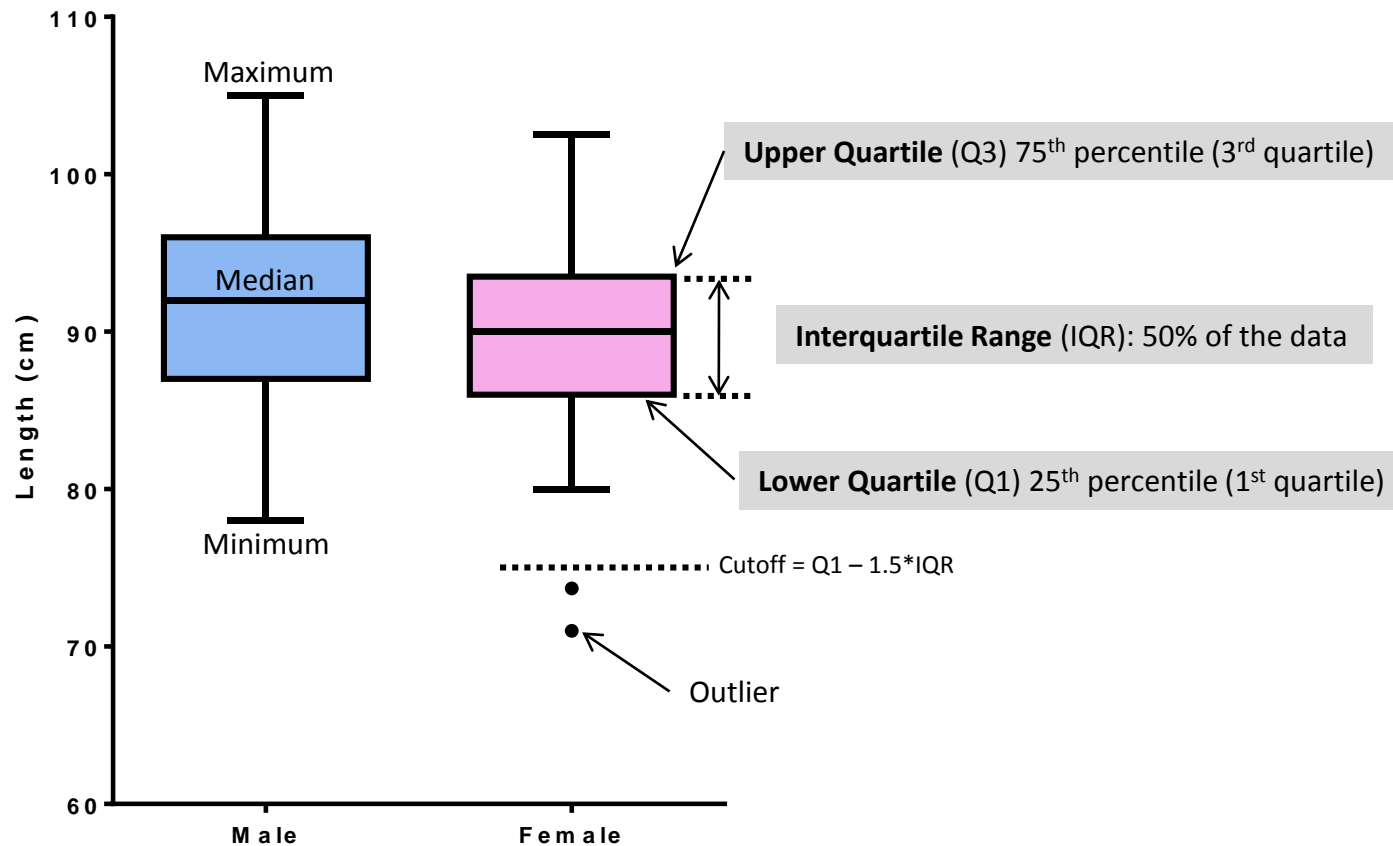
... and the type of data you are dealing with.



- Histograms are great but careful with discrete data.

Plot types – Distribution/Exploration

Boxplots and Bean plots



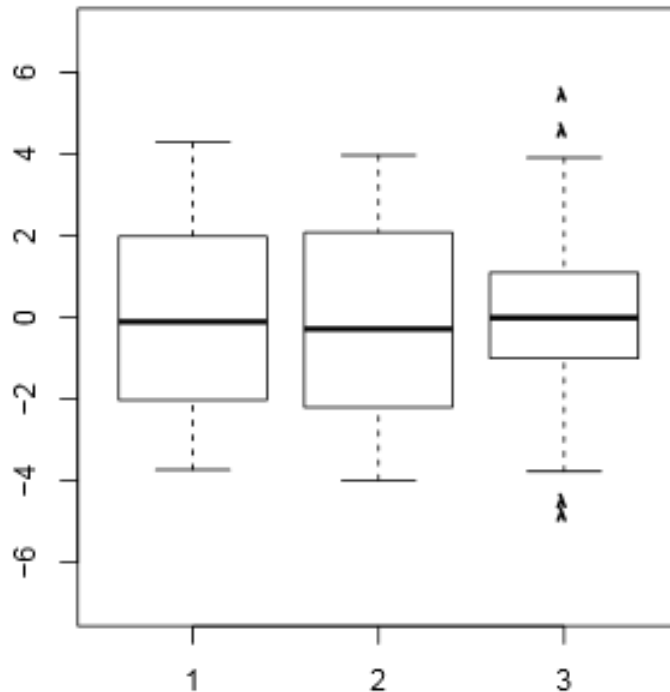
Plot types – Distribution/Exploration

Boxplots and Bean plots

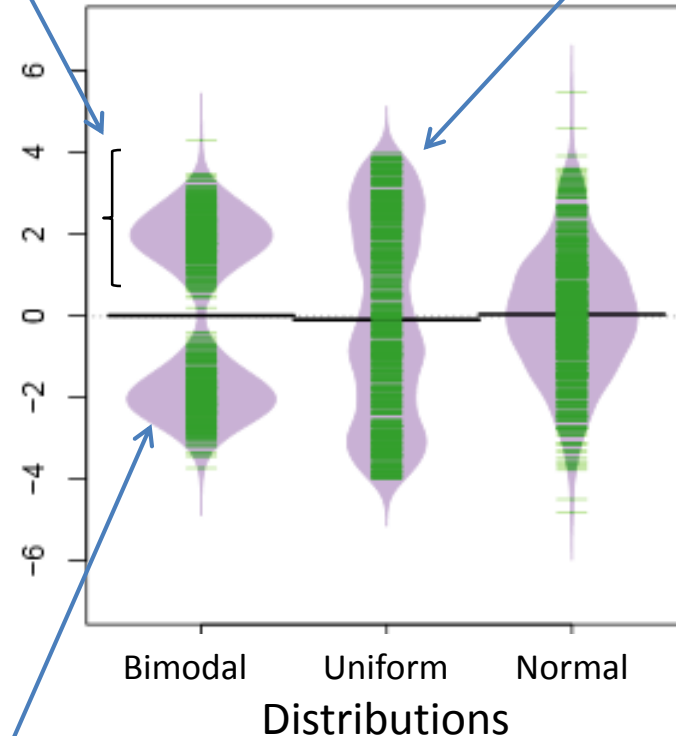
Scatterplot shows individual data

A bean = a 'batch' of data

boxplot



beanplot

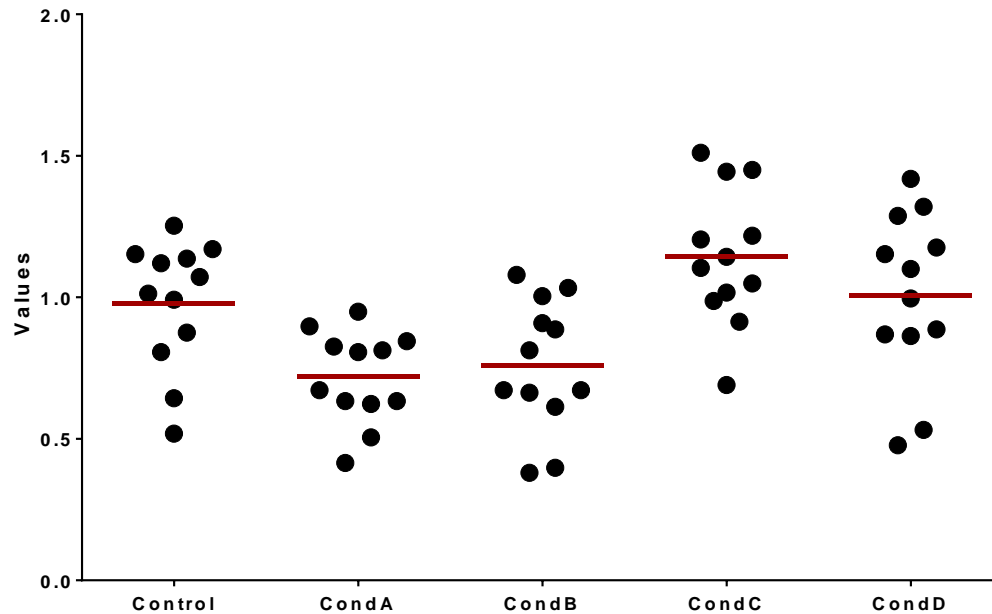


Data density mirrored by the shape of the polygon

- **Very good for** exploring data. Better on medium size dataset.
- **Boxplots are great but** be careful with underlying distribution.

Plot types – Exploration/Comparison

Stripcharts/Scatterplots

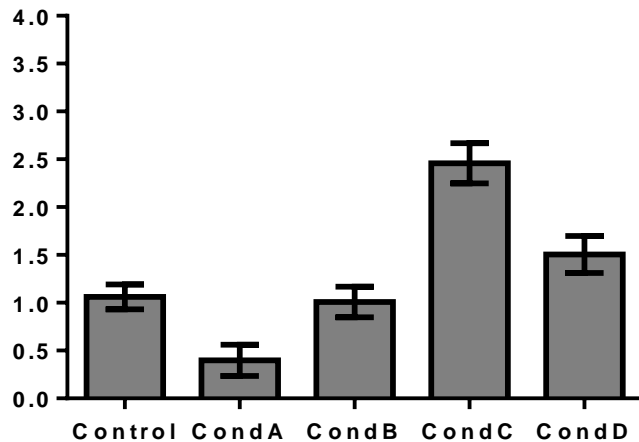


- **Very good for** exploring data. Better on small/medium dataset (up to 100-ish).
- Very informative: exploration AND comparison.
- Very hard to cheat with these.
- **Stripcharts are great but** they don't work so well with big samples.

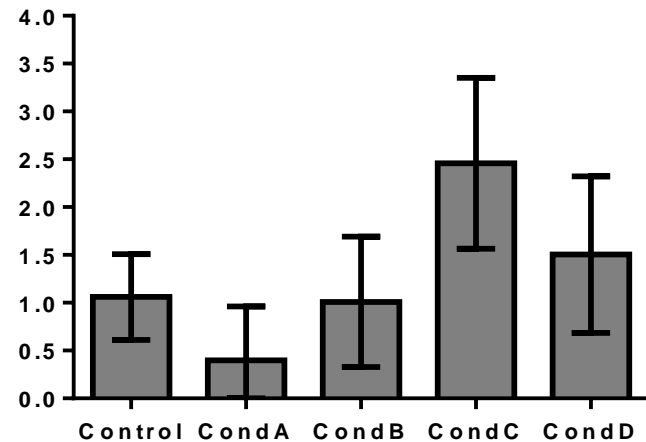
Plot types – Comparisons

Barcharts *(not for exploration!)*

Standard error



Standard deviation

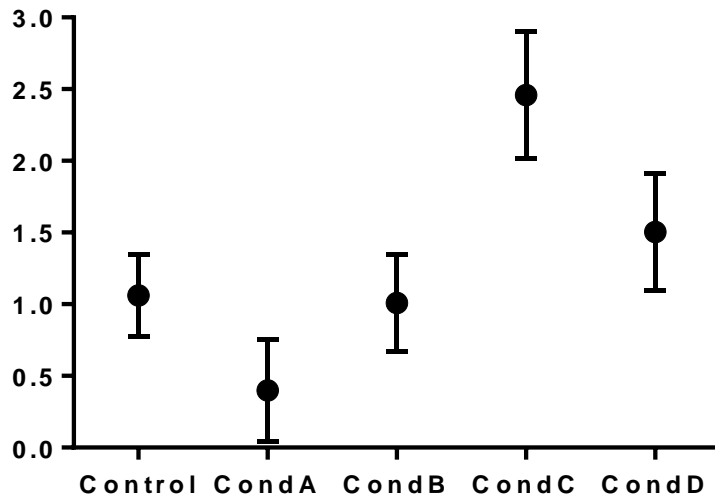


- **Very good for** presenting results and emphasizing differences.
 - Effectiveness: most important info with the most effective channel.
- **Barcharts are great but** higher bars get more attention.

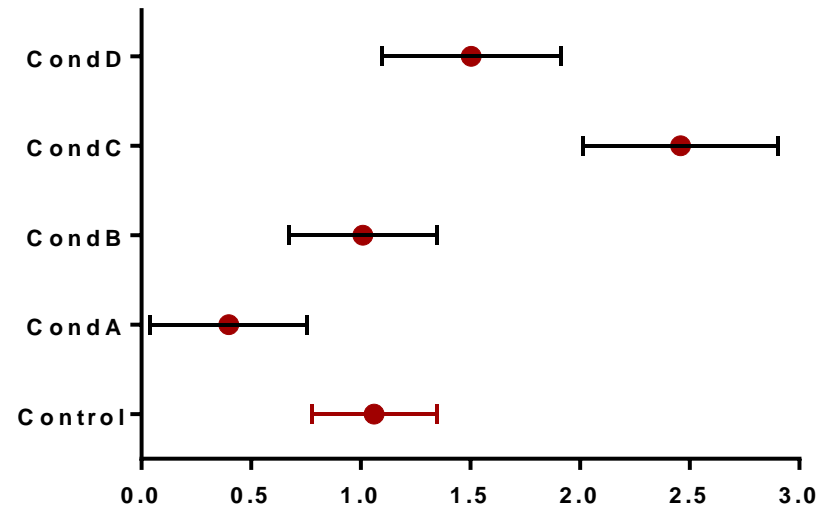
Plot types – Comparisons

Alternative to **Barcharts**

Confidence interval



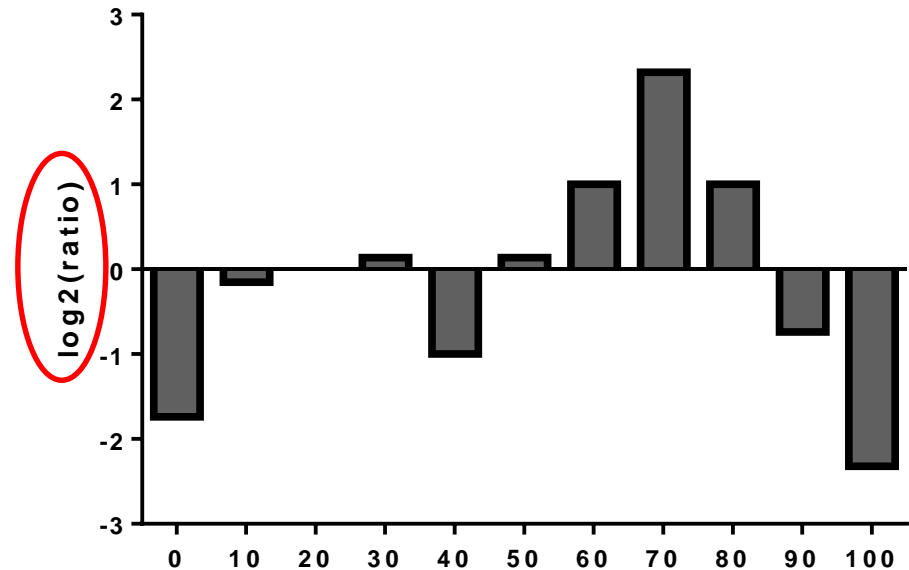
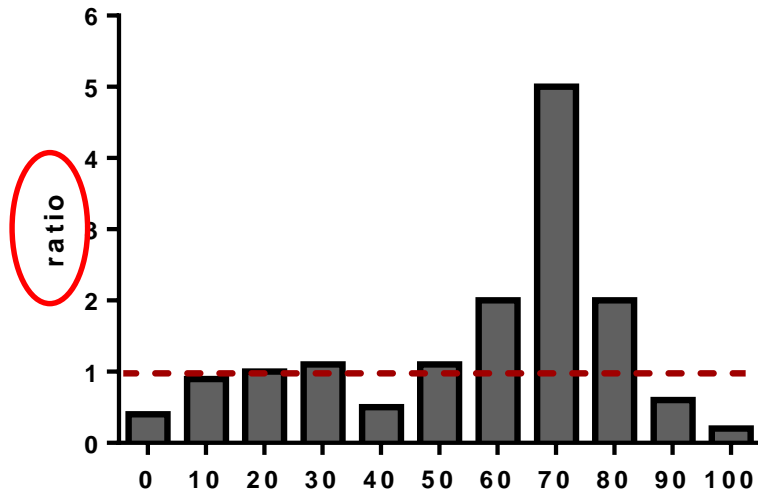
Star wars



Plot types – Comparisons

Barcharts

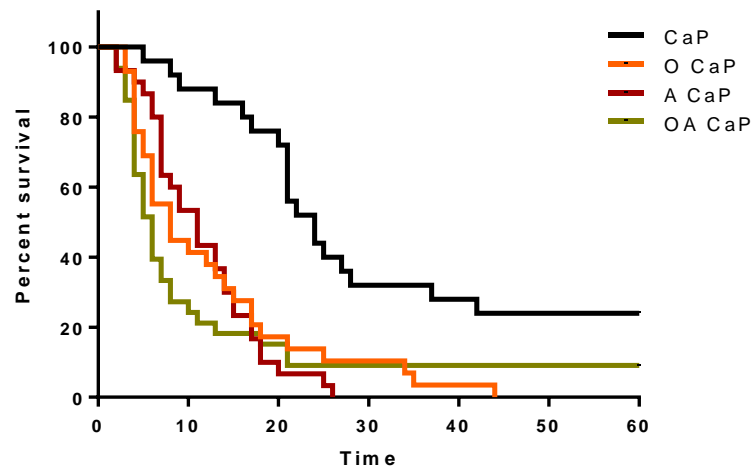
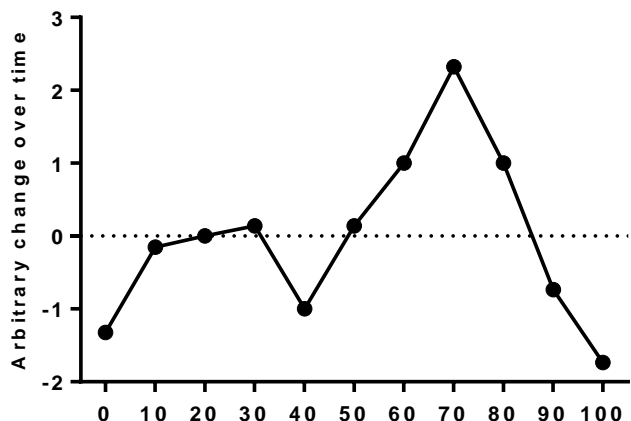
- Be careful with the scale when plotting ratio



- **Very good for** presenting results and emphasizing differences.
- **Barcharts are great but** after data exploration and the y-axis needs to be chosen wisely.

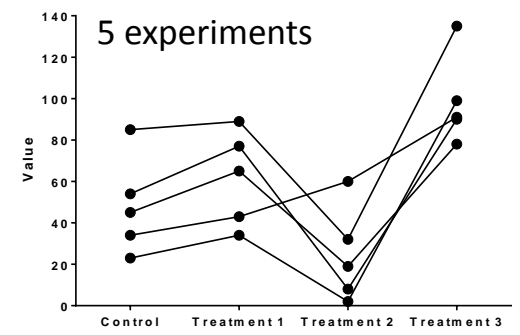
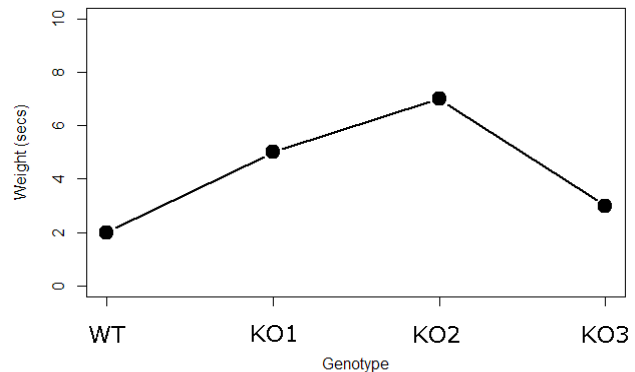
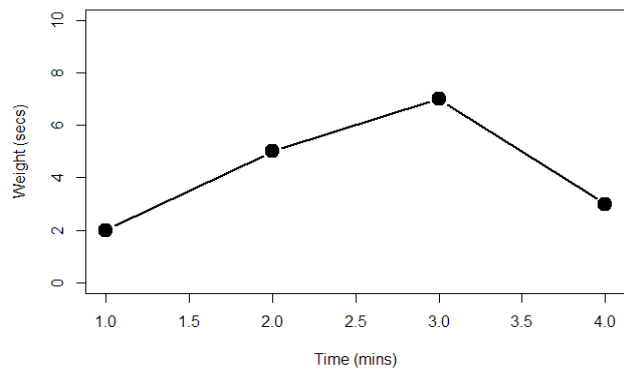
Plot types – Relationship/Comparison

Line graphs



Good line graph

Bad line graph



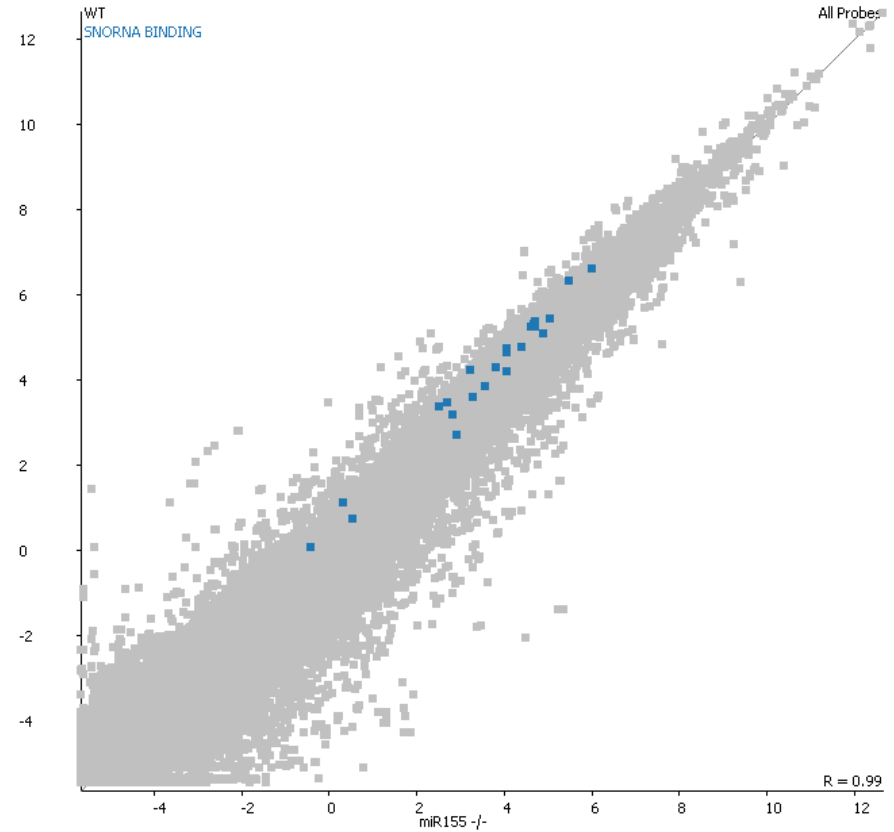
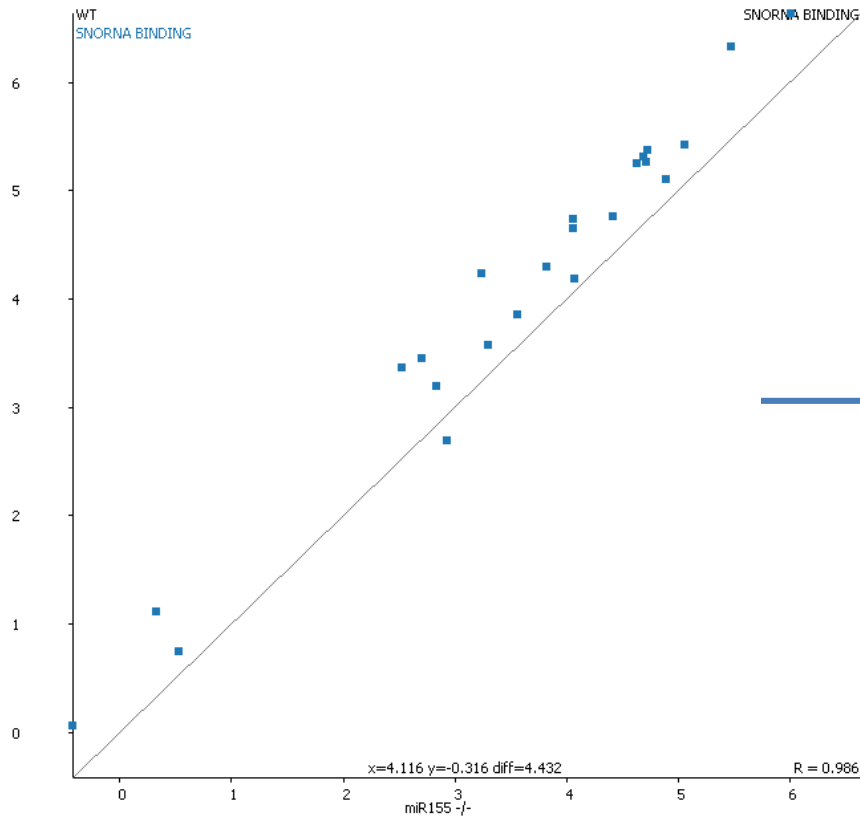
Except for exploration ...



- **Very good for** presenting results of matched/paired/repeated data.
- **Linecharts are great** but careful with the axes.

Plot types – Relationships

Two conditions - **Scatterplot**

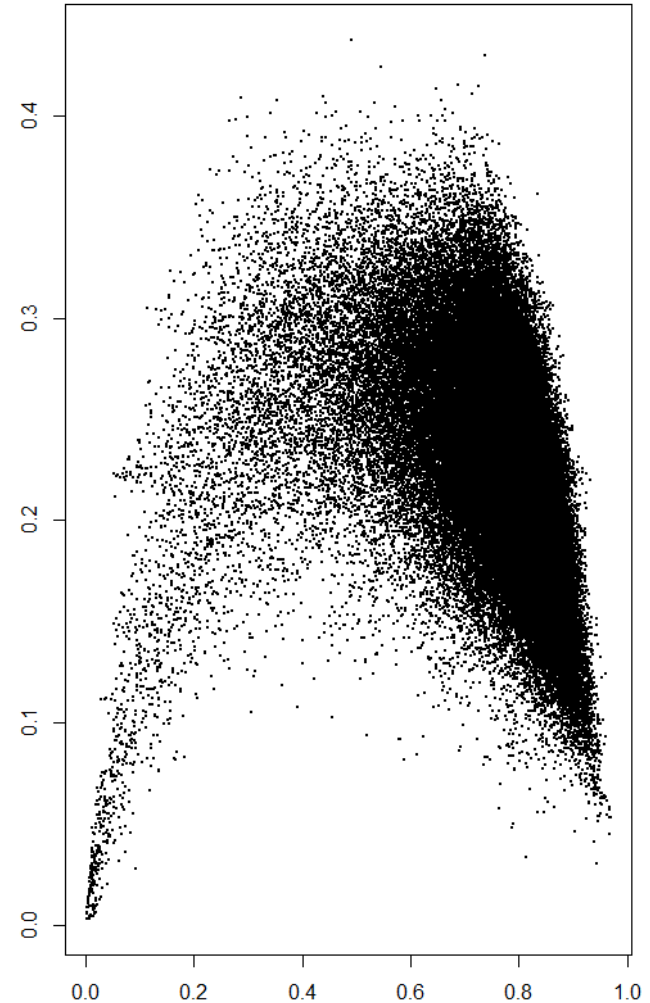
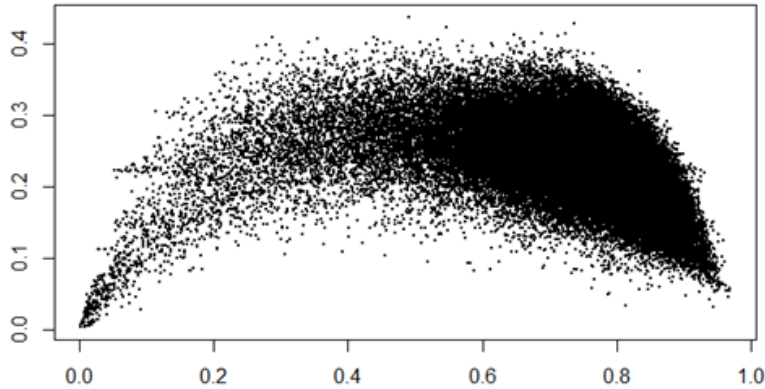


- Very good for understanding the relationship between quantitative variables.

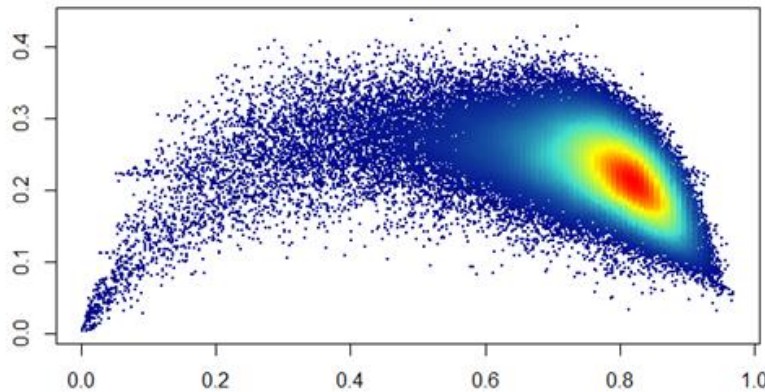
Plot types – Relationships

Scatterplots

- Problem: very big dataset



- Solution: smoothed **densities** colour representation

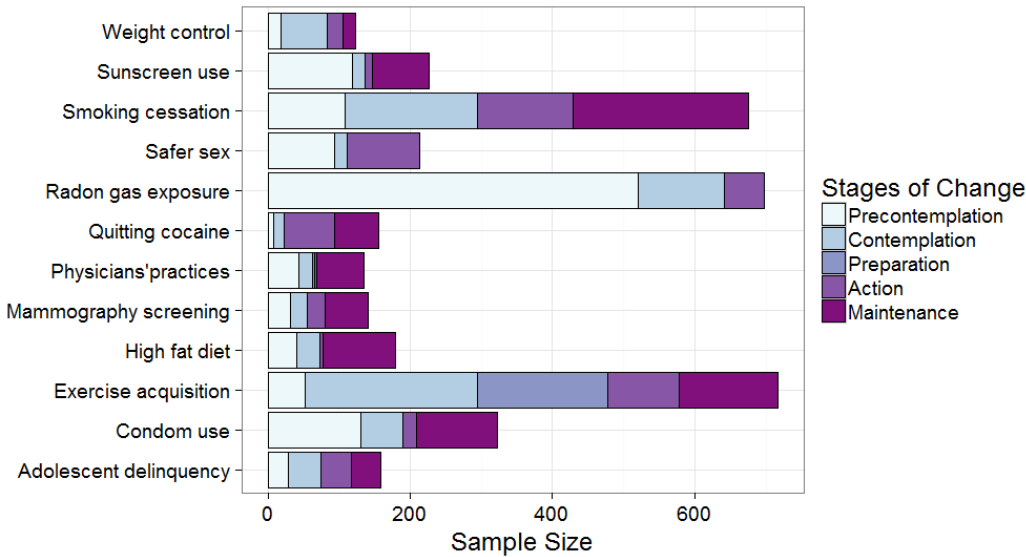


- Scatterplots are great but big data can be tricky.

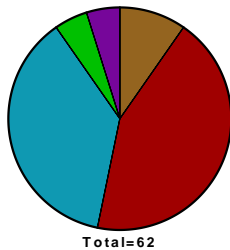
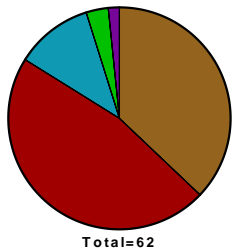
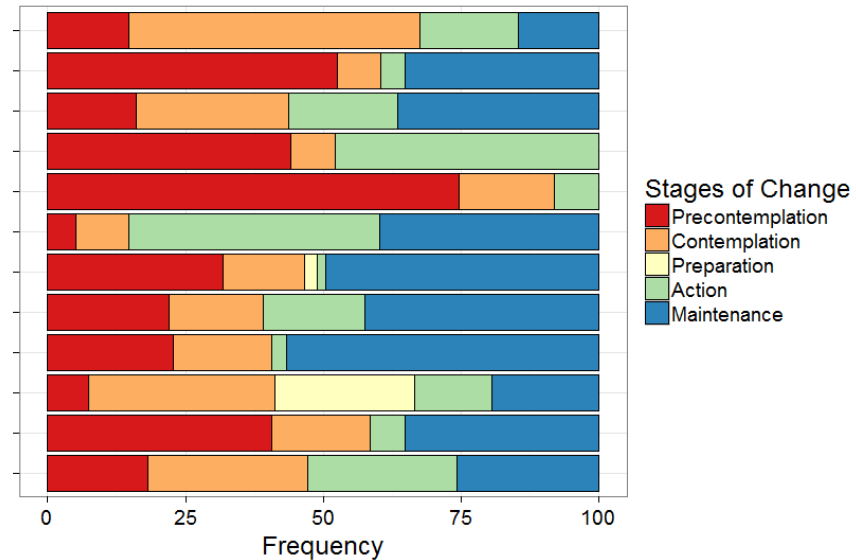
Plot types – Composition

Stack charts/Pie charts

Stages for Each of the 12 Problem Behaviours



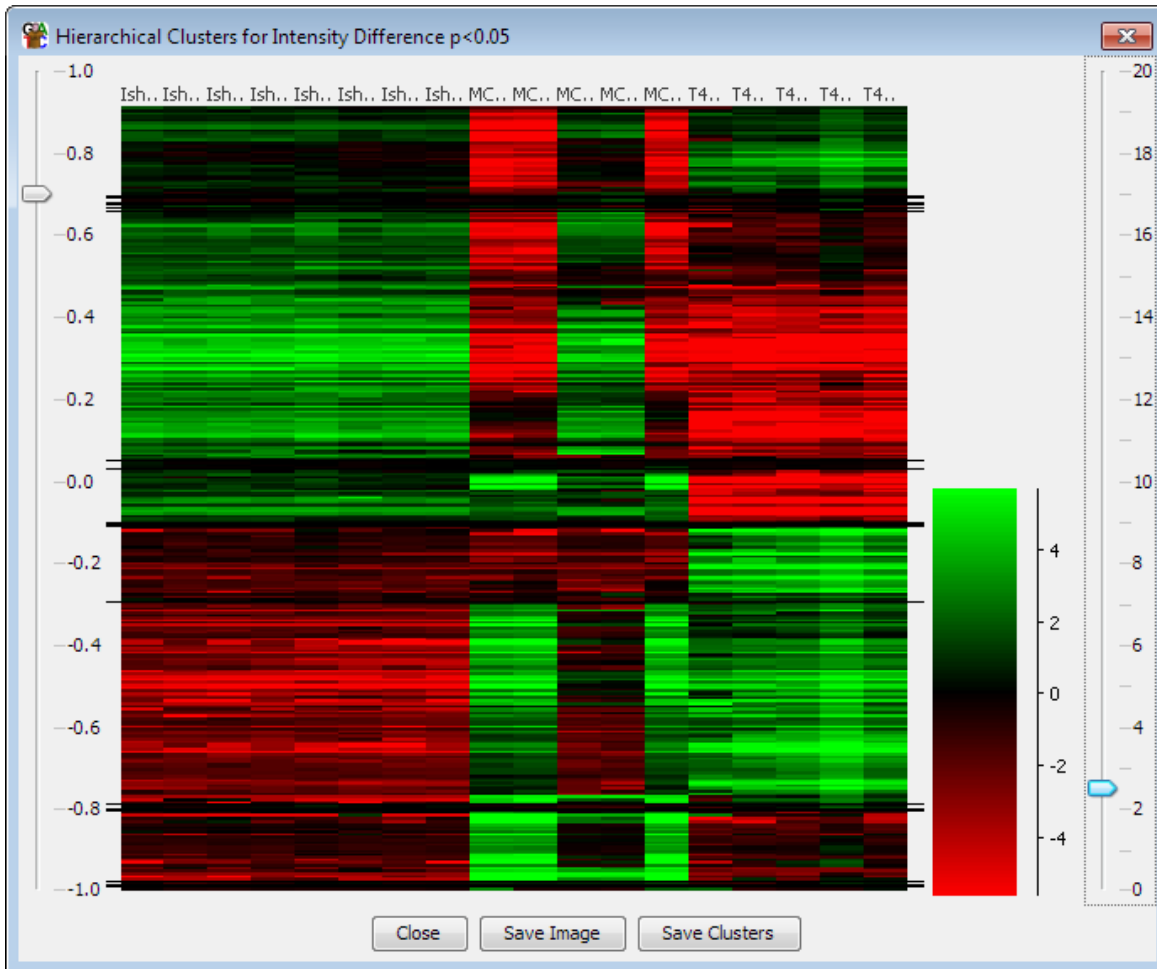
Stages for Each of the 12 Problem Behaviours



- Very good for presenting categorical data.
- **Stack /pie charts are great but keep an eye on the sample size.**

Plot types – Relationships

Many conditions - **Heatmaps**

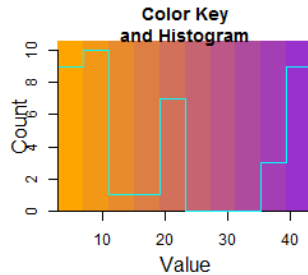


- Normalisation
- Colouring
- Filtering
- Clustering

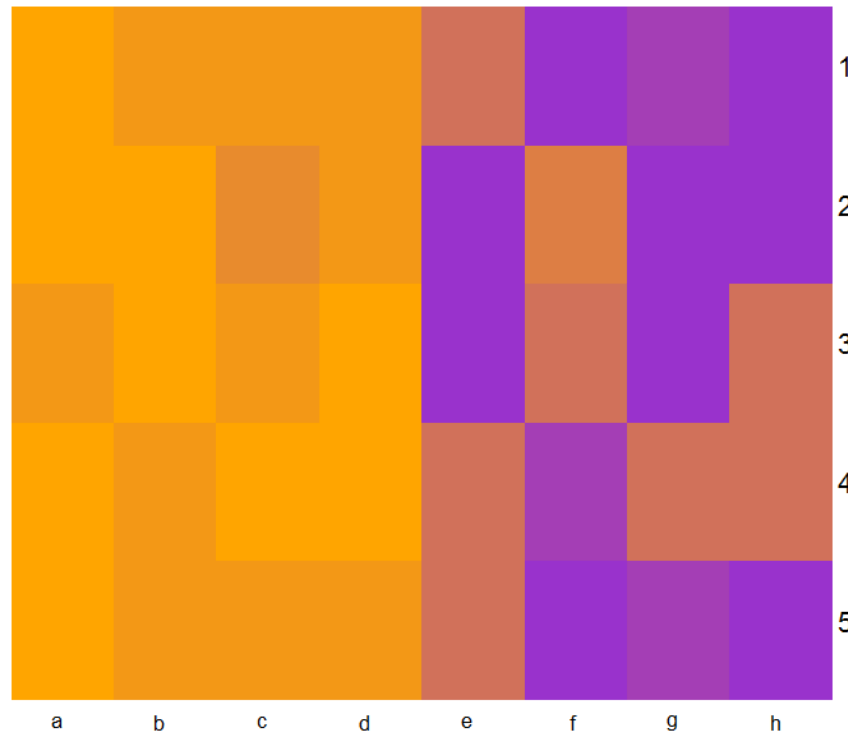
Plot types – Relationships

Heatmaps

- A heatmap is basically a table that has colors in place of numbers.

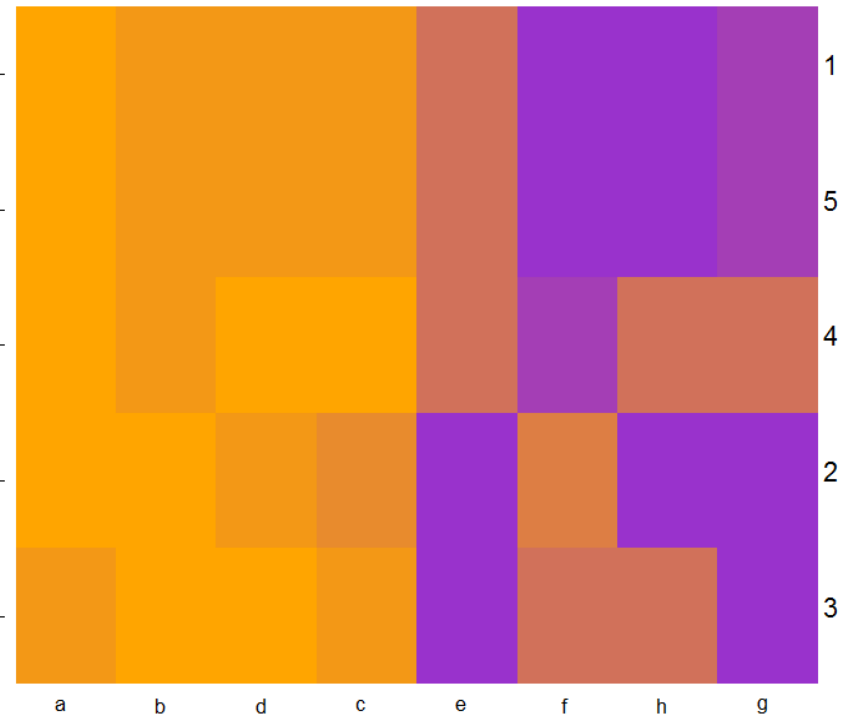
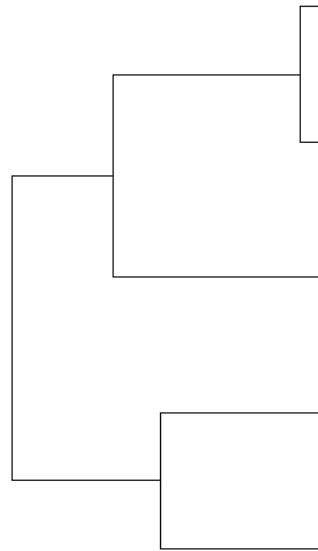
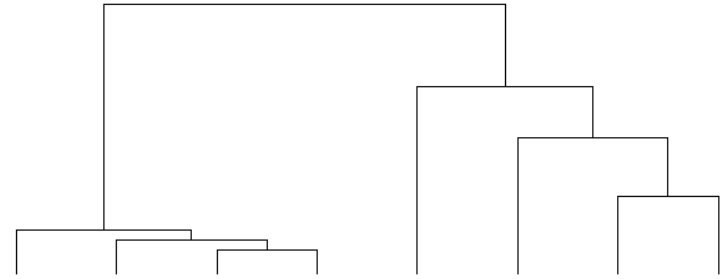
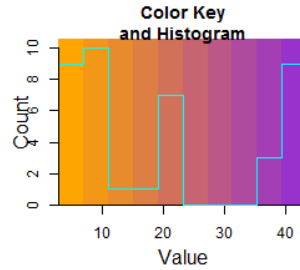
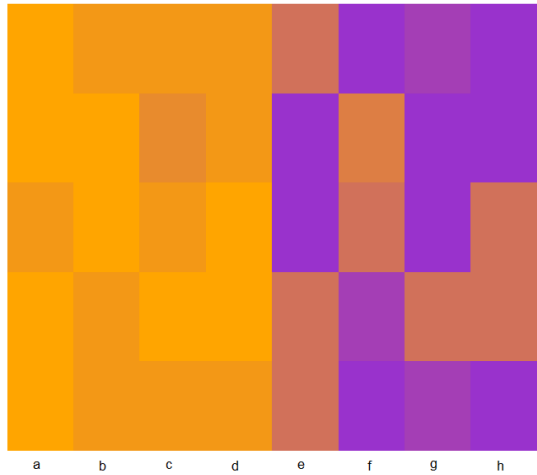


	a	b	c	d	e	f	g	h
1	3.7	10.8	8.4	10.4	19.5	40.6	39.0	40.1
2	6.9	4.7	11.7	10.7	40.8	18.7	39.9	40.4
3	9.9	6.2	9.7	4.0	43.4	19.9	39.9	20.3
4	5.6	7.8	5.2	6.0	19.8	39.2	20.5	20.0
5	3.0	8.9	9.3	8.7	20.7	41.2	38.6	39.8



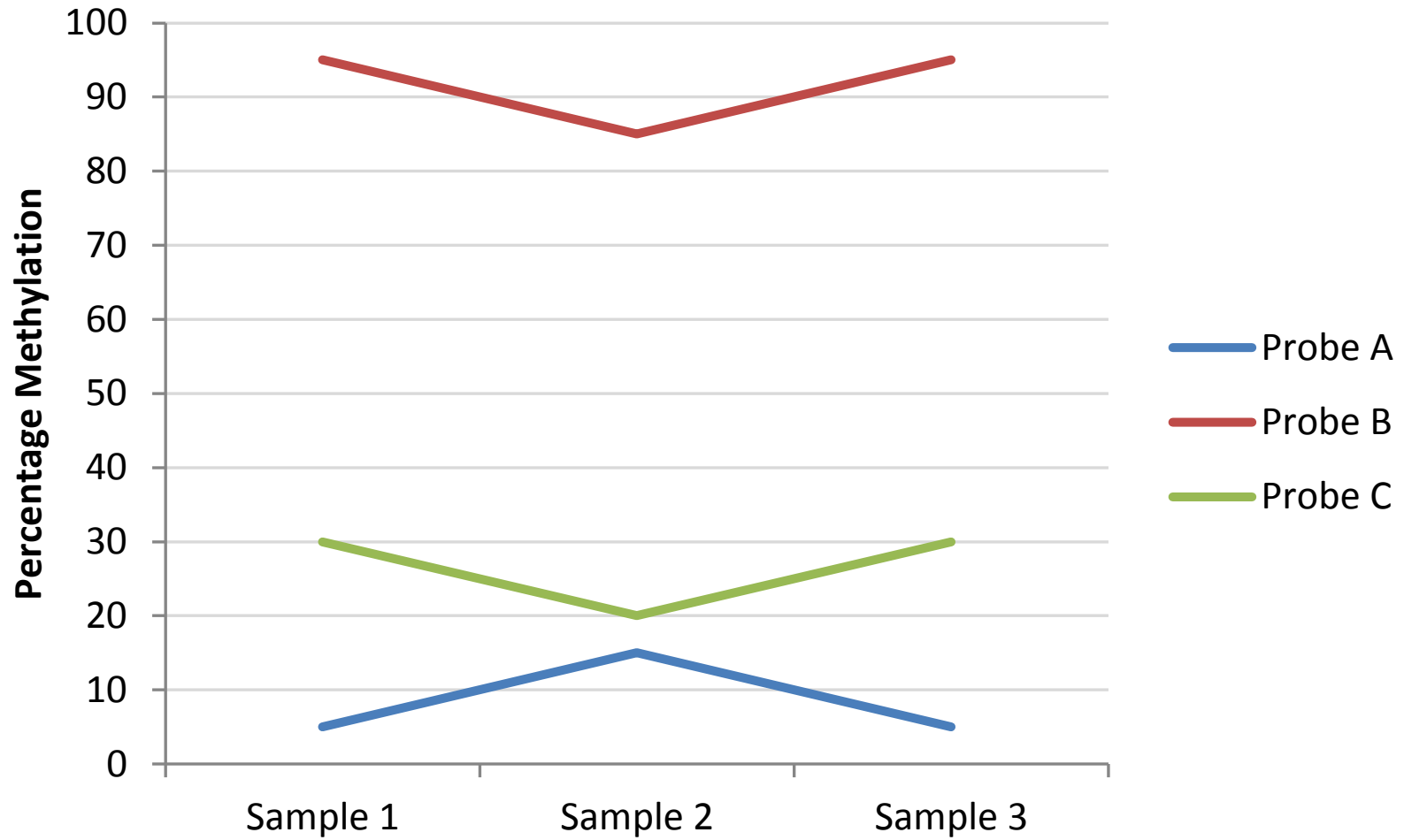
Plot types – Relationships

Heatmaps



- Colour scheme for grouping.
Clustering

Clustering Heatmaps



Plot types – Relationships

Heatmaps

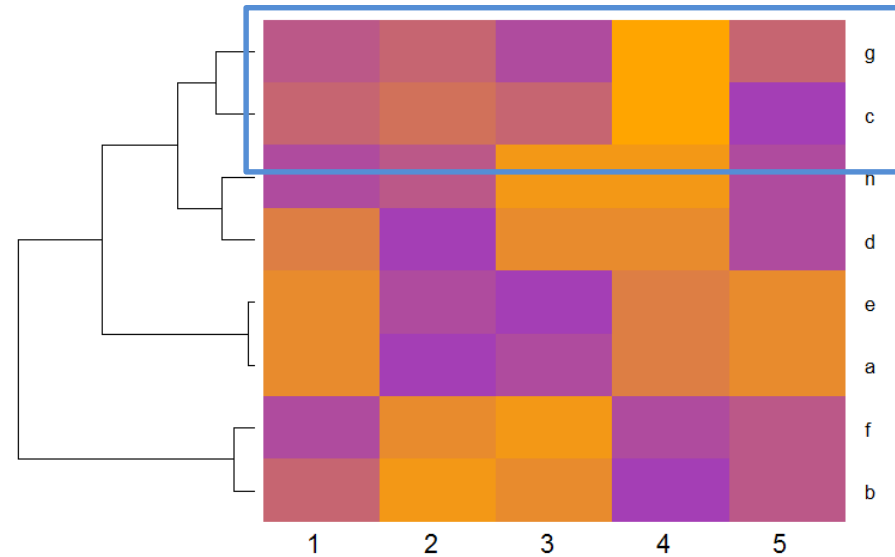
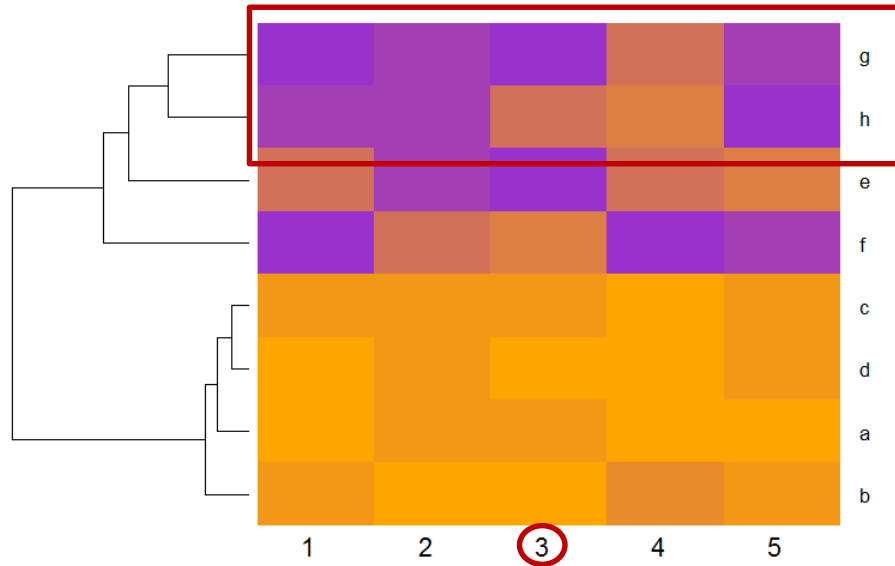
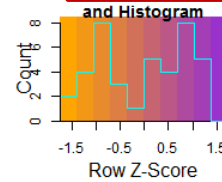
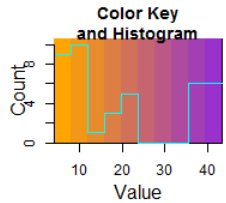
Euclidian distance:
differences
between values

Correlation:
consistency
of pattern

	a	b	c	d	e	f	g	h
1	3.7	10.8	8.4	10.4	19.5	40.6	39.0	40.1
2	6.9	4.7	11.7	10.7	40.8	18.7	39.9	40.4
3	9.9	6.2	9.7	4.0	43.4	19.9	39.9	20.3
4	5.6	7.8	5.2	6.0	19.8	39.2	20.5	20.0
5	3.0	8.9	9.3	8.7	20.7	41.2	38.6	39.8

Pattern Difference

0 0 =1.1
 ↑↑ =0.5
 ↓ - =19.6
 ↓↓ =0.5
 ↑↑ =1.2



- Heatmaps are great but plot data that are changing.

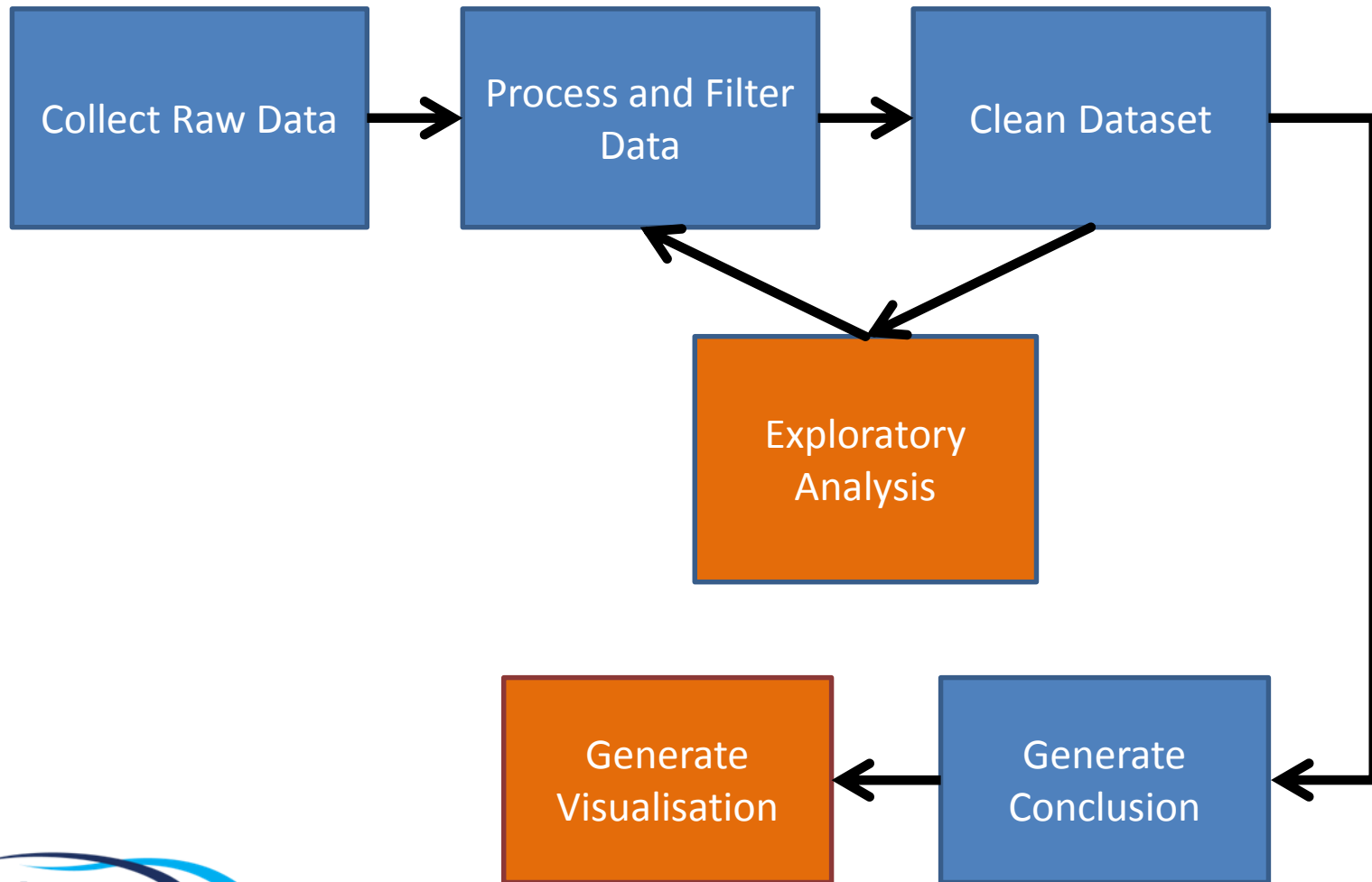
Ethics of data representation

v2017-06

Simon Andrews, Anne Segonds-Pichon

simon.andrews@babraham.ac.uk
anne.segonds-pichon@babraham.ac.uk

Data Visualisation Process



What is Ethics

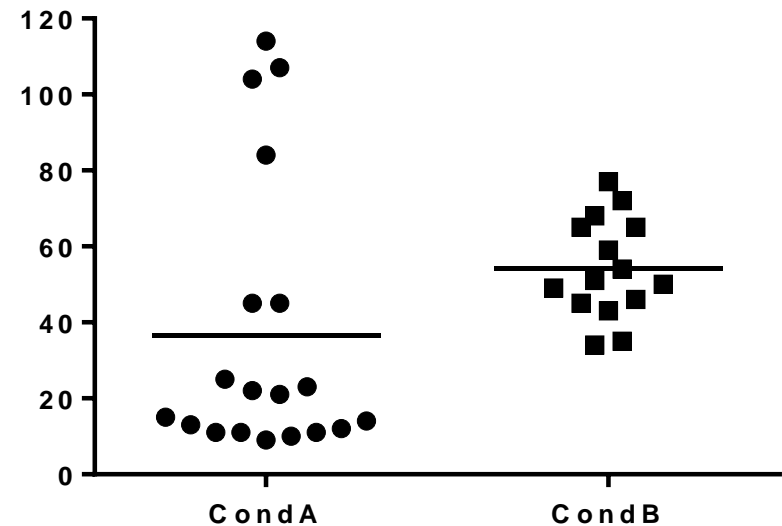
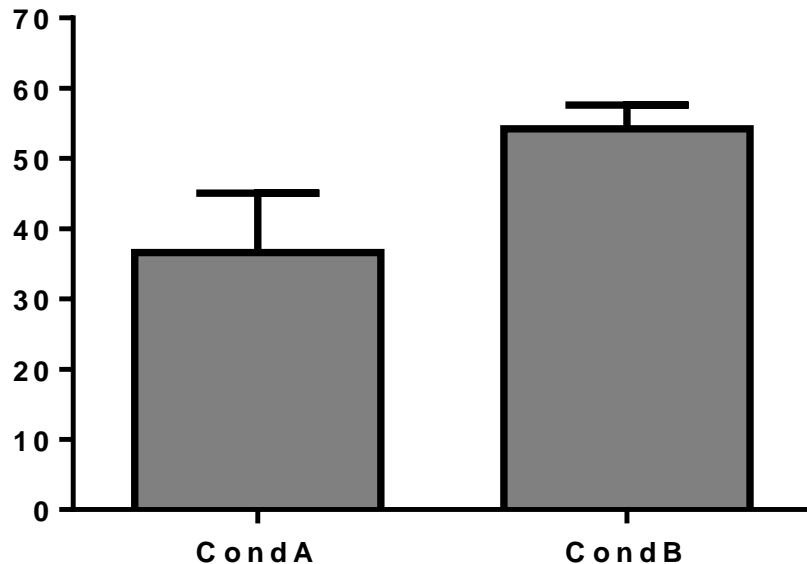
when it comes to data visualisation?

- 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

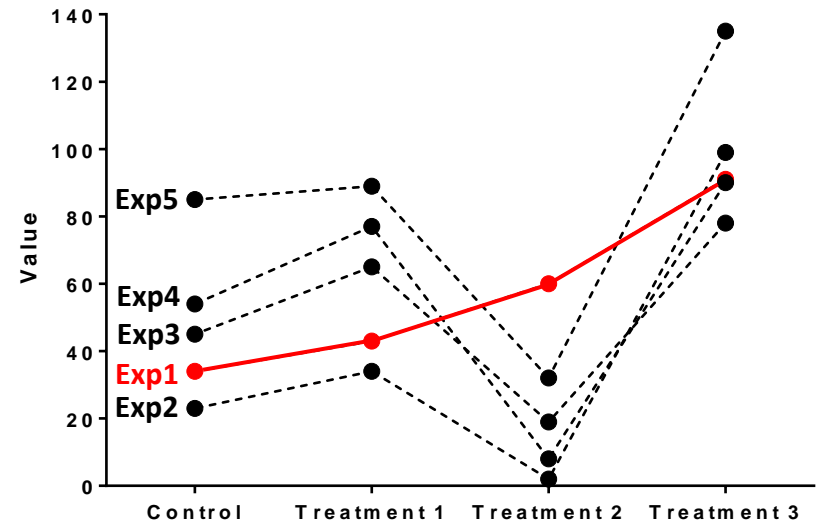
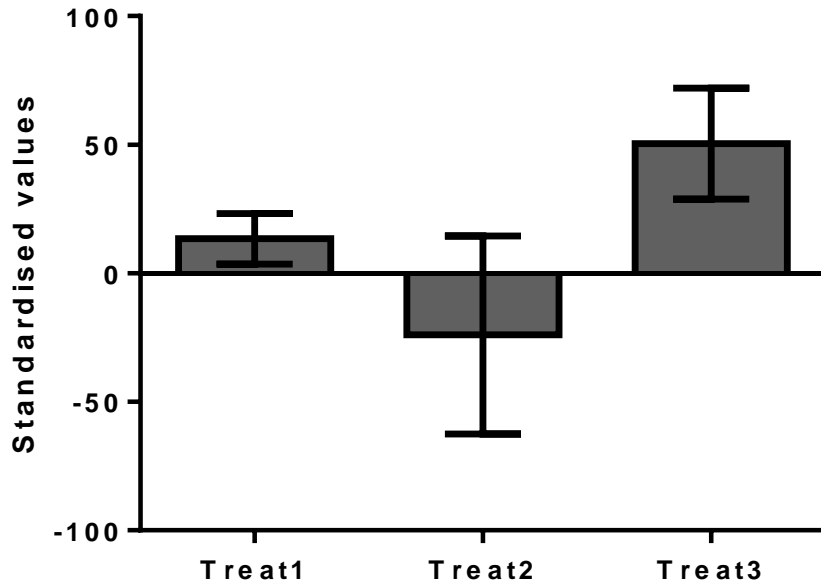
- One experiment: change in the variable of interest between CondA to CondB.
 - ❖ Data plotted as a **bar chart**.



Not exploring/getting to know the data well enough

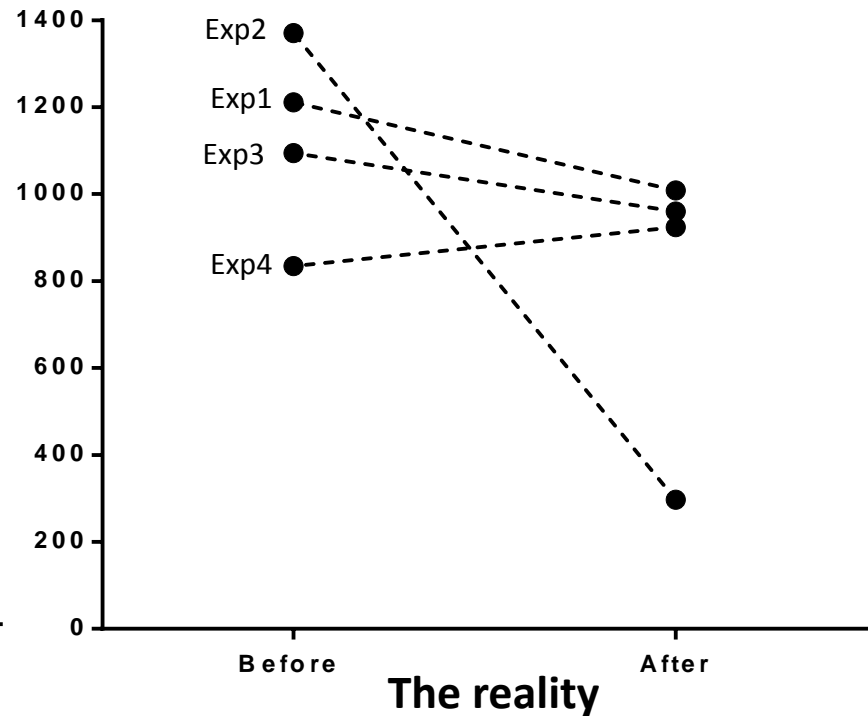
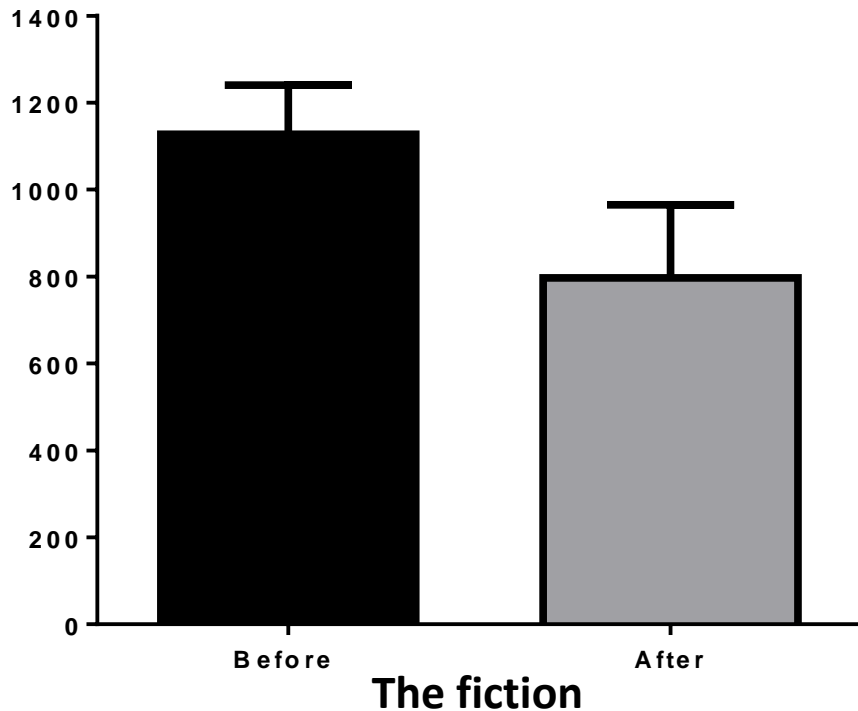
Example 2

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



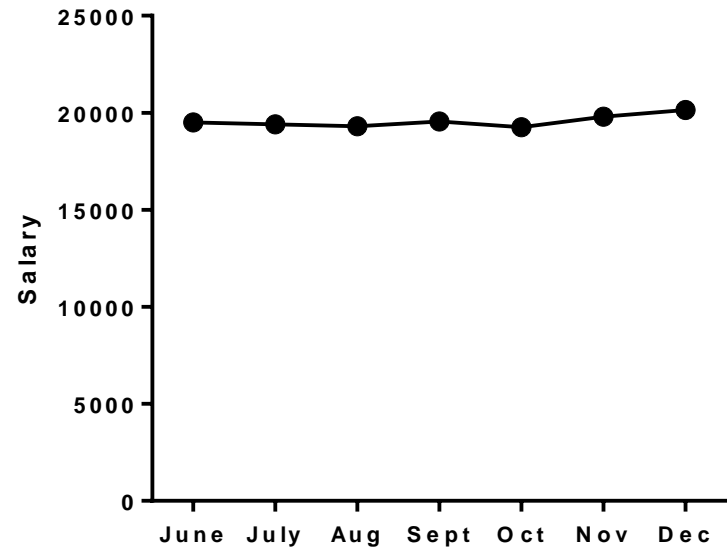
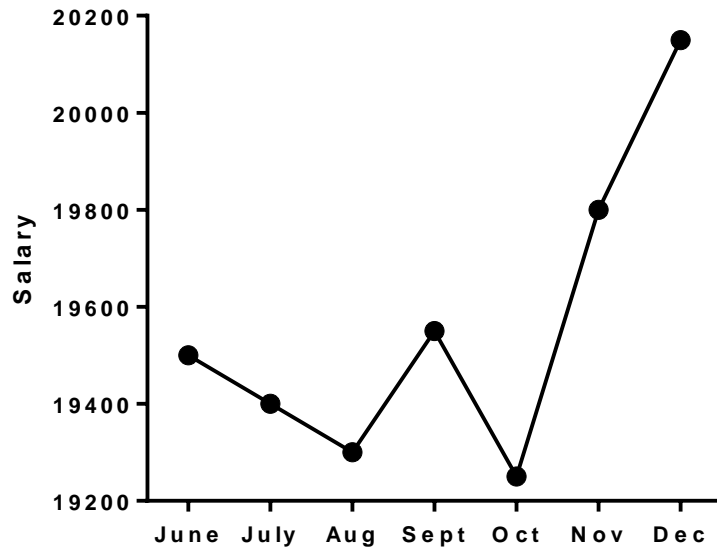
Choosing the wrong graph to present the data

- 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**.



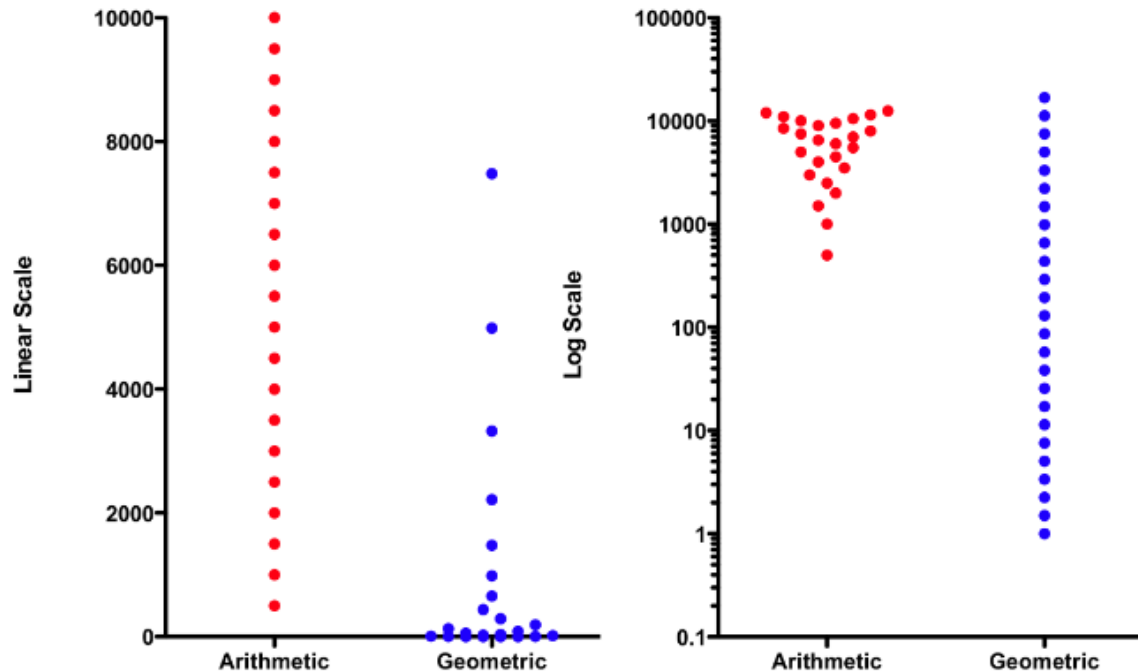
Choosing the wrong axis/scale

- Example: increase in salary in the last term.



Choosing the y-axis/scale

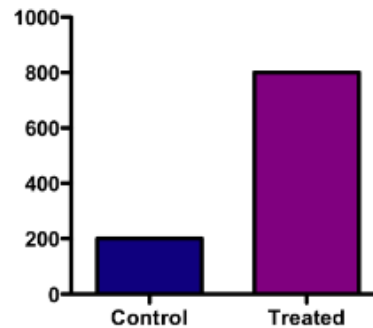
- Be careful with Linear vs. logarithmic scale.



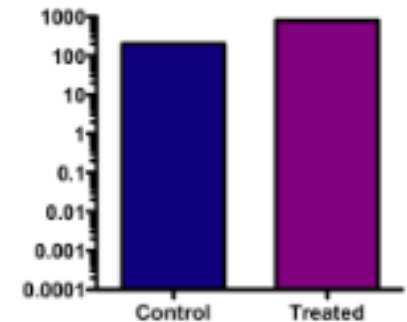
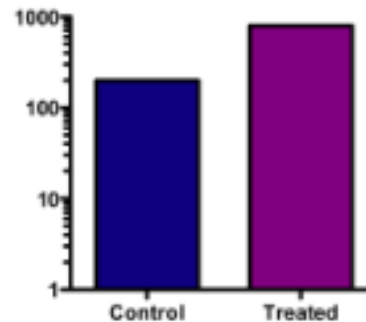
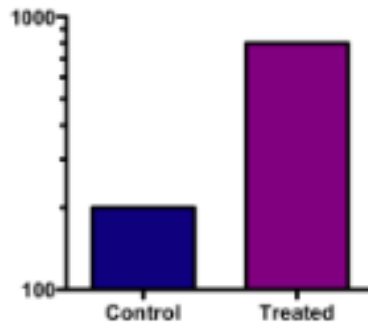
Choosing the y-axis/scale

- 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

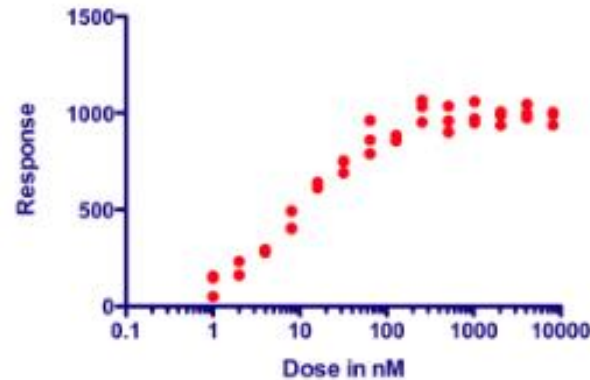
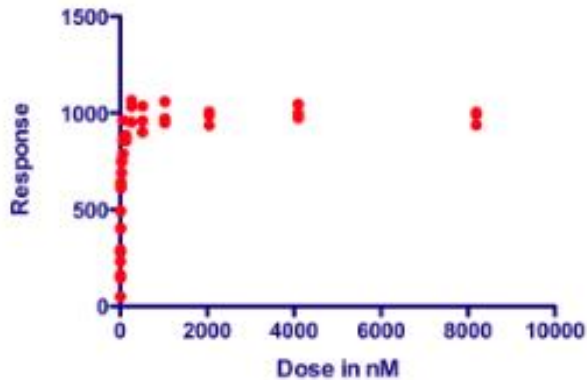


Logarithmic scale

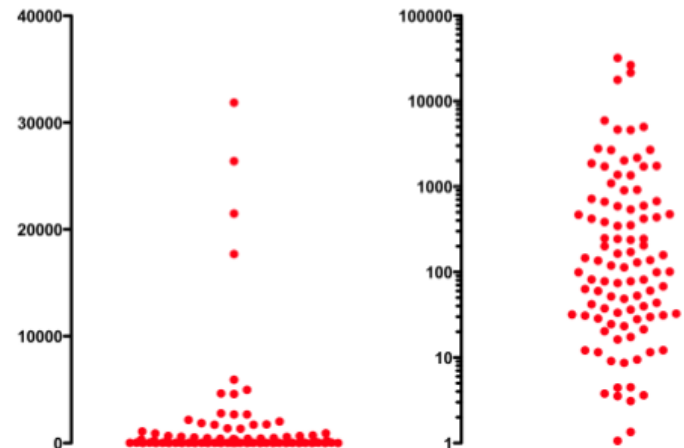


Choosing the y-axis/scale

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



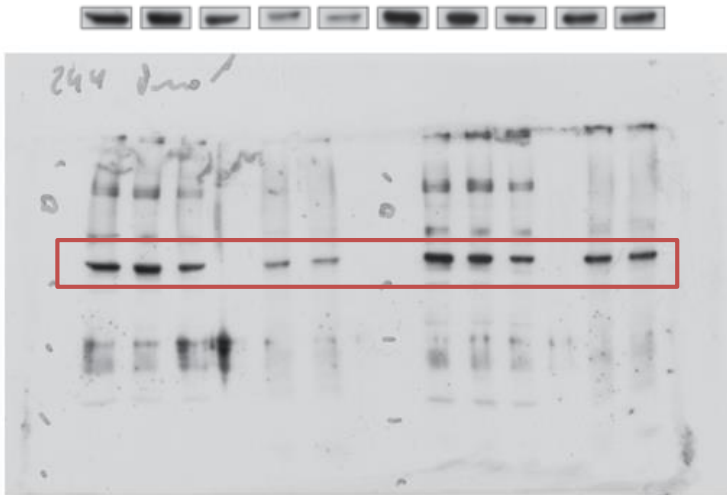
Logarithmically spaced values



Simply Cheating:

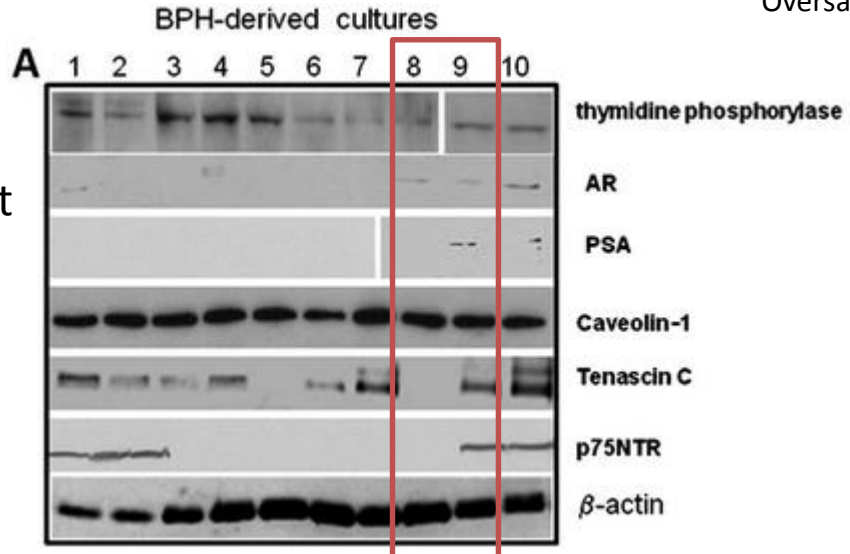
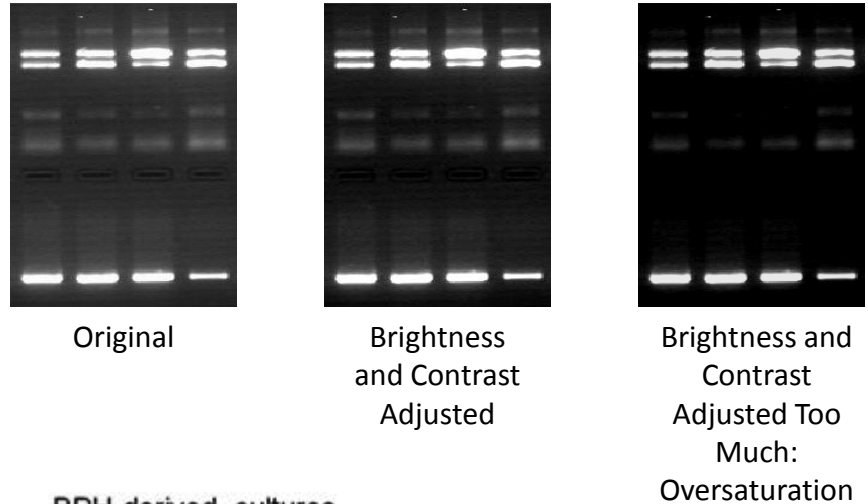
Manipulating images: Western blot

- 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?