

**Cancer Research UK**

# **Designing more effective scientific figures**

(II)

Aiora Zabala

PhD Environment. VTP Graphic Design  
az296, [aiora.zabala@gmail.com](mailto:aiora.zabala@gmail.com)

# Structure of this course

## THEORY

- Morning 1
- Why figure design?
  - Principles of figure design
  - Elements of a figure
  - Colour & ethics

## PRACTICAL

- 2
- Gimp** – *bitmap* (e.g. jpg)
  - Setting up a canvas
  - Layers and importing files
  - Editing colour
  - Export formats and qualities

- Afternoon 3
- Dealing with complexity
  - Choosing the right figure
  - Typography
  - Composition & layout

- 4
- Inkscape** – *vectorial* (e.g. pdf)
  - Document properties
  - Create & manipulate objects
  - Composition
  - Import & export for publication

# Key ingredients: the tools

## Elements: marks and channels

- Data
- Points, lines, areas
- Colour
- Typography

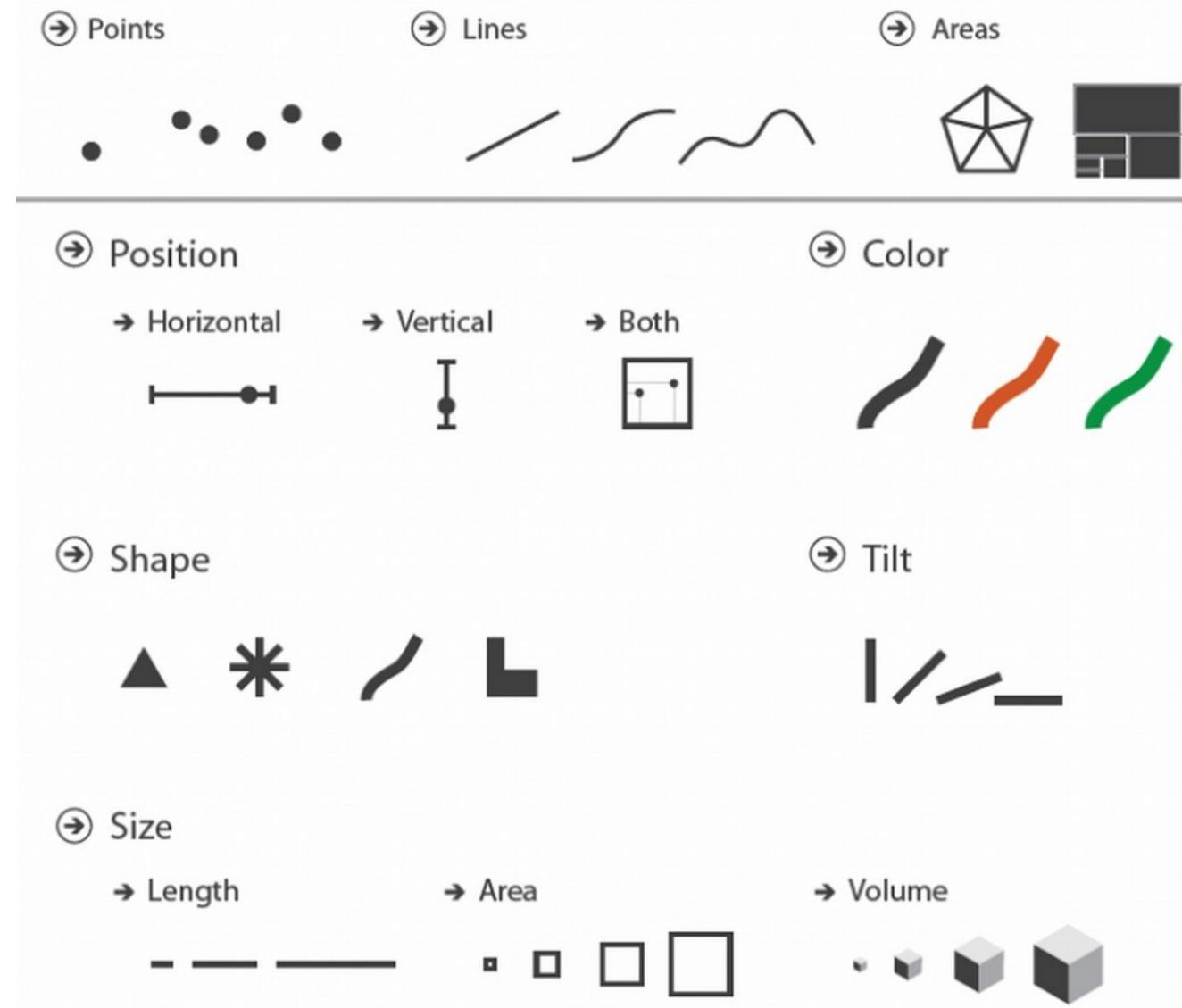
## Composition

- Grid and alignments
- Balance
- Hierarchy and focus

# Elements: Marks and channels

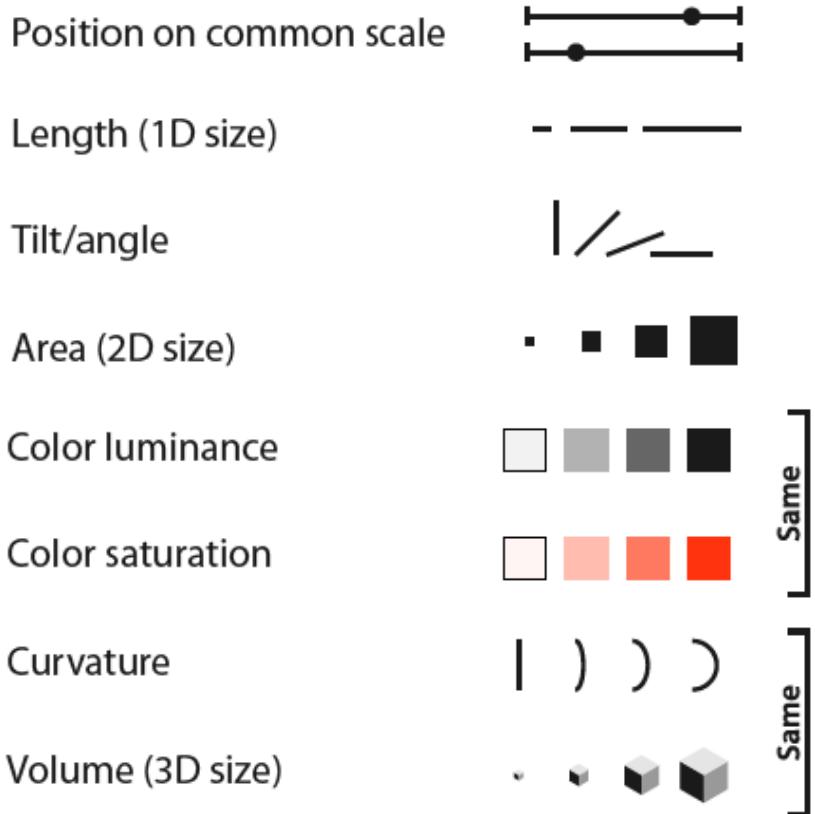
**Marks** (geometric primitives): used to **represent** data

**Channels** control the graphical appearance of marks: used to **encode** data, can be combined



# Types of channel

Identity channels: categorical/  
**qualitative** attributes

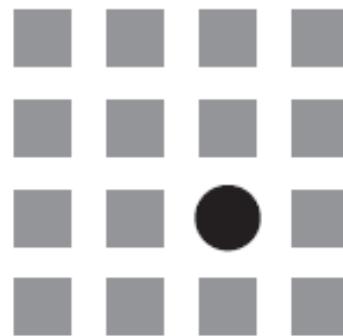


Magnitude channels: ordered/  
**quantitative** attributes



# Types of channel (continued)

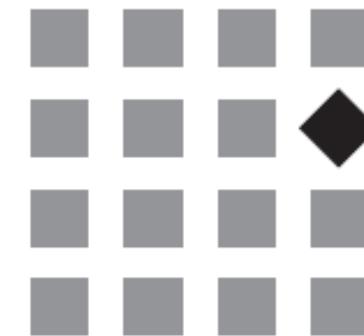
SHAPE



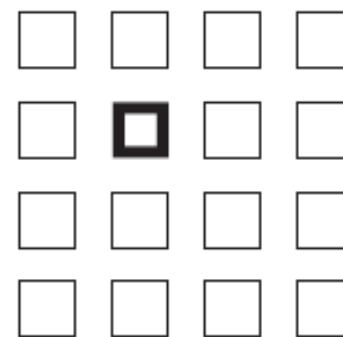
SIZE



ORIENTATION



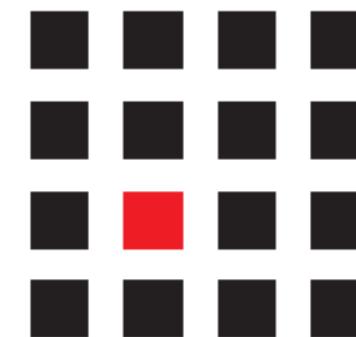
WEIGHT



POSITION



COLOR

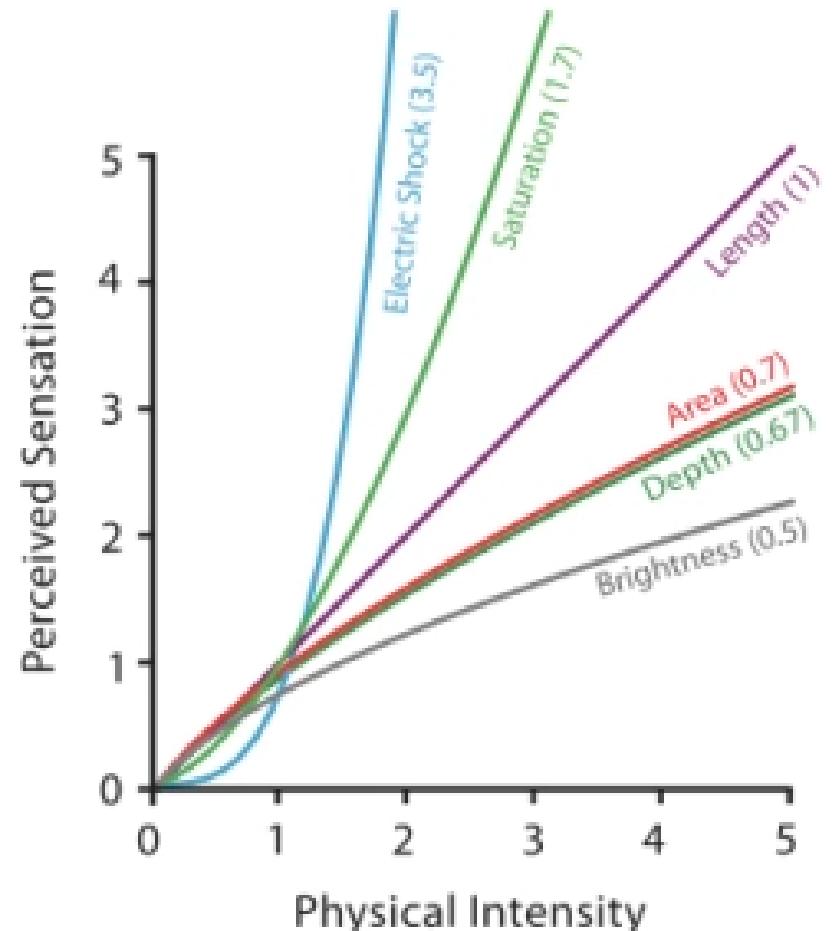


# Effectiveness of each channel: Quantitation perception

The perceived magnitude of sensory channels follows a power law:  $S = I^N$

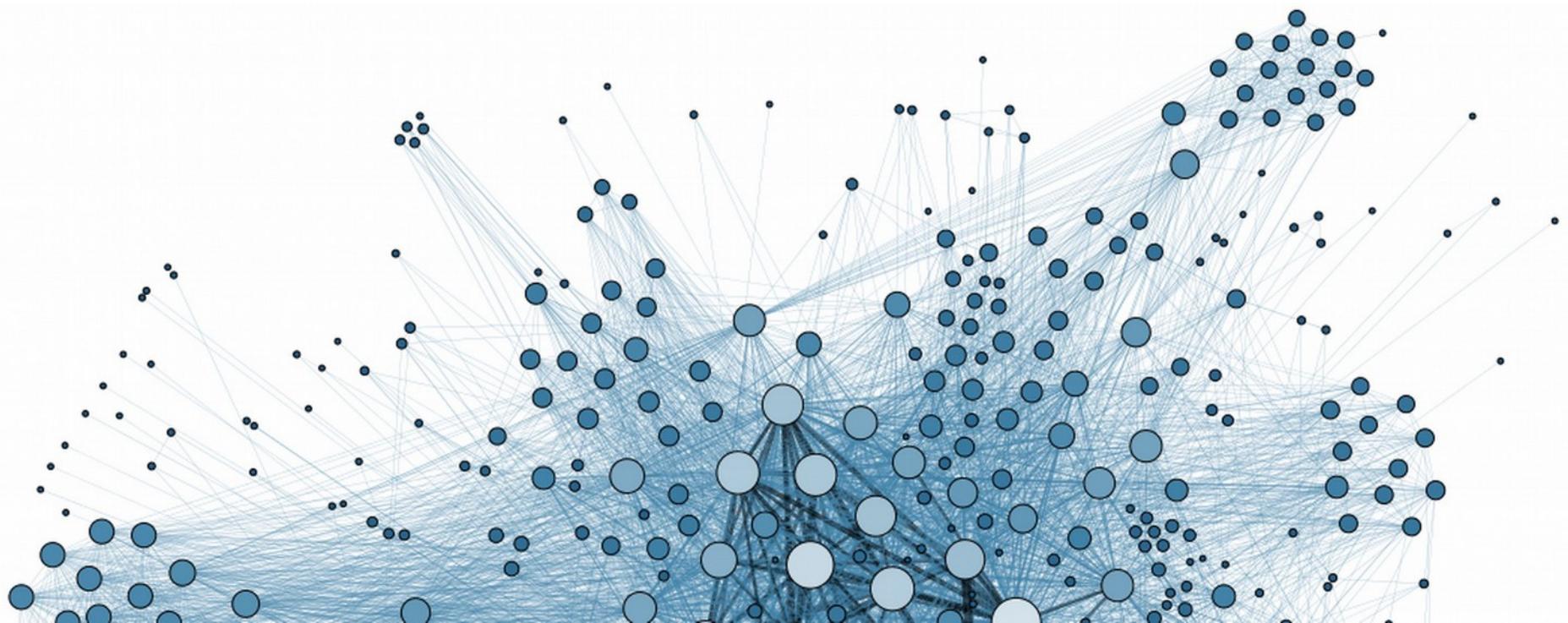
Depending on the **N** of a given type of sensation, its perception is magnified (*e.g. colour saturation*) or compressed (*e.g. brightness*)

Steven's Psychophysical Power Law:  $S = I^N$

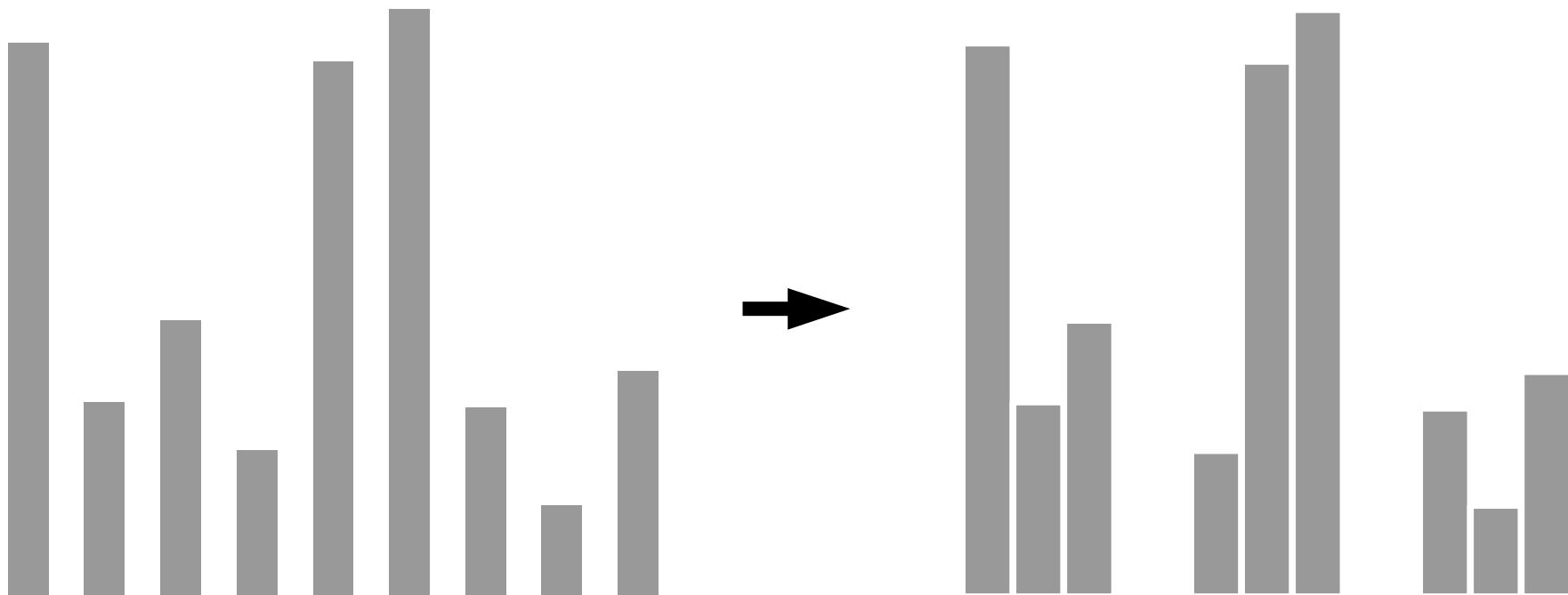


# Dealing with complexity

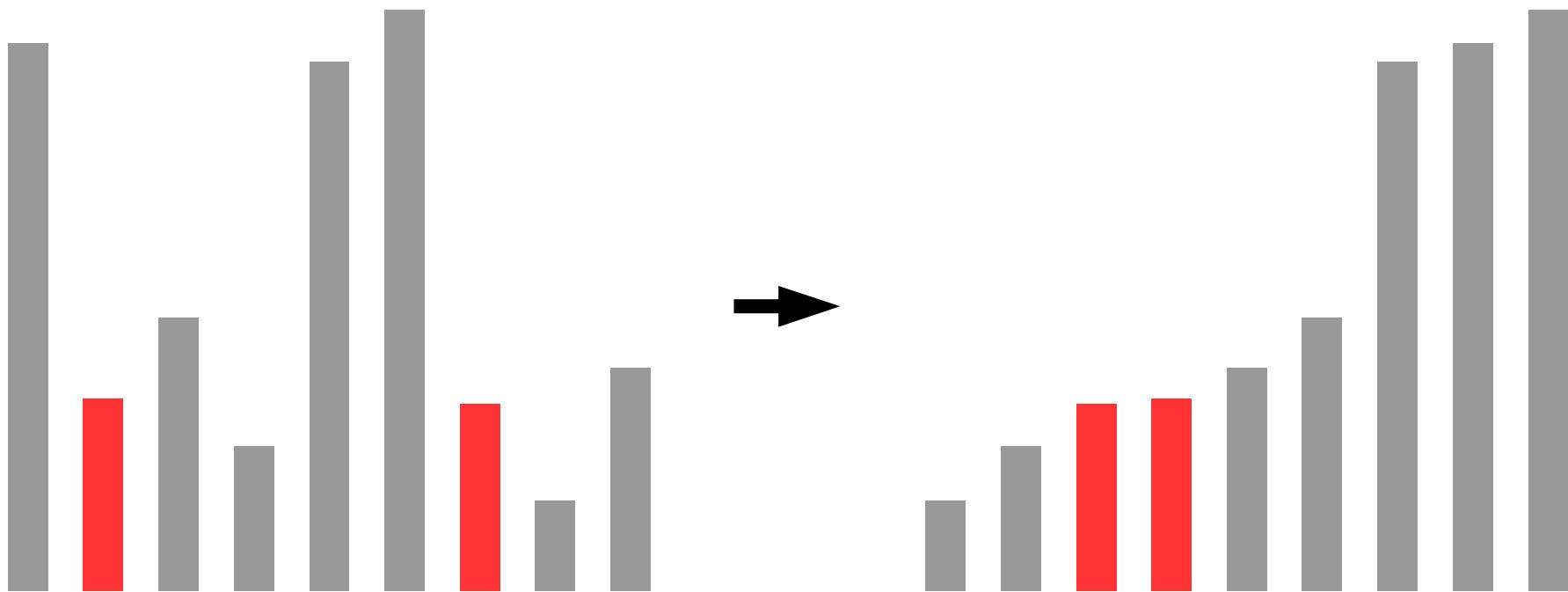
- To focus the viewer's attention onto the main point you want to convey (e.g. on specific subsets of data)
- To require less cognitive load for the viewer to understand the message



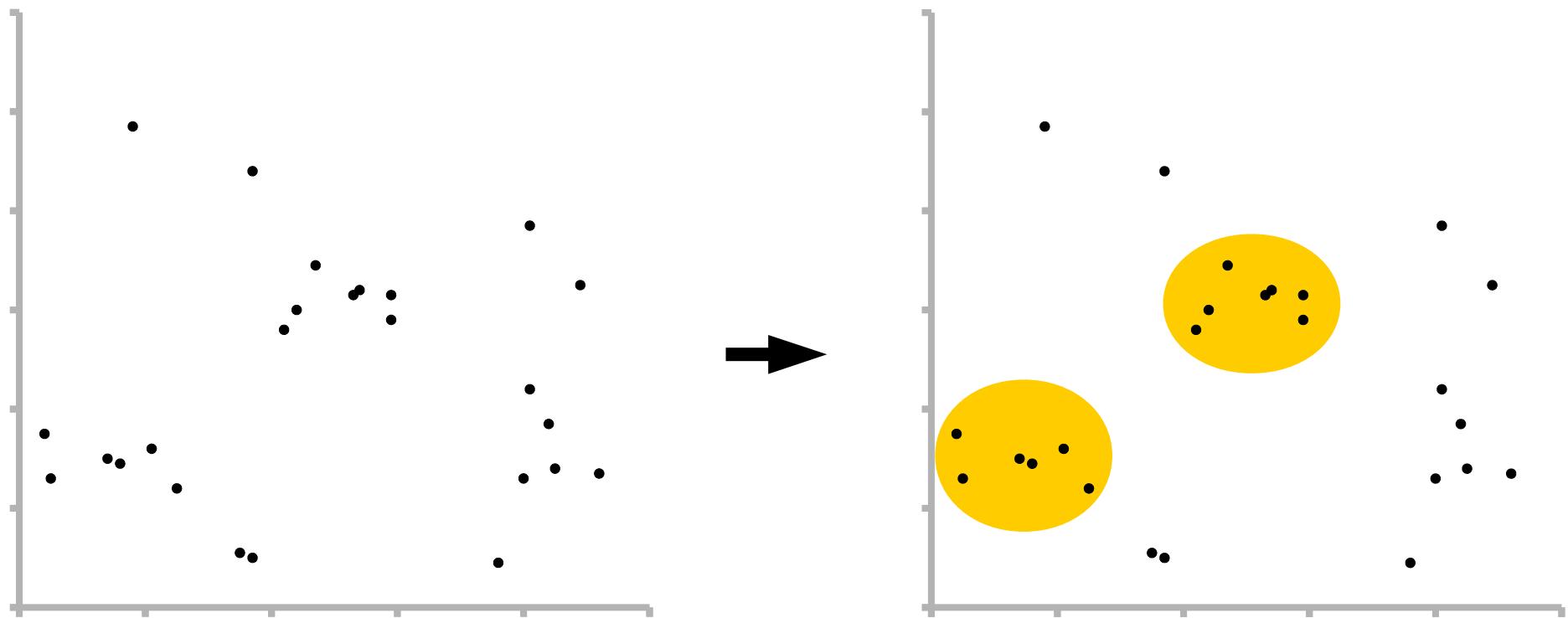
# Grouping



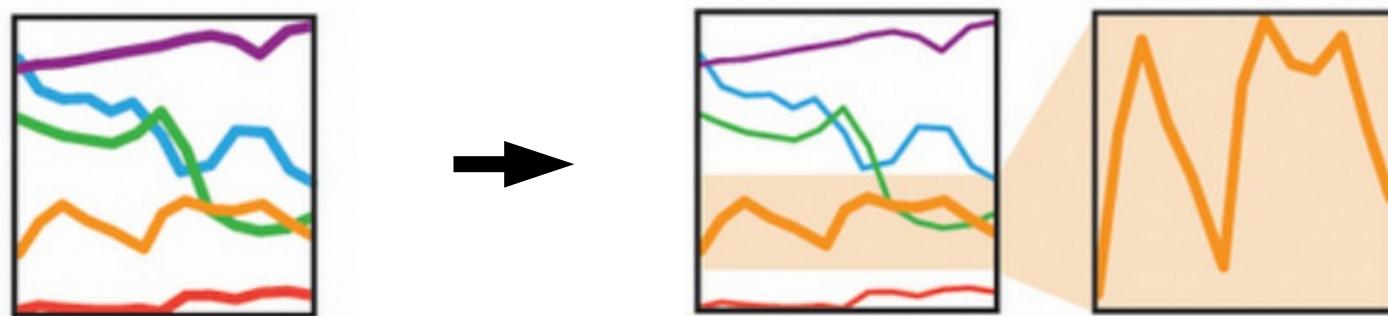
# Ordering (only for categories)



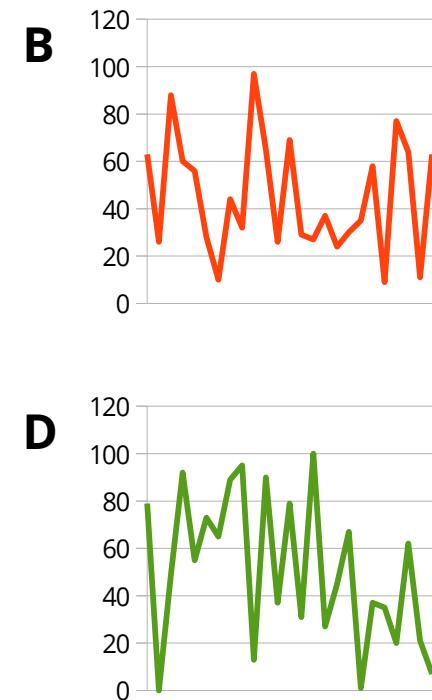
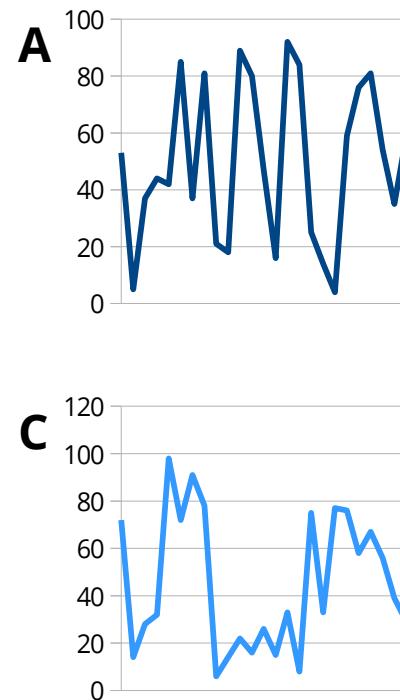
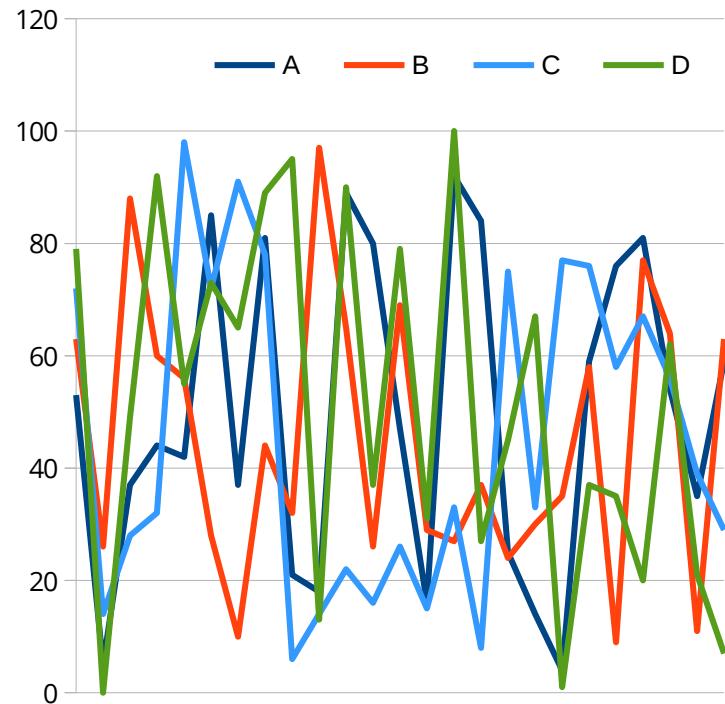
# Containment



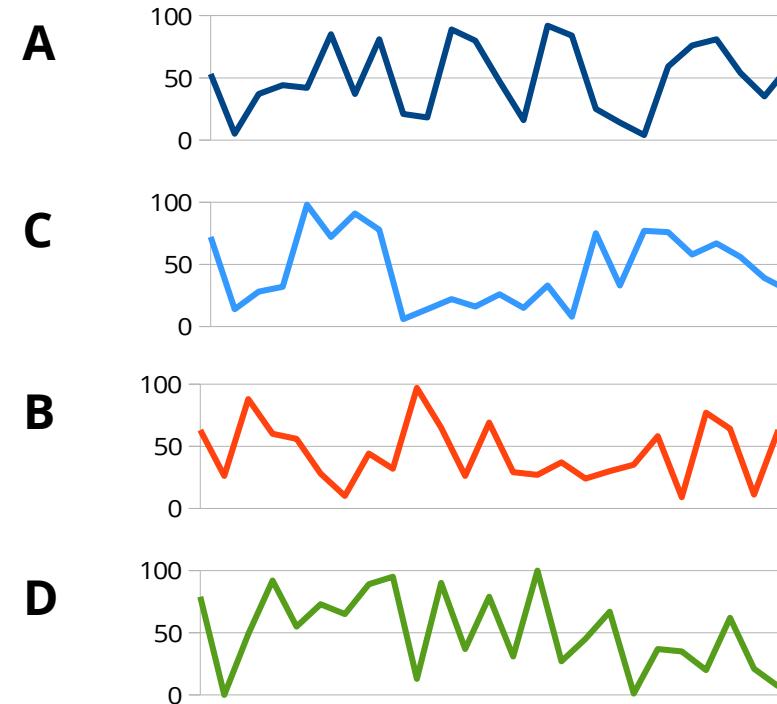
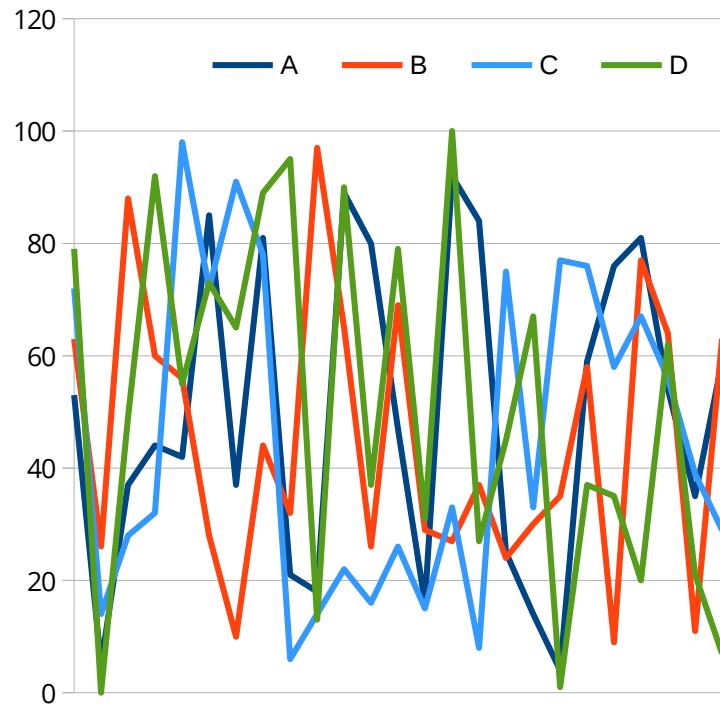
# Filter, link, embed



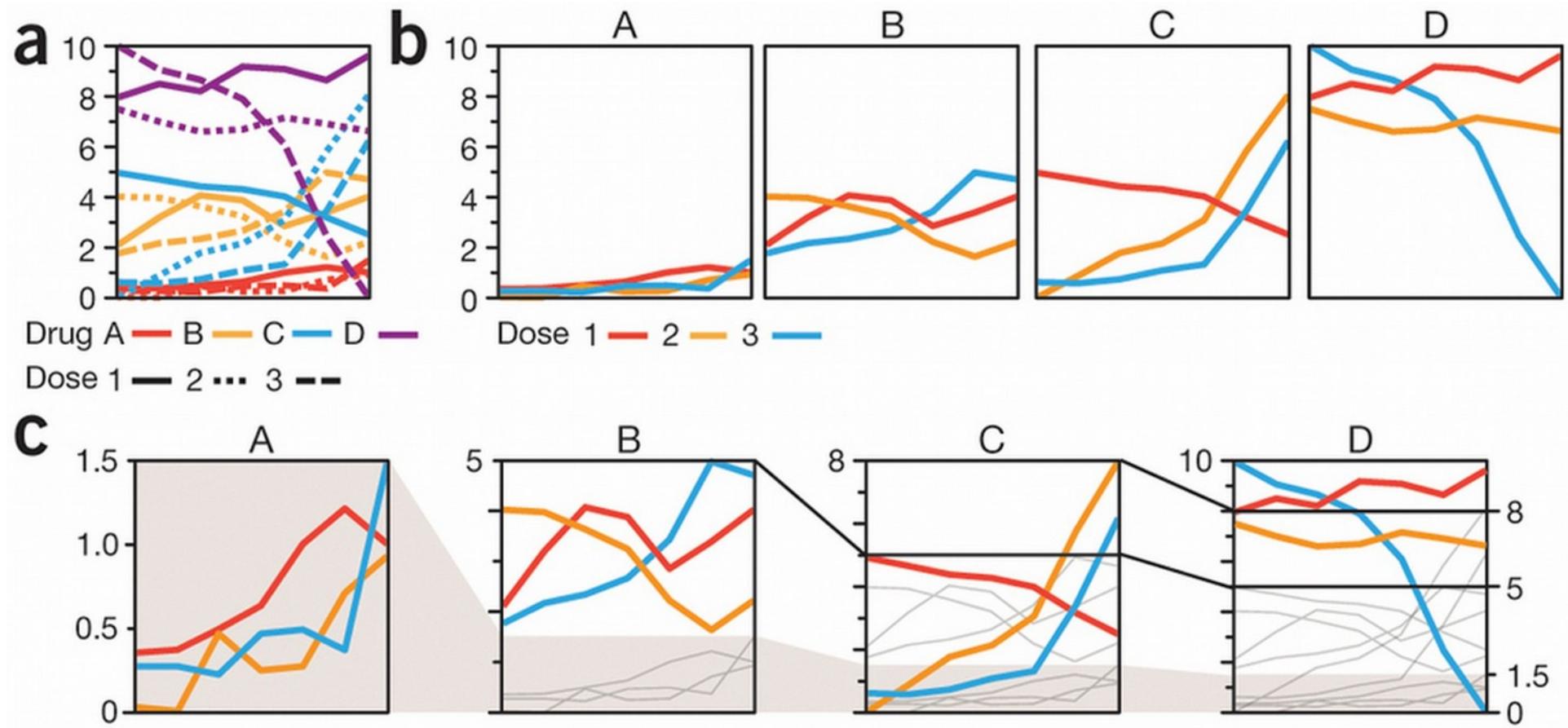
# Small multiples



# Small multiples



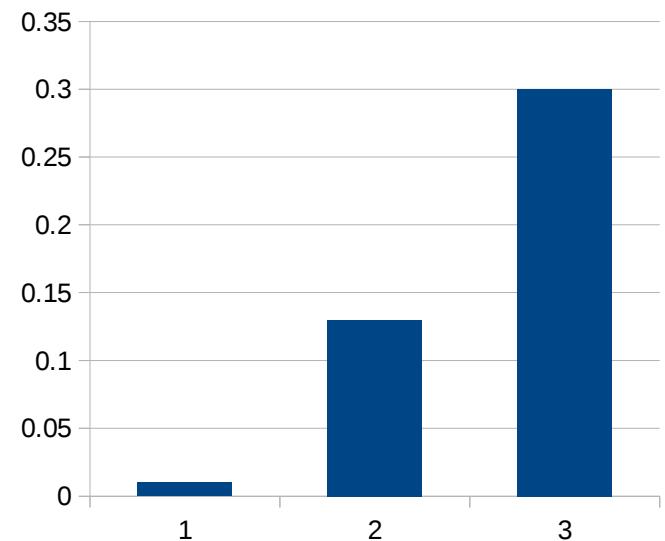
# Small multiples



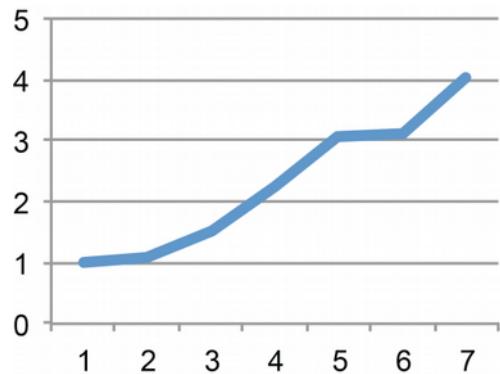
# Choosing the type of figure

- Text, table or figure?
  - Text: one or two numbers
  - Table:
    - Exact numerical values
    - Small datasets (a figure may be best avoided if it has low data density)
    - When the data presentation requires many localised comparisons

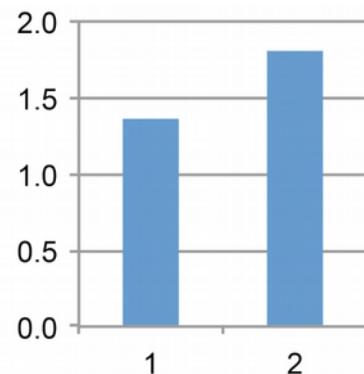
Treatment 1	0.01
Treatment 2	0.13
Treatment 3	0.30



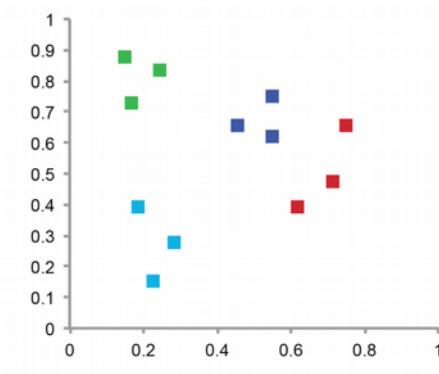
# Things you can illustrate



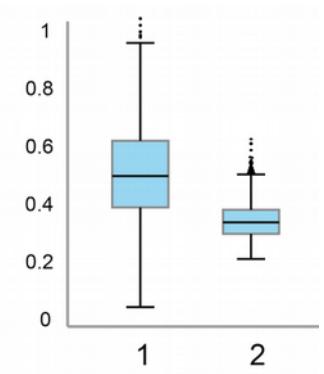
Relationship



Comparison

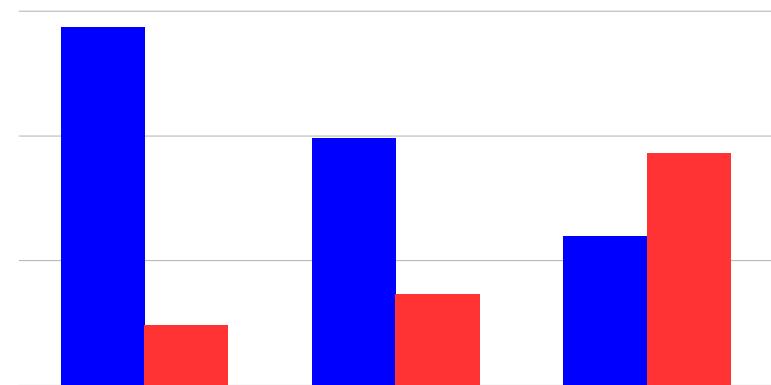
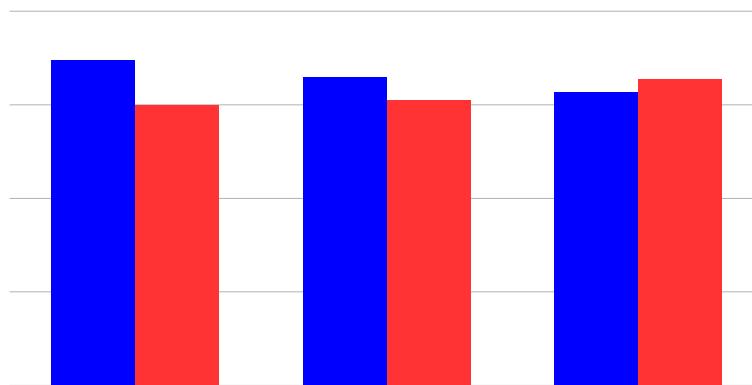
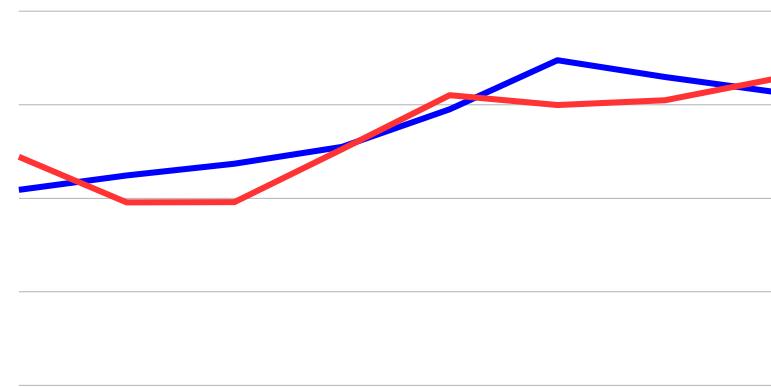
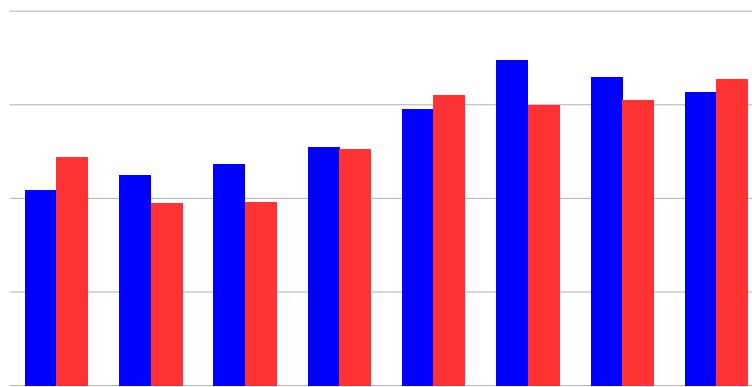


Composition

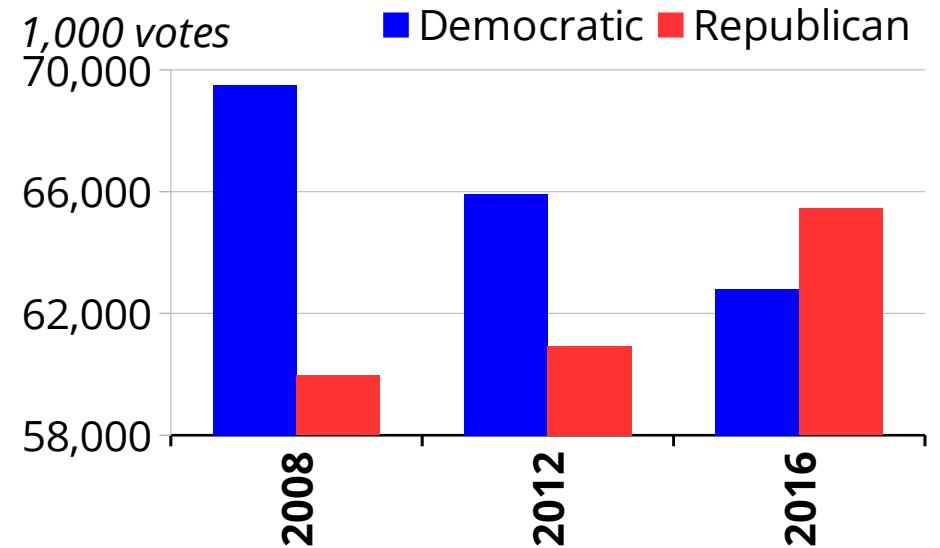
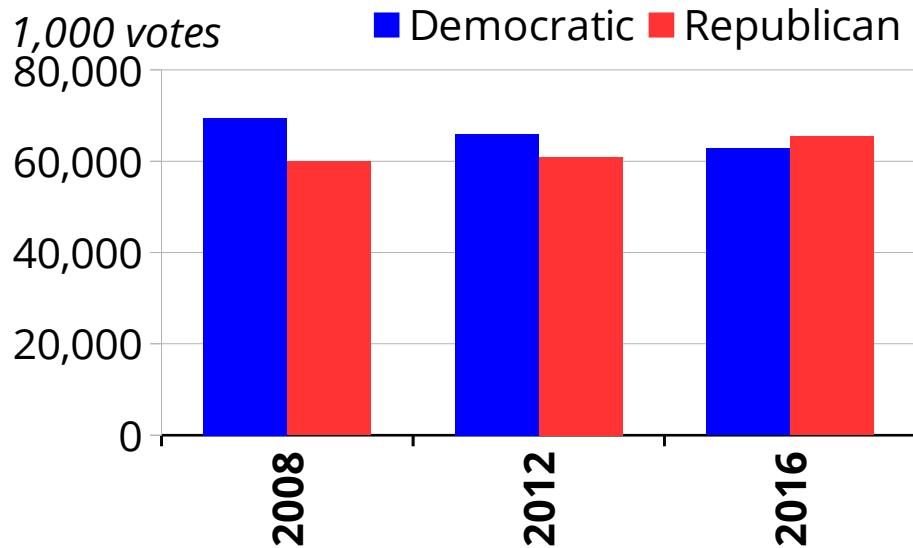
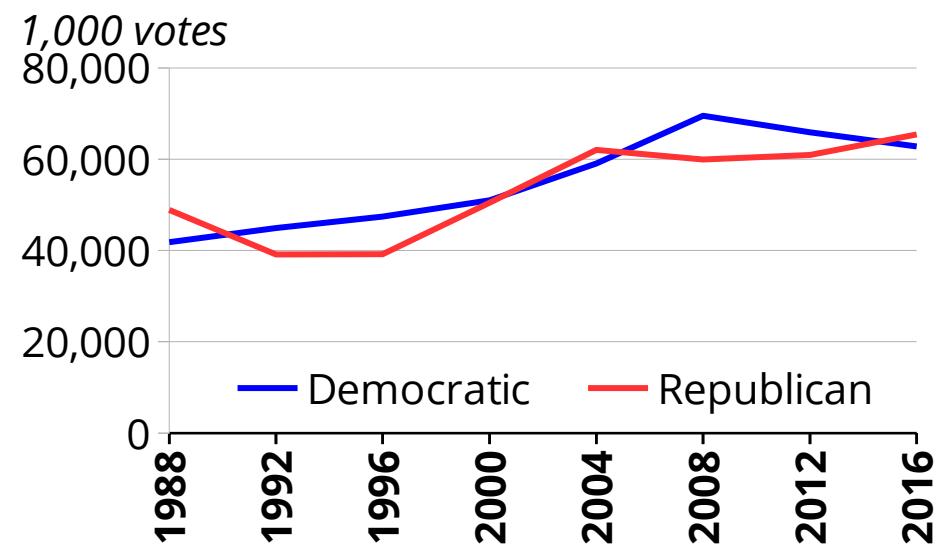
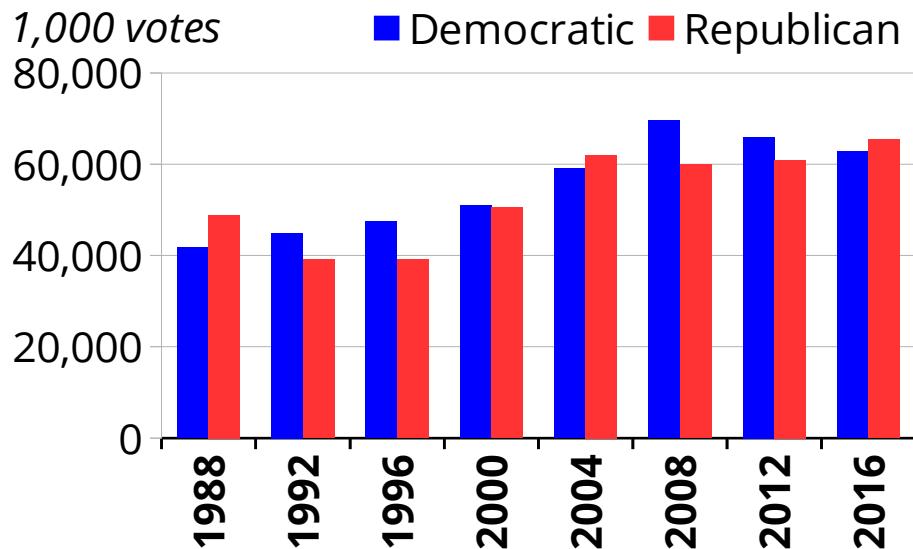


Distribution

# Each figure tells a different story

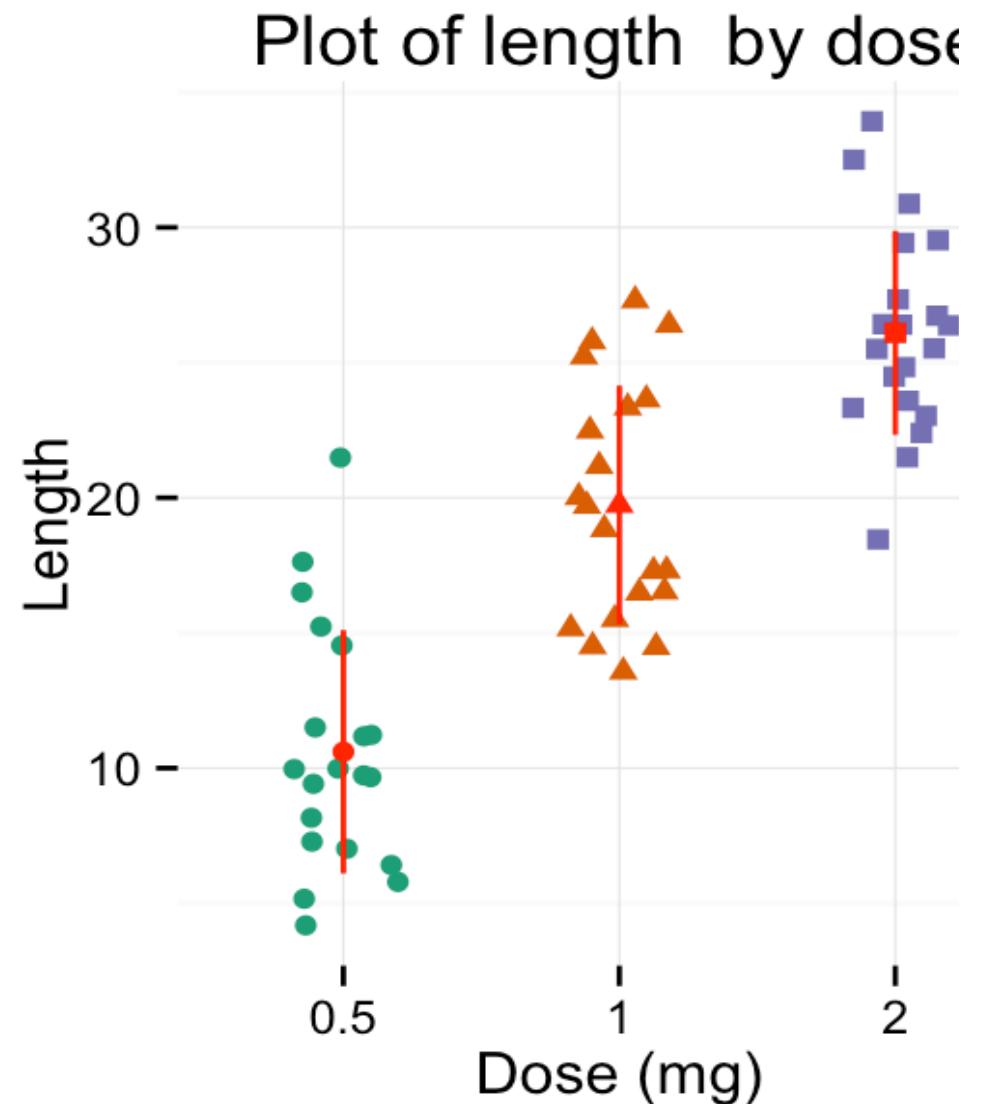


# Each figure tells a story differently



# Stripchart – comparison

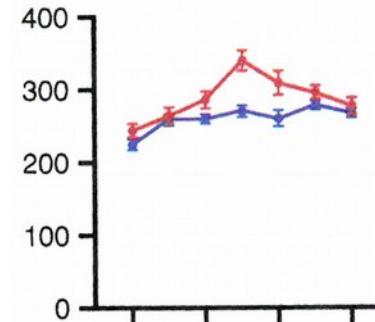
- Only one of the axis is meaningful
- To explore small datasets ( $n < 100$ ) and compare categories
- The most basic plot (rarely in publications)



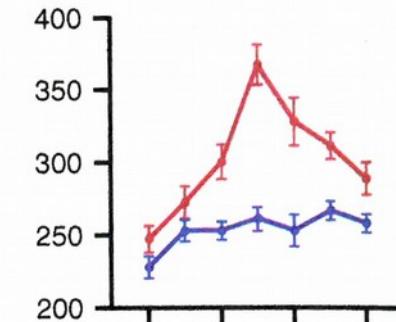
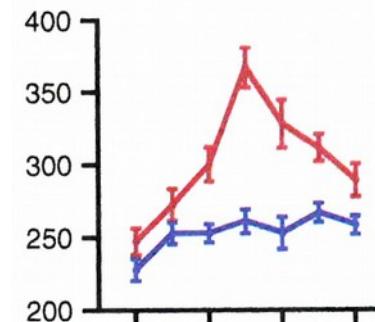
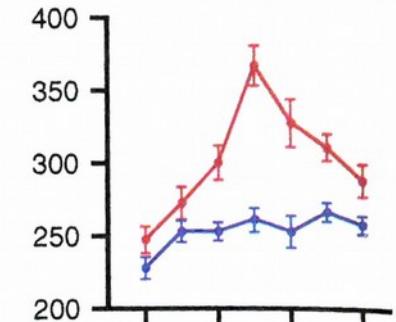
# Line chart – relationships

- To show a trend of **continuous** data (usually over time)
- For matched, paired or repeated data, and for time-series
- To tell a story: how data change, rather than the discrete values of the data

Before



After

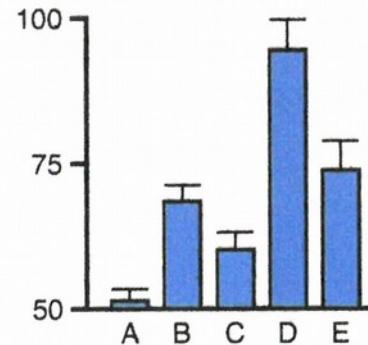


Carter 2013

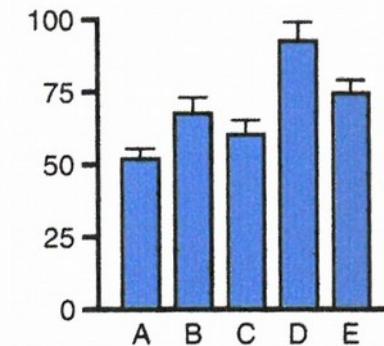
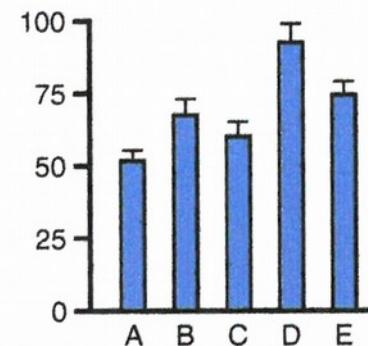
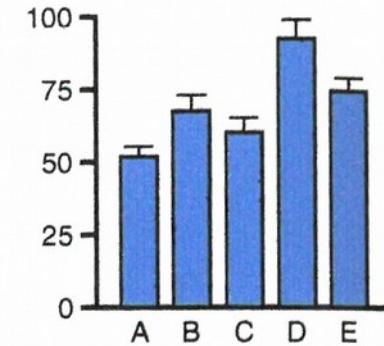
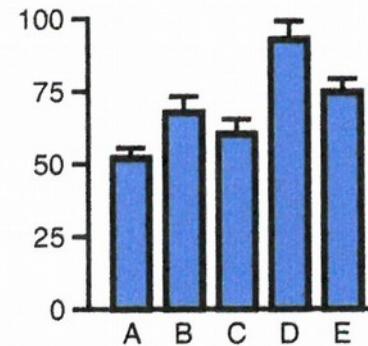
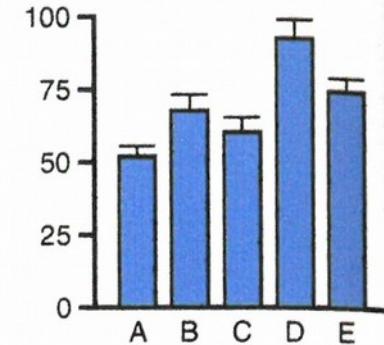
# Bar chart – comparison

- To **compare** discrete quantities of **non-continuous** data
- For presenting results and emphasise differences (not so much to explore)

Before

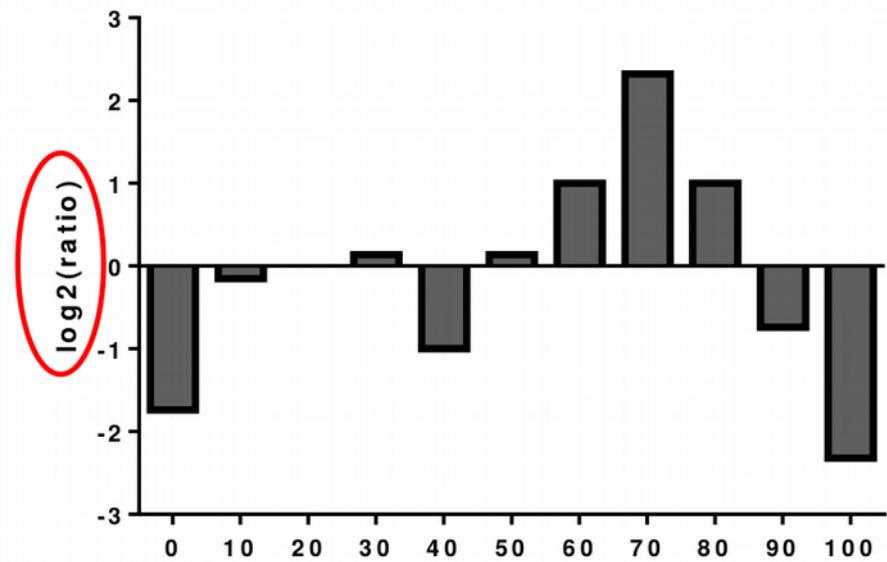
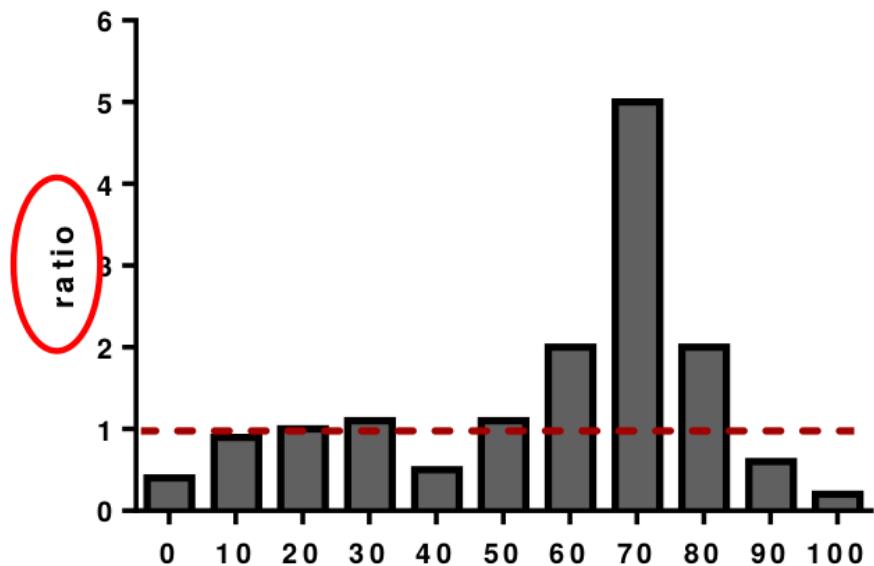


After



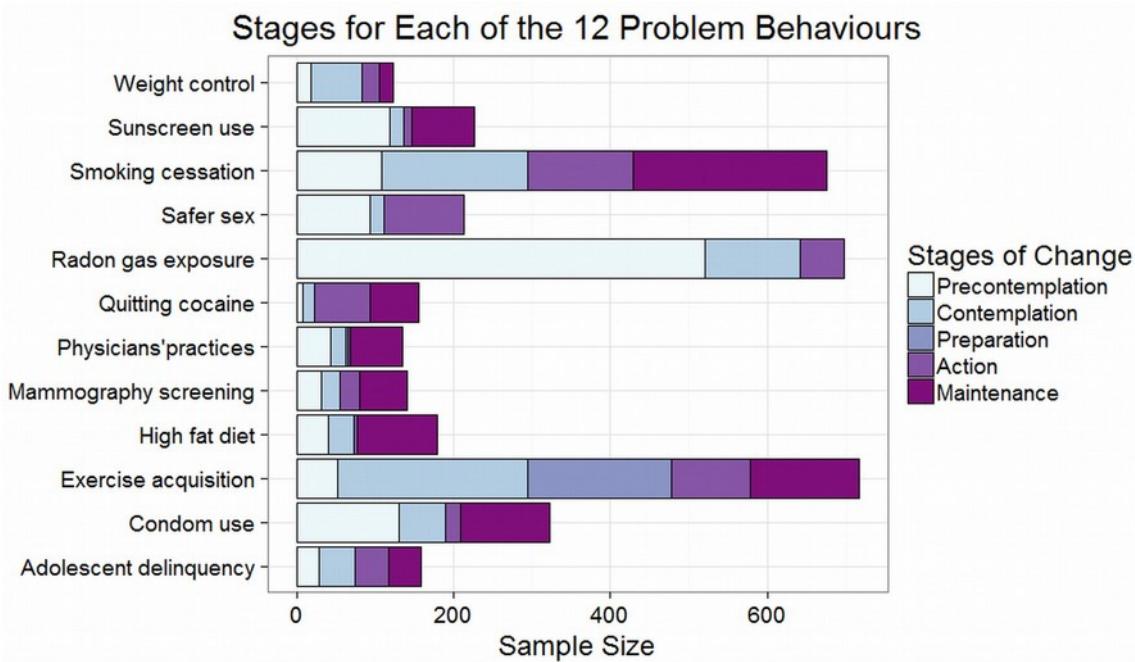
# Bar chart – comparison

The choice of the x axis and of point of reference can affect how comparisons are perceived

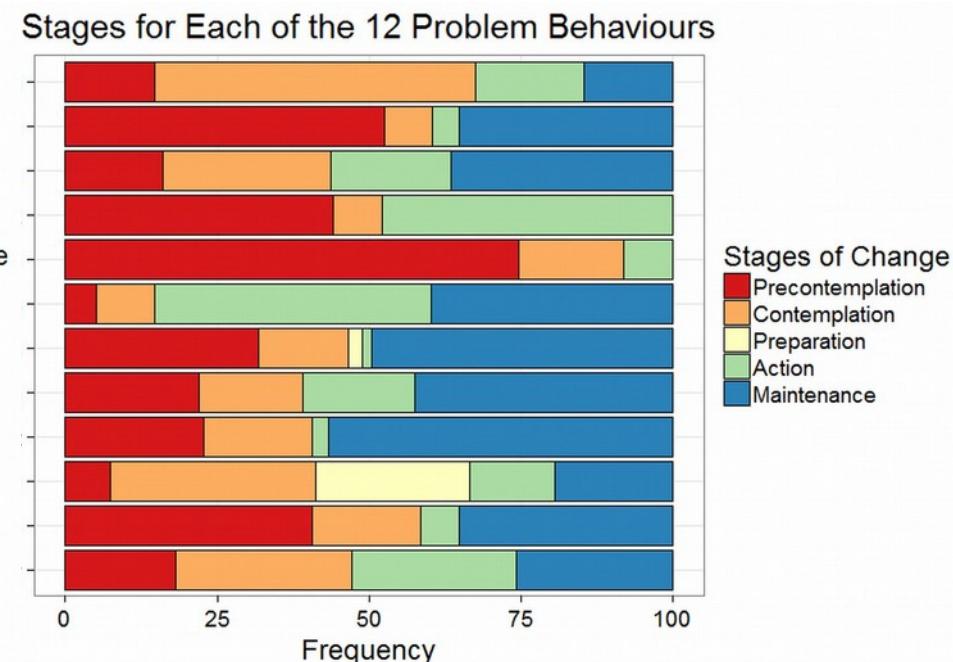


# Bar chart variations

## Stacked bar chart



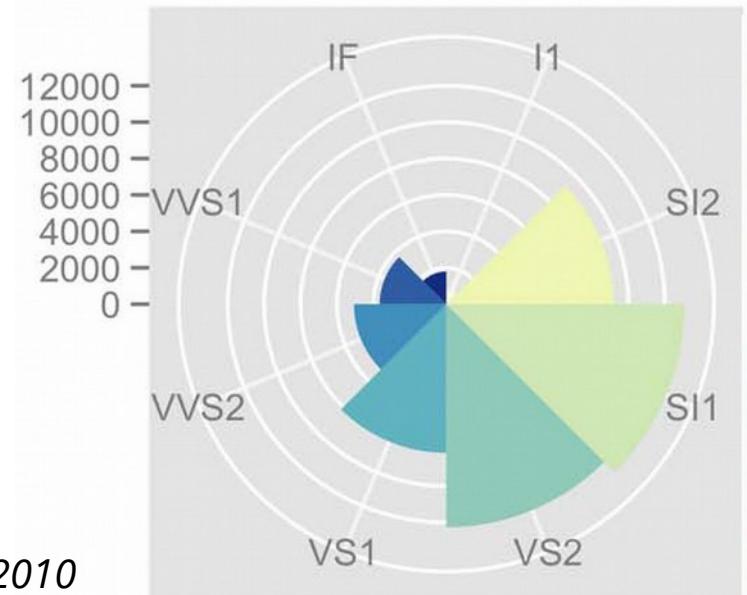
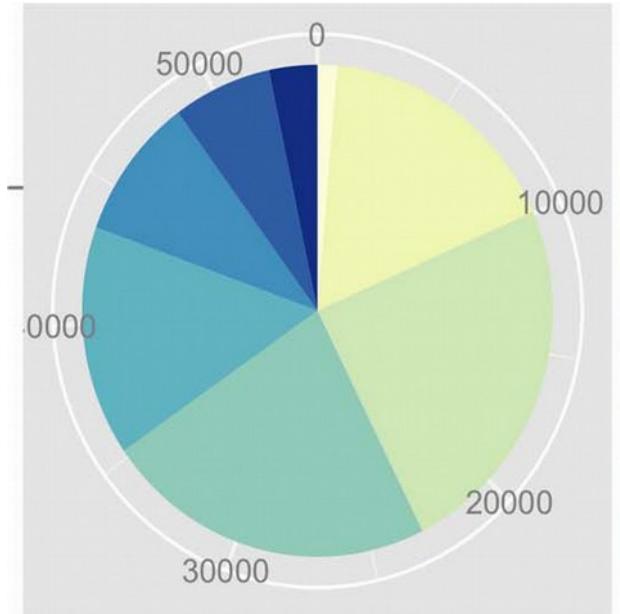
## Normalised stacked bar chart



- For categorical data; heed the sample size

# Pie chart – composition/ proportion

- To show relative proportions of a whole
- Not a great idea, ‘given their low data-density and failure to order numbers along a visual dimension’ (Tufte)



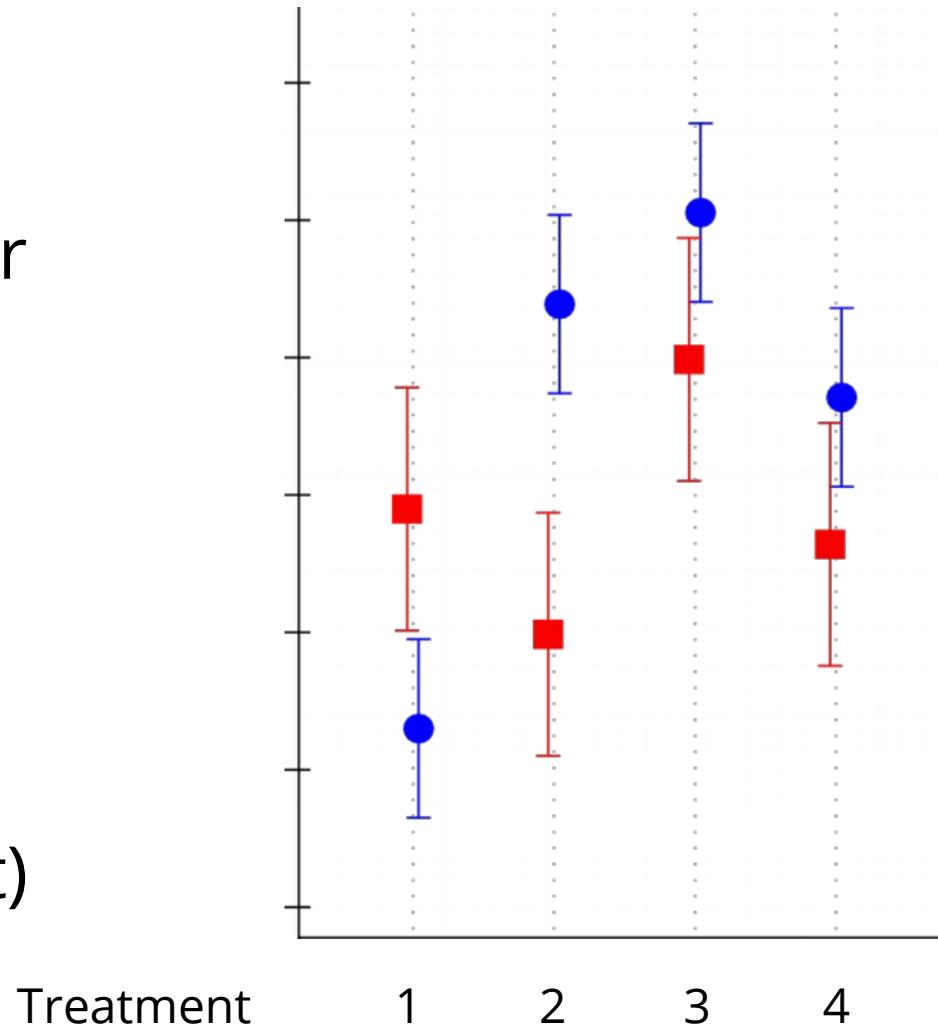
Alternative:

## Polar area chart

*Wickham, 2010*

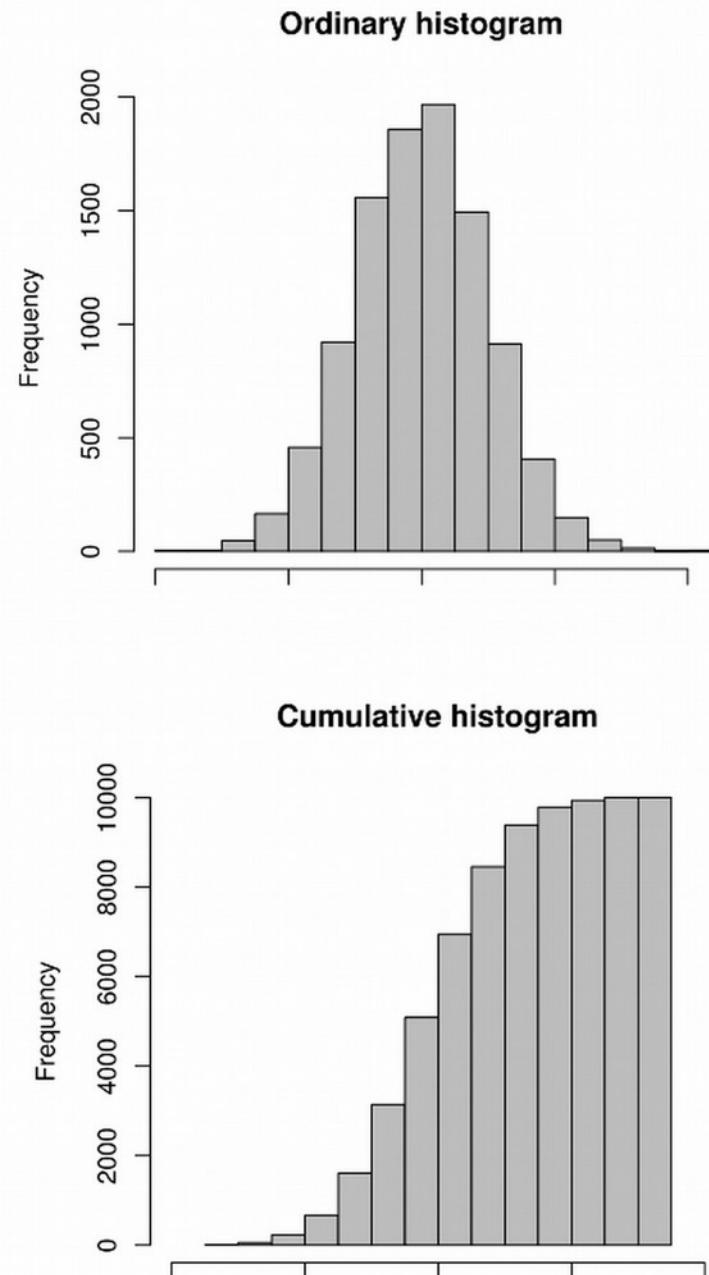
# Bar chart alternative for comparisons: Dotchart with confidence intervals

- Focuses attention on the **relative values** and their measure of **variability**, rather than on the absolute values
- (absolute values are better conveyed using the heights – in a barplot)



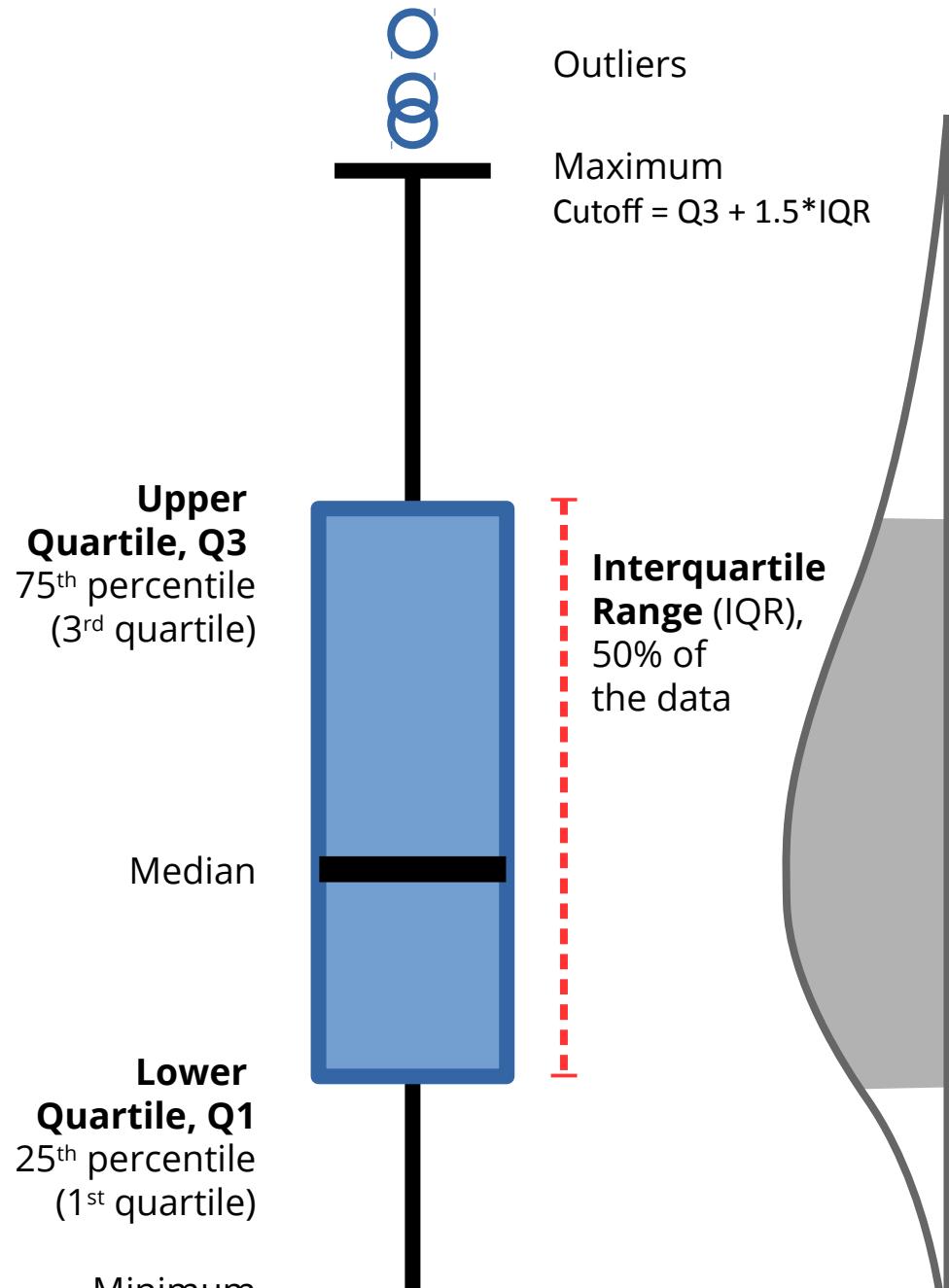
# Histogram – distribution

- To show the distribution of a variable and the relative frequency of values; to explore the data
- Better on big datasets
- Estimate of the probability distribution of the variable
- The number of bins (resolution) affects the perceived shape of the distribution; the same perceptive distortion can occur when using histograms with discrete data
- Rules: Number of intervals  $\approx \sqrt{N}$  and Interval width  $\approx \text{Range} \div \sqrt{N}$



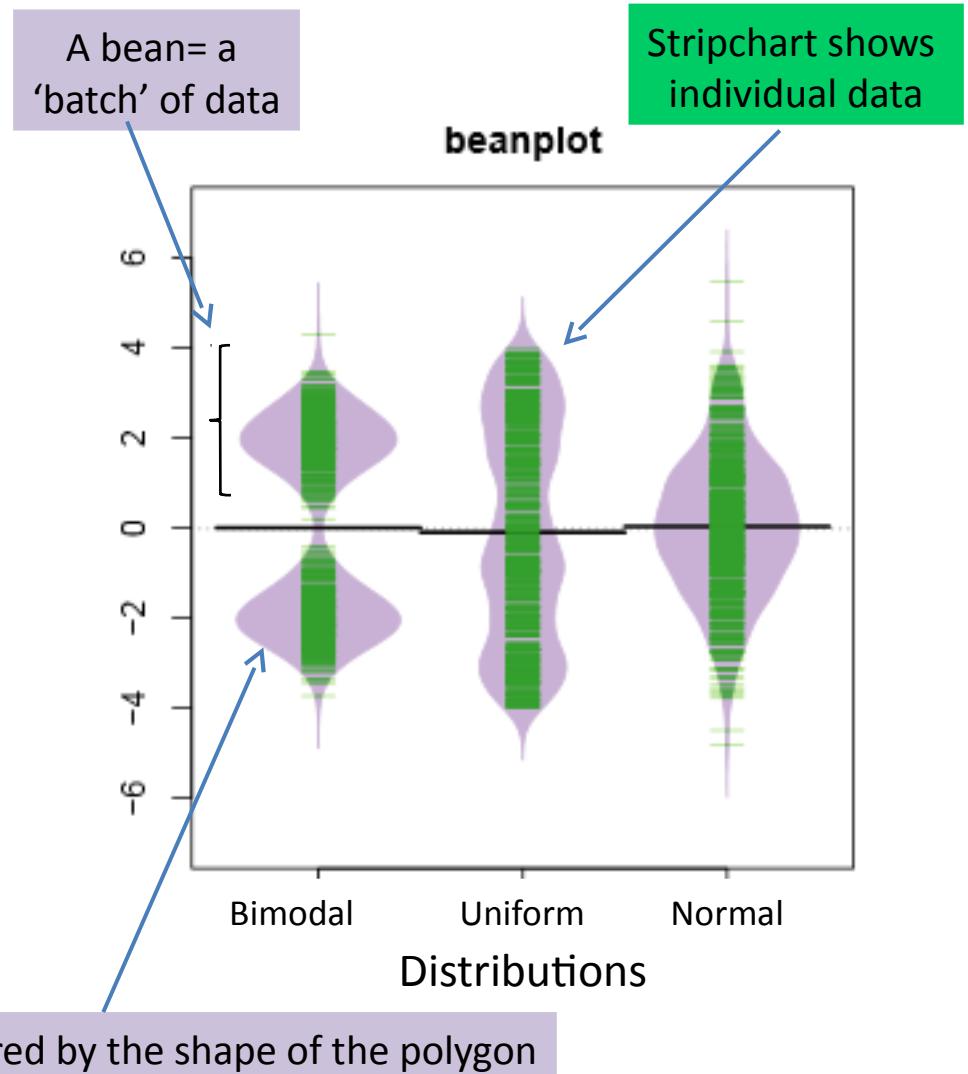
# Boxplot – distribution

- Also ***box-and-whisker*** plot
- Shows the central value, the extremes, and the area where 50% of the values are located.
  - Usually median, minimum, maximum, lowest and highest quartiles
- Particularly useful to understand distribution of not-normal data



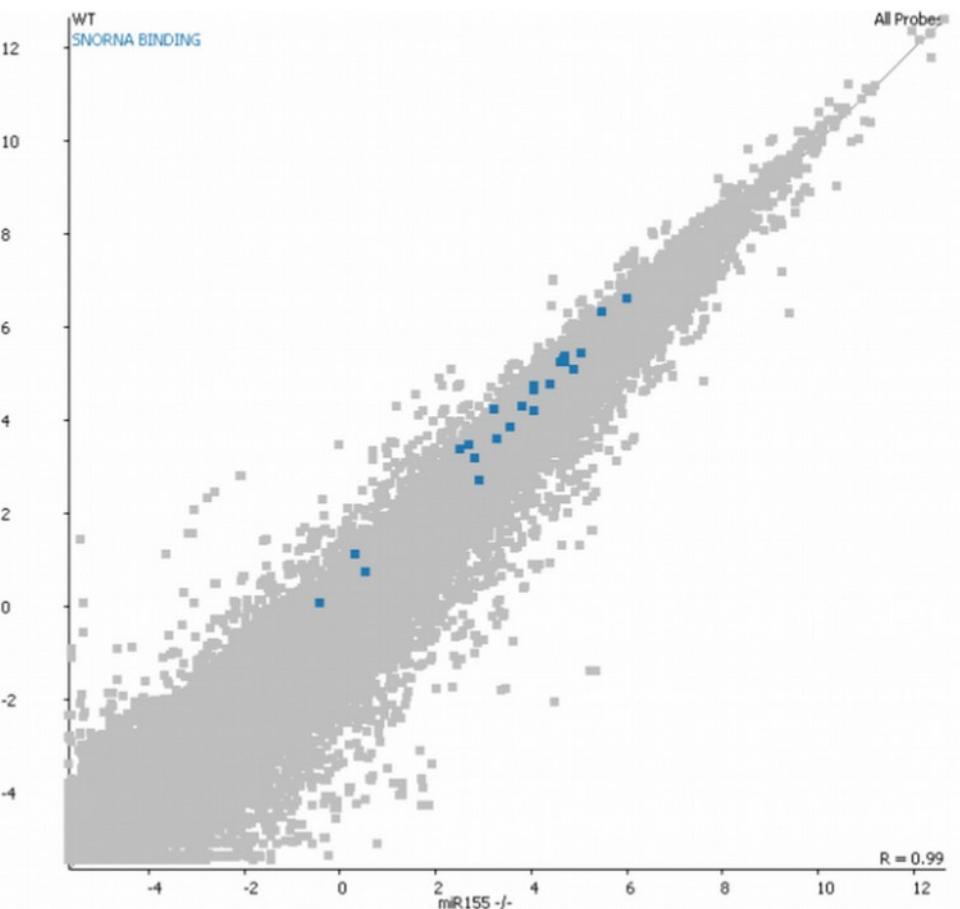
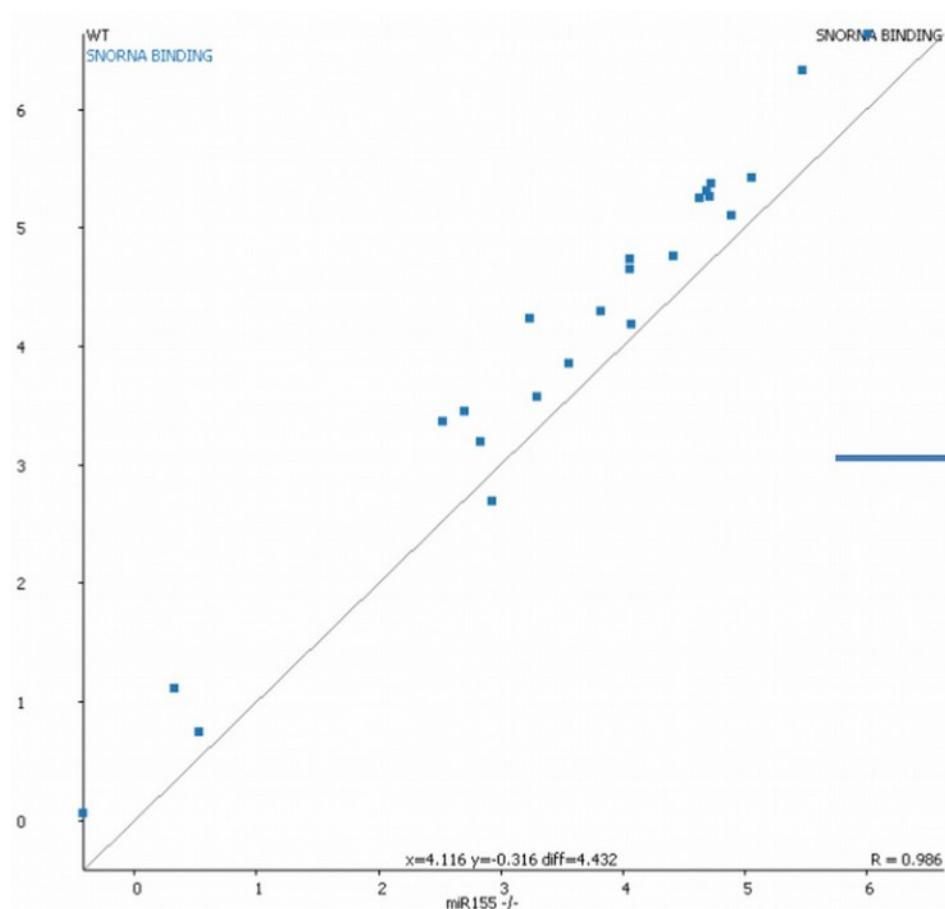
# Boxplot variation: Violin/ Bean plots

- To the above, it adds a **stripchart** of the actual datapoints
- Shows the data **density**
- To understand the distribution in more detail



# Scatterplot – relationships

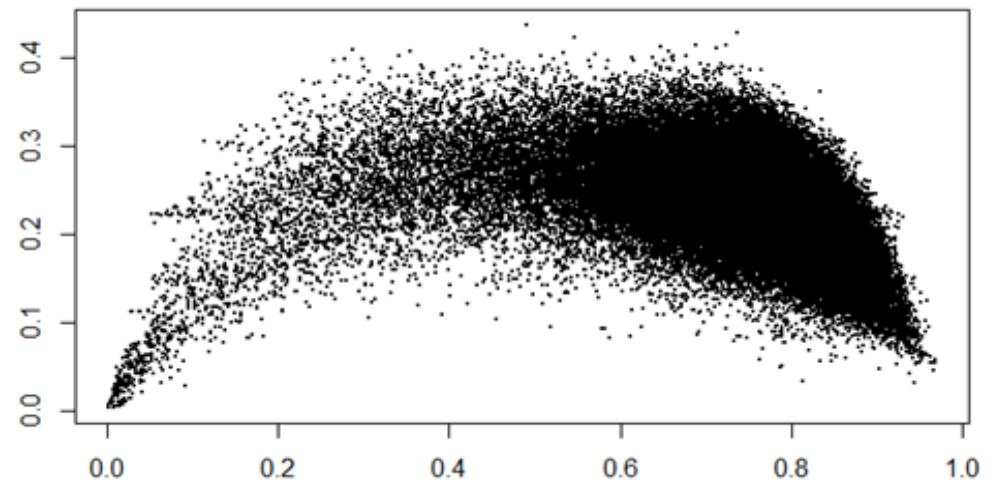
To show the relationship between two continuous variables



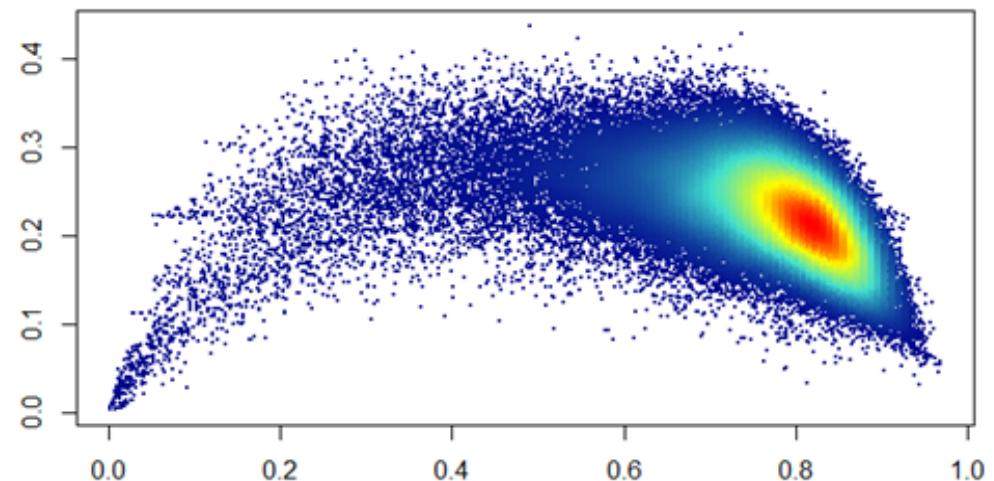
# Scatterplot – relationships

For high-density data: use  
colours or transparency

Problem: very big dataset



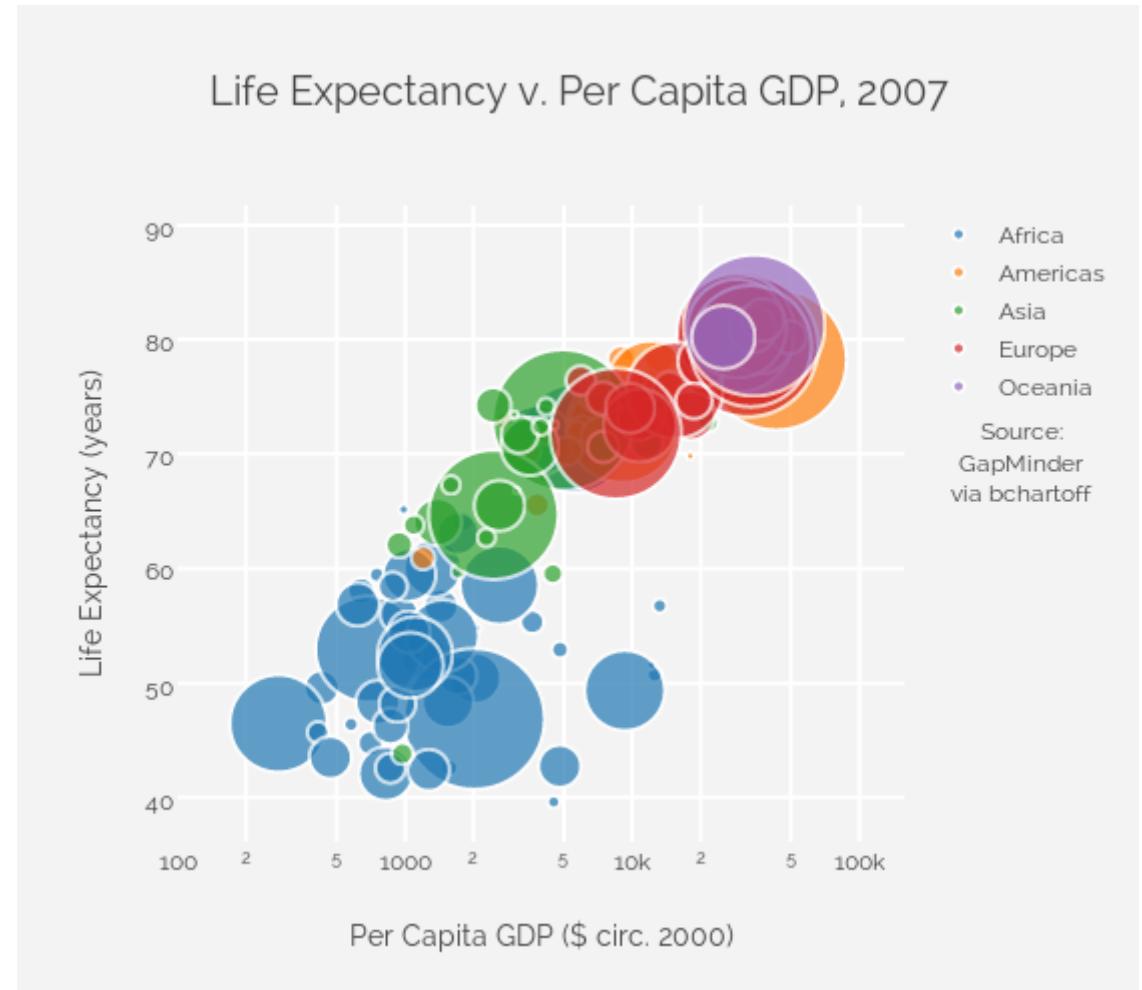
Solution: smoothed **densities**  
colour representation



# Scatterplot variations

## Bubble scatterplot

It adds a 3<sup>rd</sup> dimension (but only for small datasets)

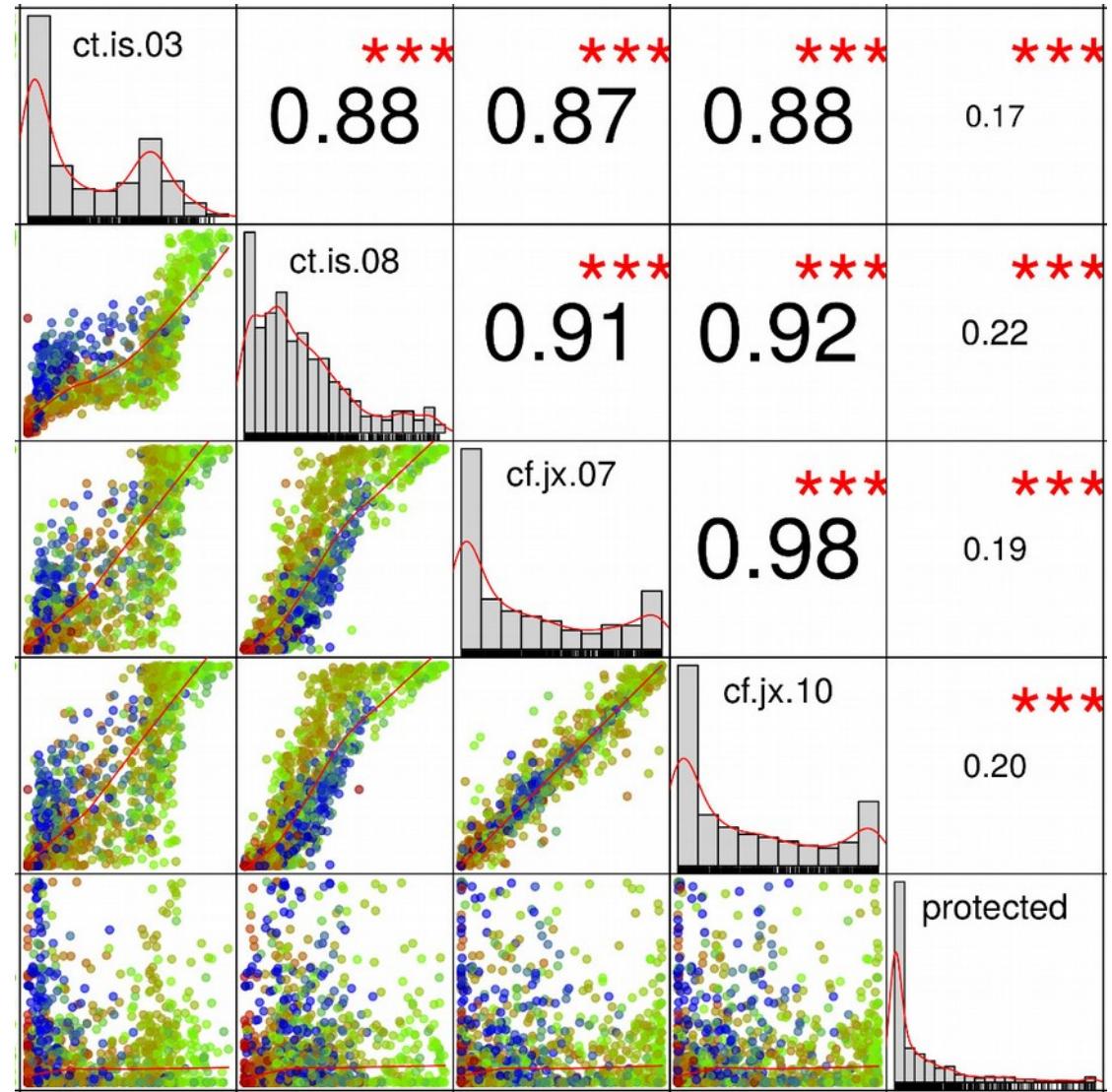


*From plot.ly*

# Scatterplot variations

## Scatterplot matrix (correlogram)

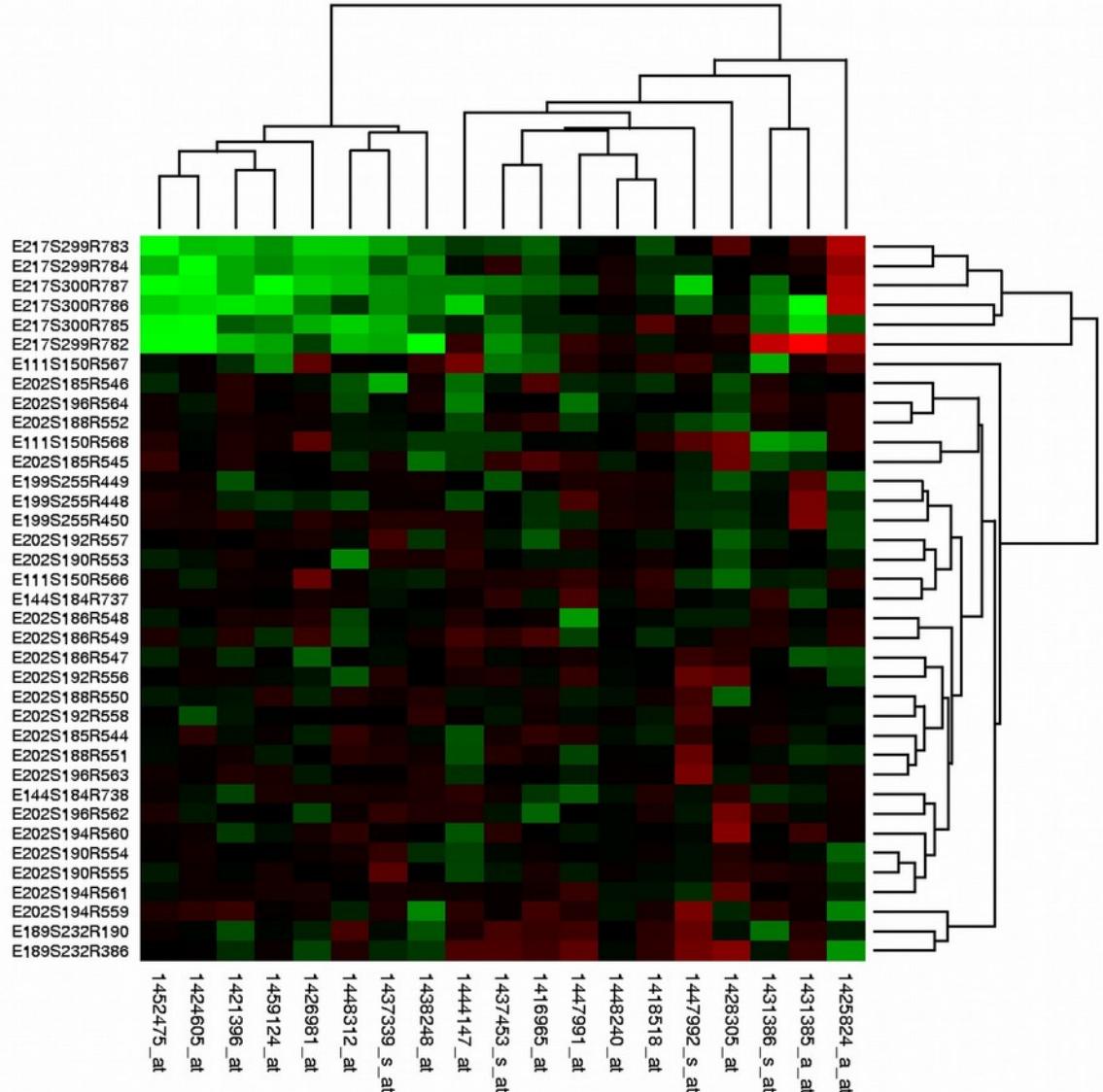
Useful to explore  
bivariate associations  
in a large dataset



Built using **corrgram** package for R

# Heatmap – relationship

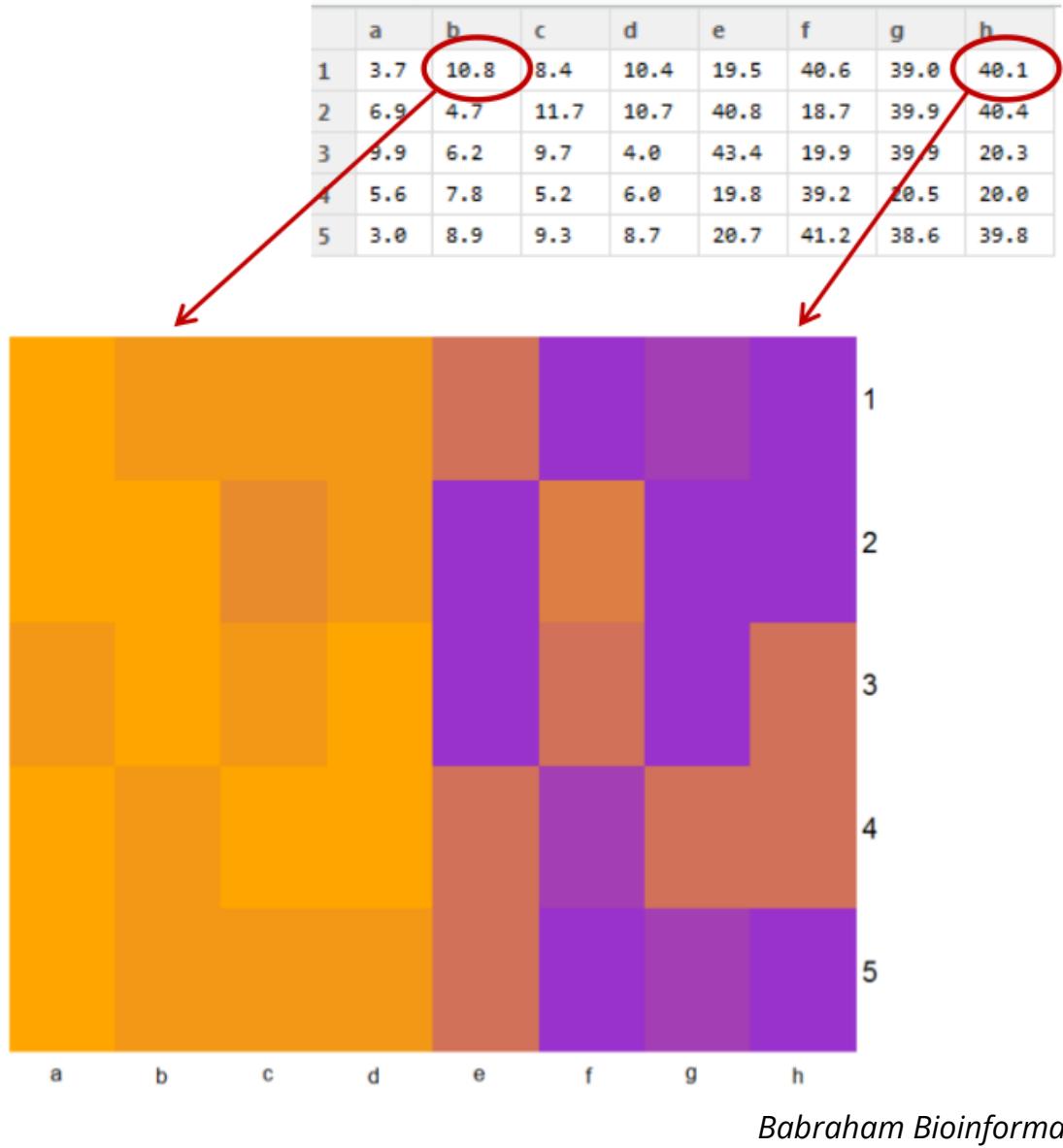
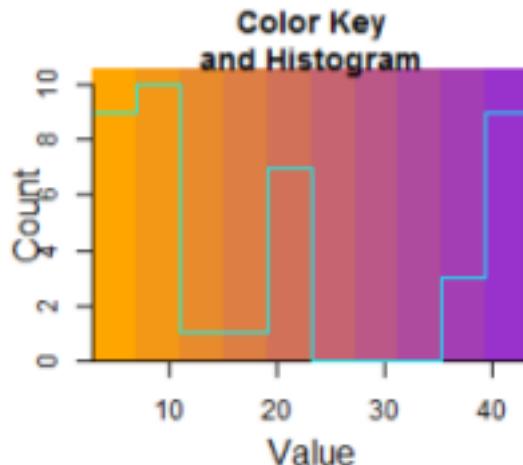
- Shows more complex relationships, e.g. many conditions
- *Steps:* normalisation, clustering
- *Representation:* colouring, filtering



# Heatmap

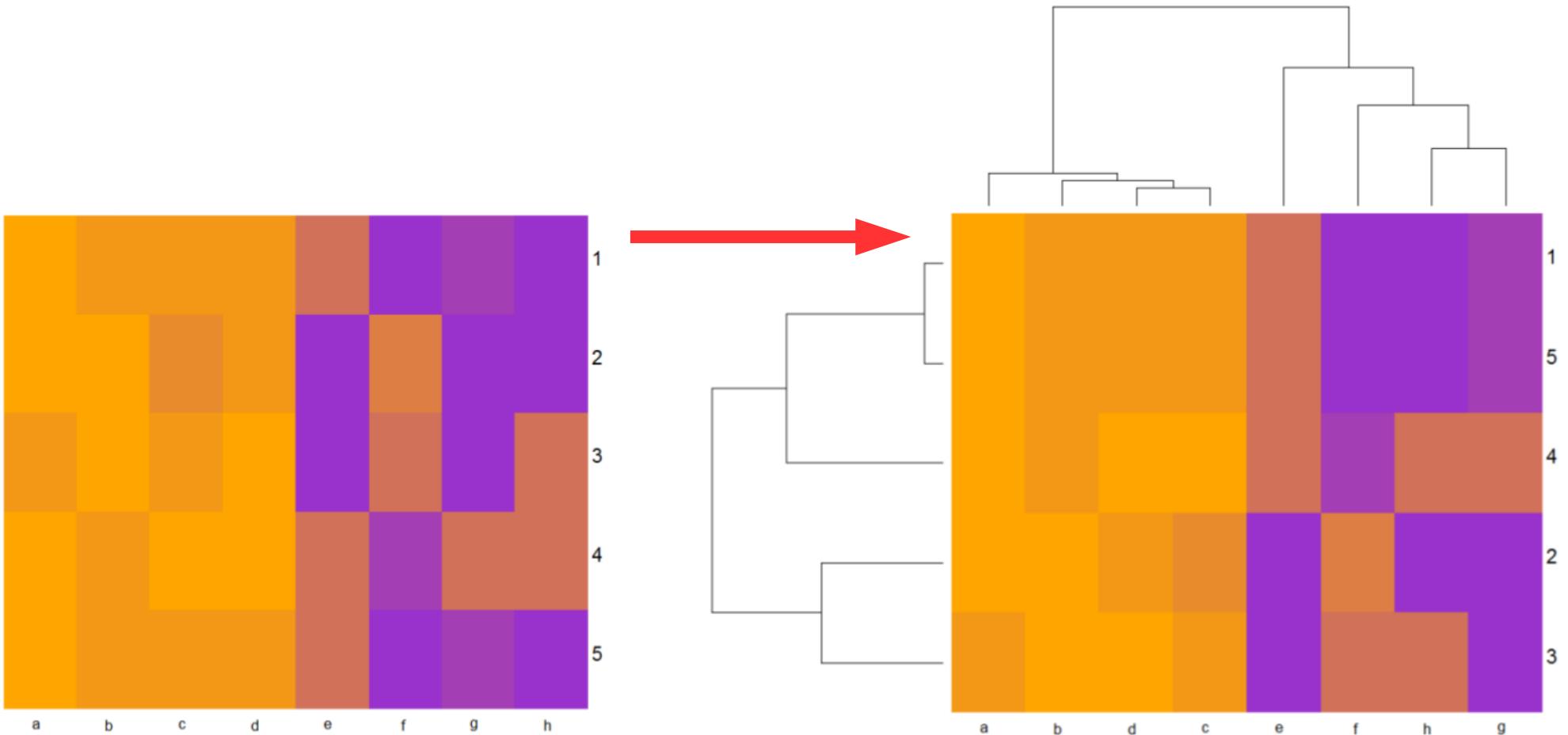
A heatmap is basically a table that has colours in place of numbers

	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	10.5	20.0
5	3.0	8.9	9.3	8.7	20.7	41.2	38.6	39.8



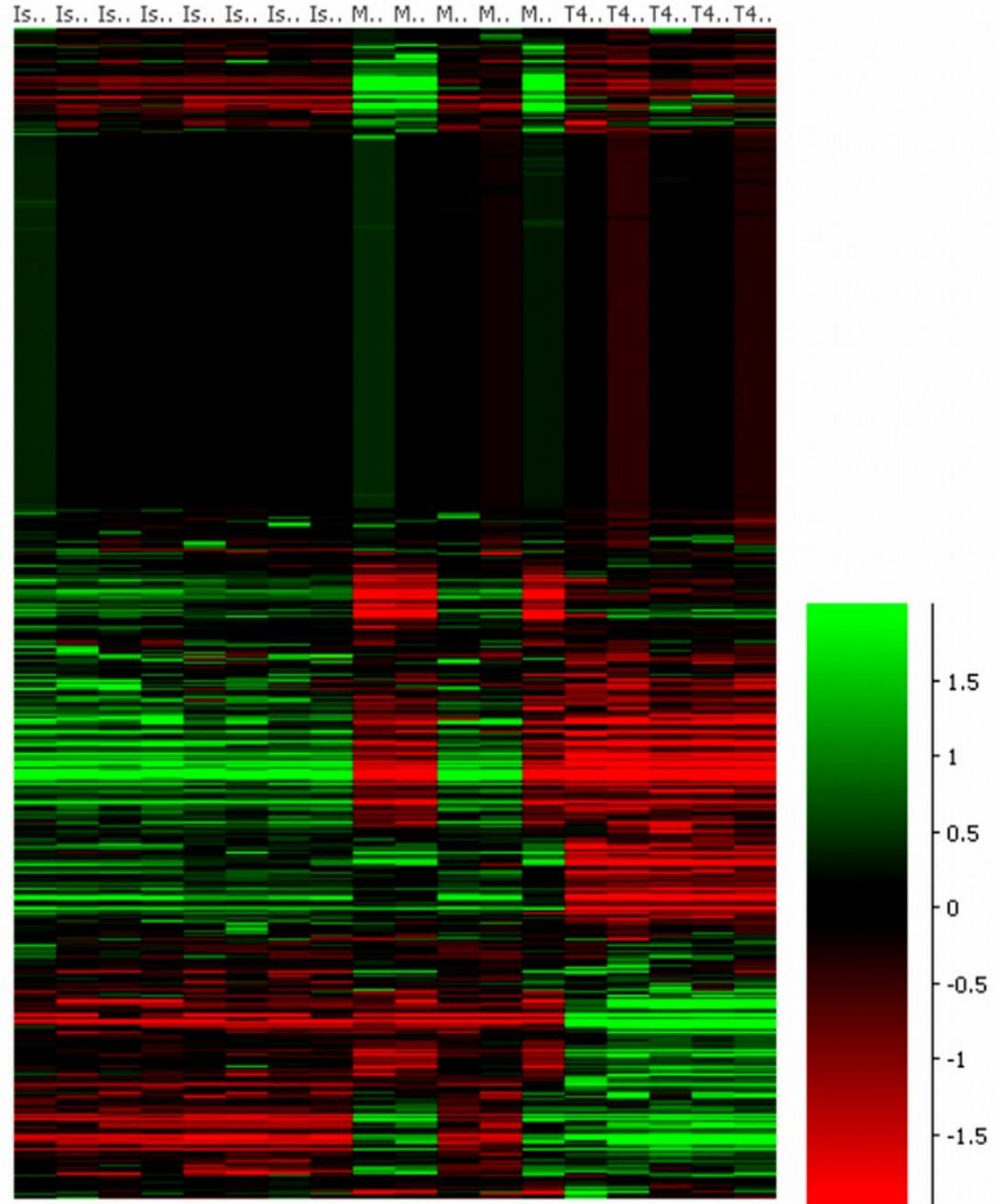
# Heatmap

Colour scheme for grouping: **Clustering** (done usually via Euclidean distances –differences between values)



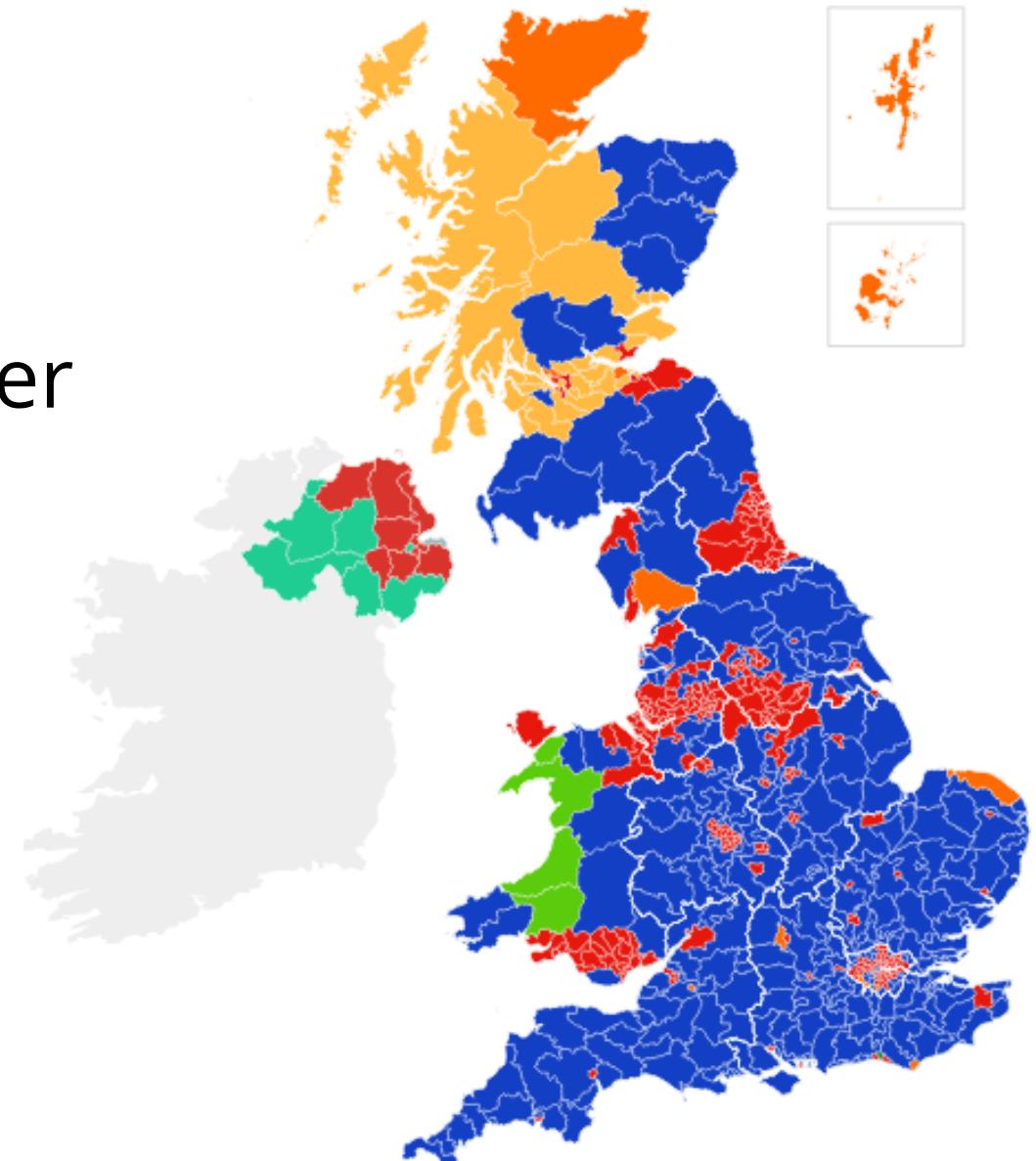
# Heatmap

- Heatmaps are great but:
  - Careful with clustering
  - Plot data that are changing
- Remove unchanging points to focus on differences



# Maps (a very quick look)

Information shown  
over maps has great  
communication power



*Telegraph*

# Maps (a very quick look)

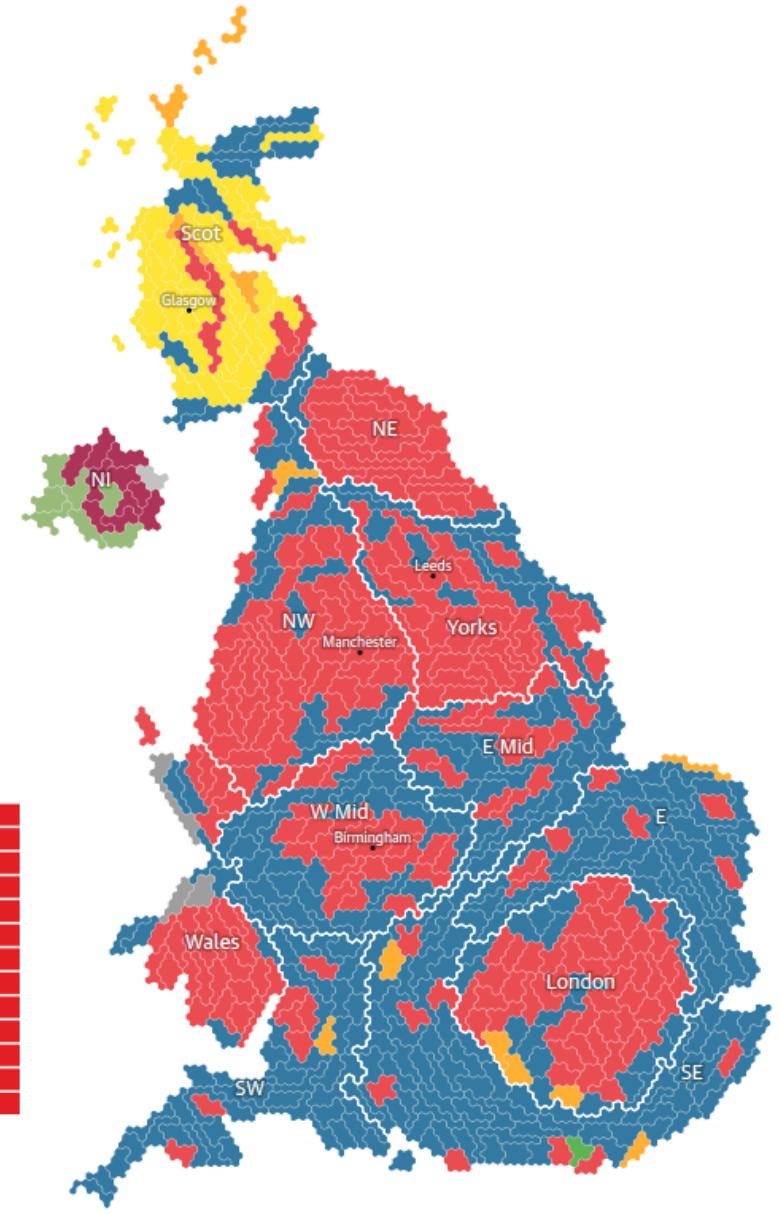
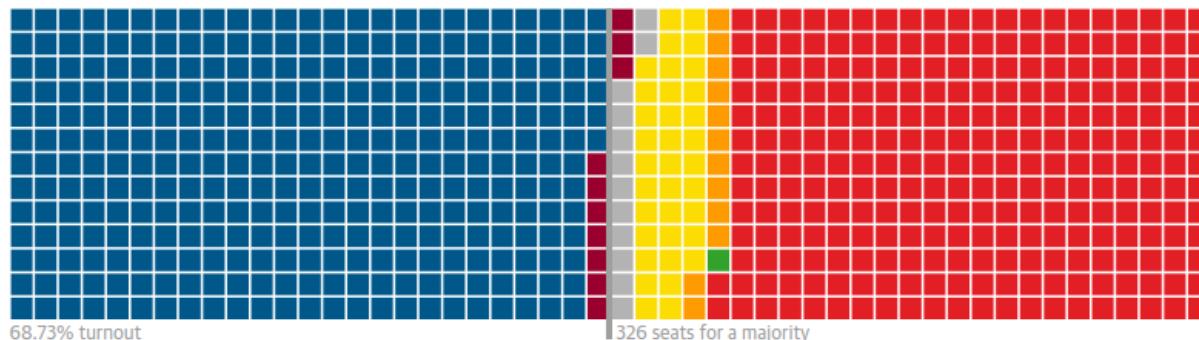
But they are also  
highly prone to  
distortions and to  
biasing perceptions



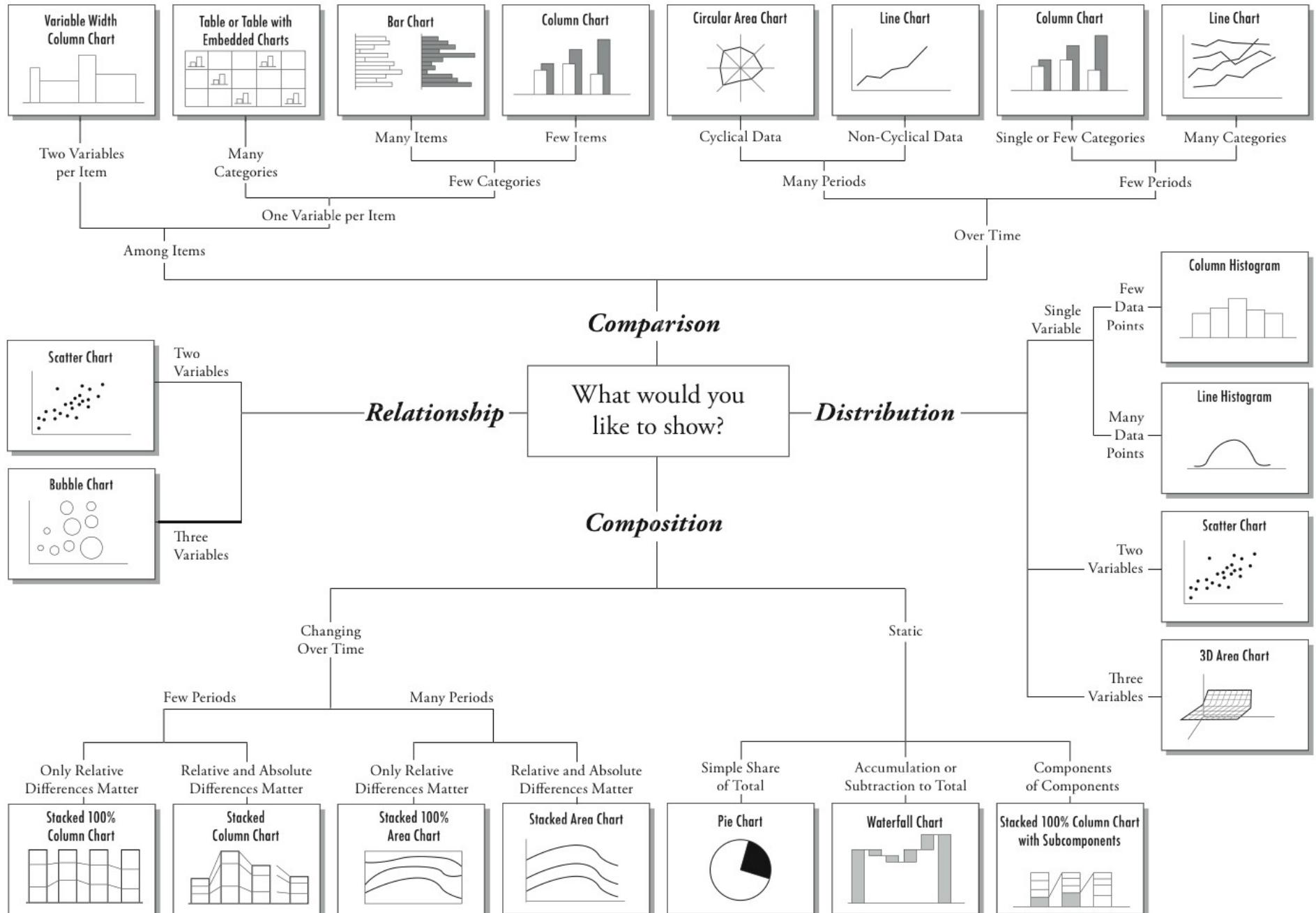
# Maps (a very quick look)

What is the message you want to emphasise?

Geographical distribution? →  
Proportions? ↓



## Chart Suggestions—A Thought-Starter

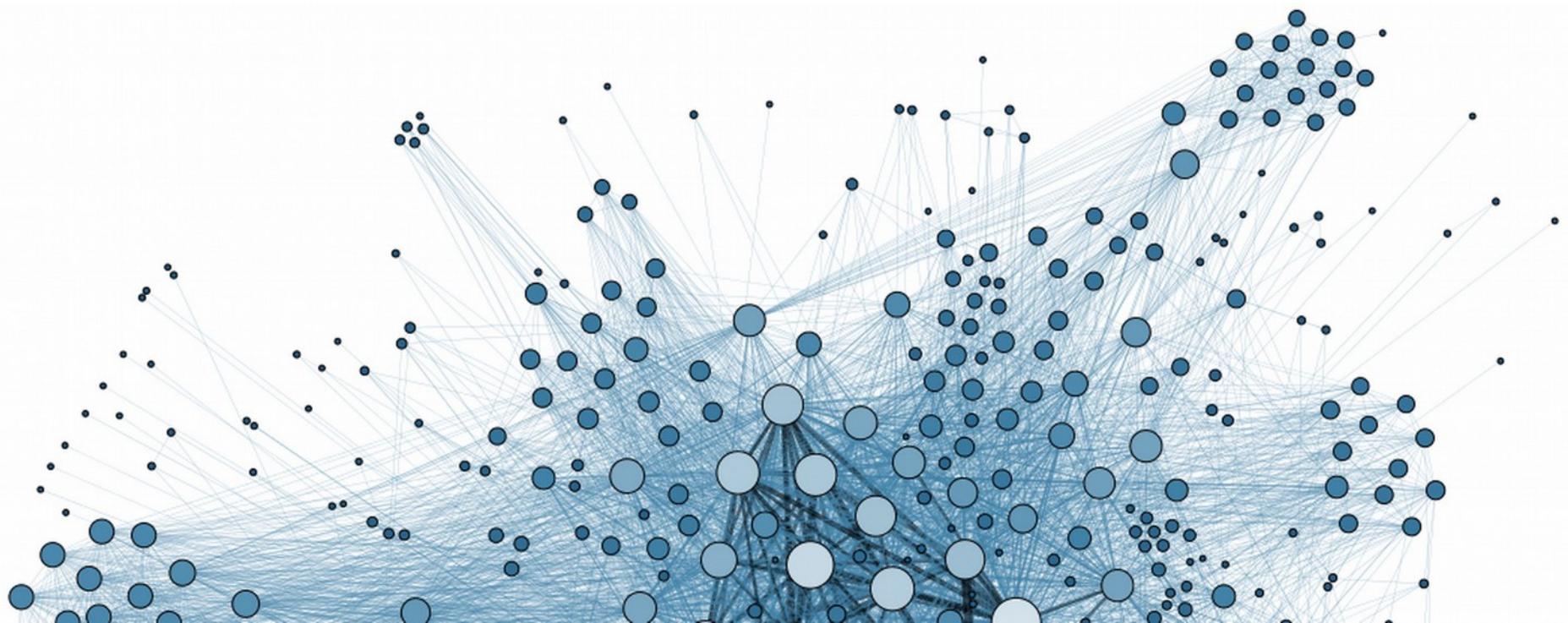


# Summary

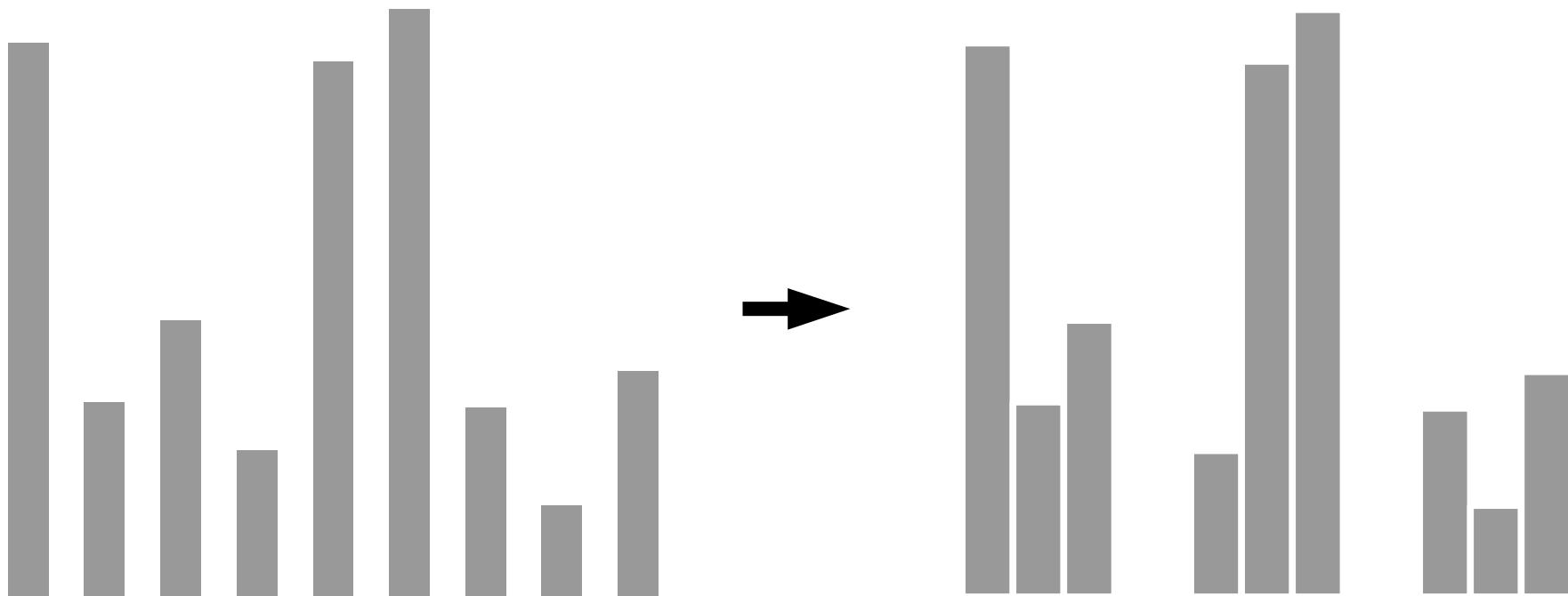
<b>Plot</b>	<b>Aim</b>	<b>Main R function</b>
<b>Stripchart</b>	distribution	<b>stripchart()</b>
<b>Line chart</b>	relationships	<b>plot(type="l")</b>
<b>Bar chart</b> (stacked, norm. stacked)	comparison (and composition)	<b>barplot()</b>
<b>Dotchart with CI</b>	comparison	<b>dotchart()</b>
<b>Histogram</b>	distribution	<b>hist()</b>
<b>Boxplot</b> (violin/ bean)	distribution	<b>boxplot(), vioplot()</b>
<b>Scatterplot</b> (correlogram)	relationships	<b>plot(x, y), corrgram package</b>
<b>Pie chart</b>	composition	<b>pie()</b>
<b>Heatmap</b>	relationship	<b>heatmap()</b>

# Dealing with complexity

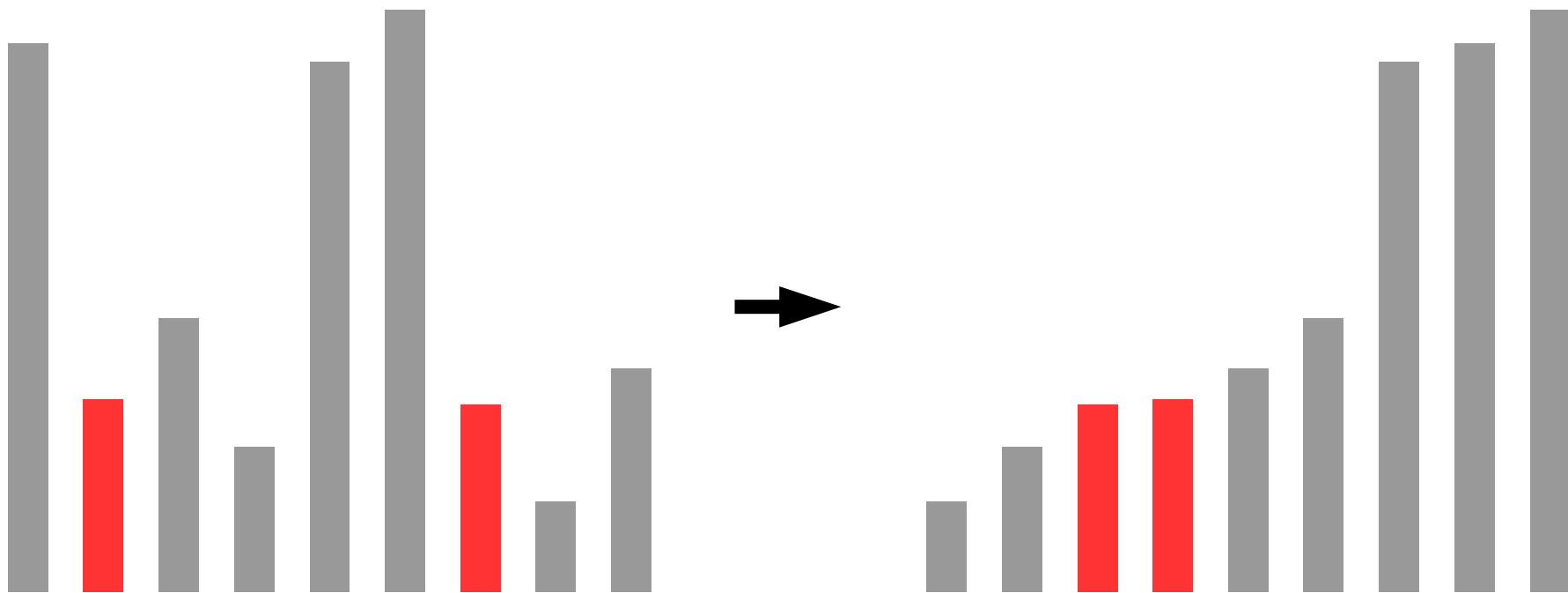
- To focus the viewer's attention onto the main point you want to convey (e.g. on specific subsets of data)
- To require less cognitive load for the viewer to understand the message



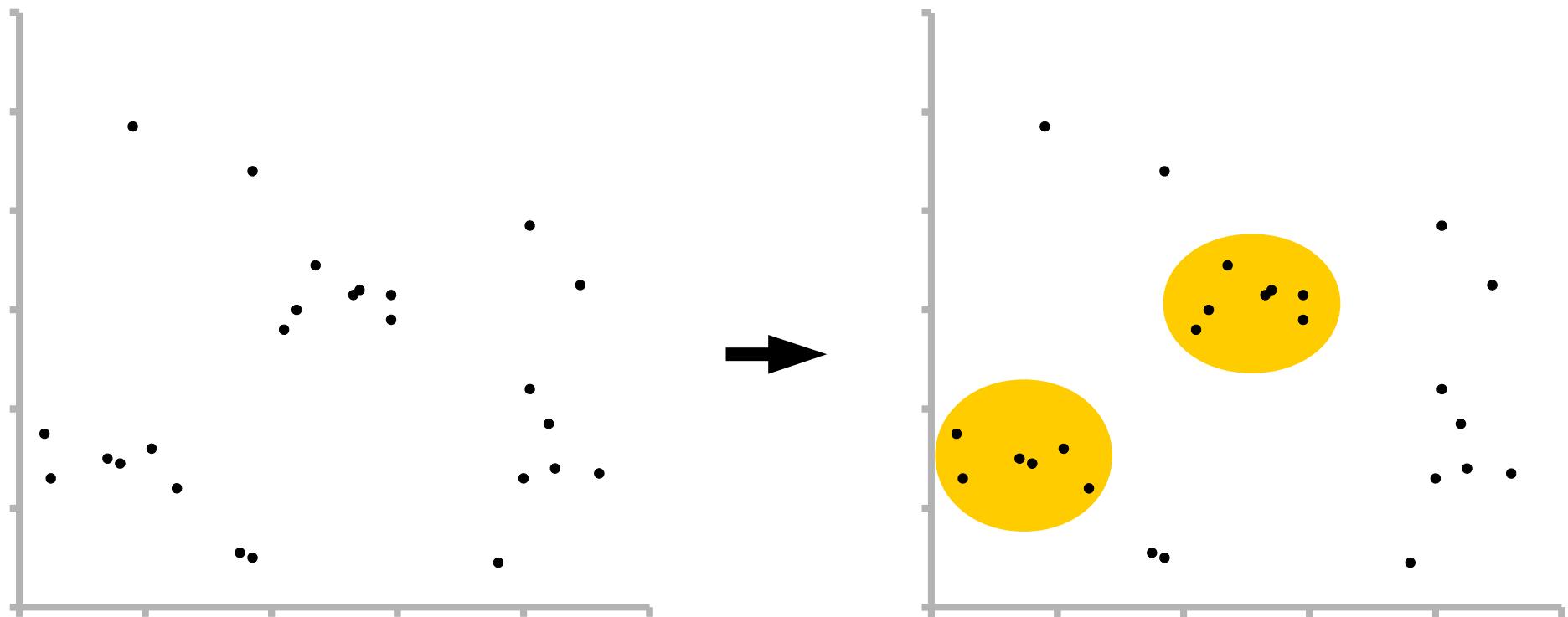
# Grouping



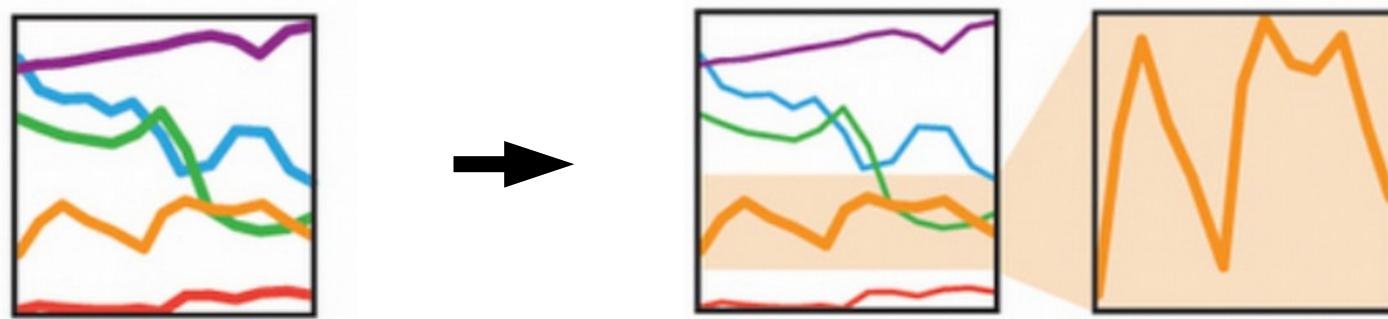
# Ordering (only for categories)



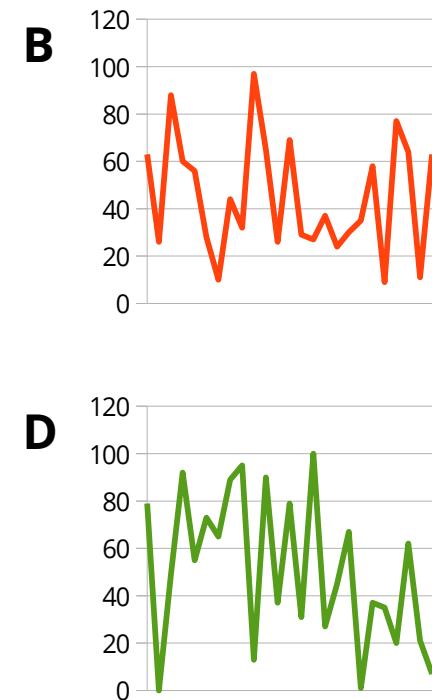
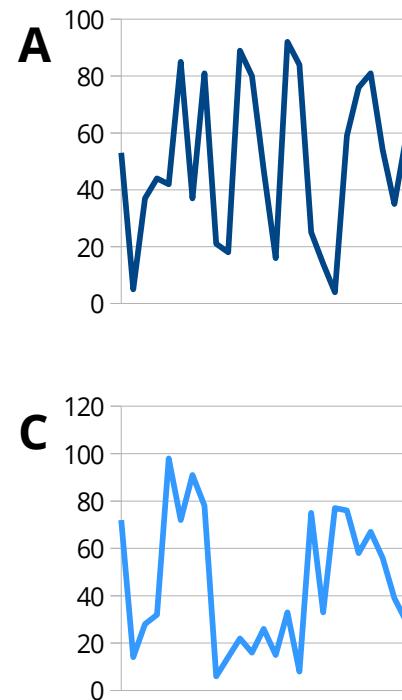
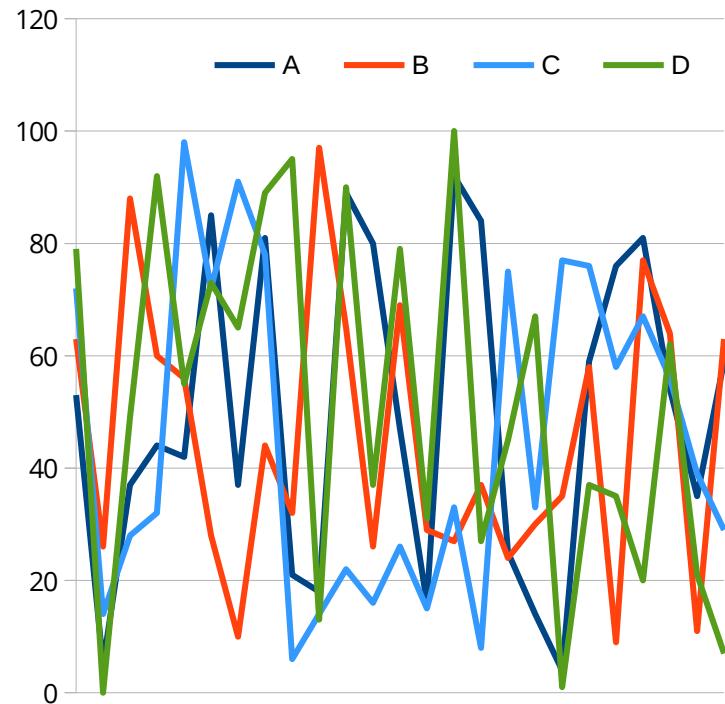
# Containment



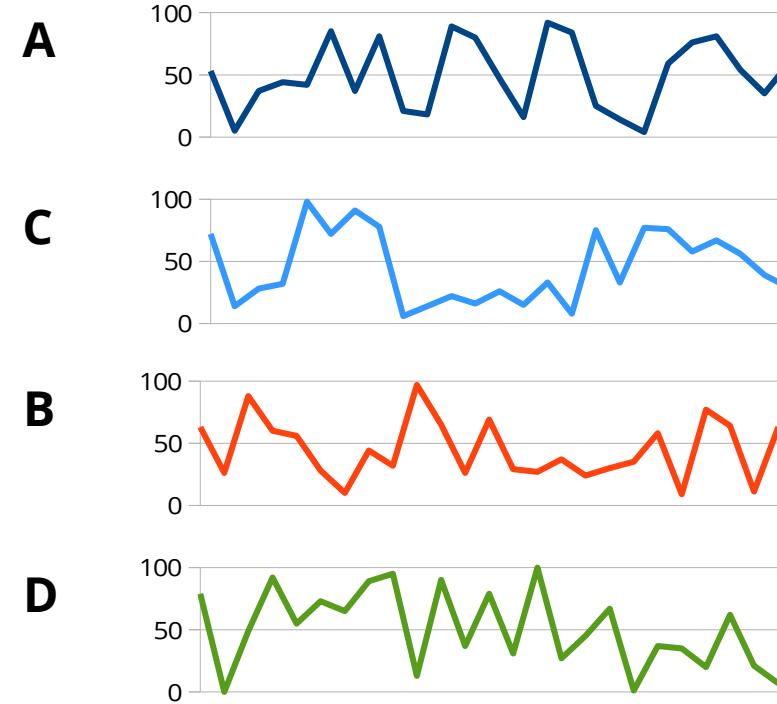
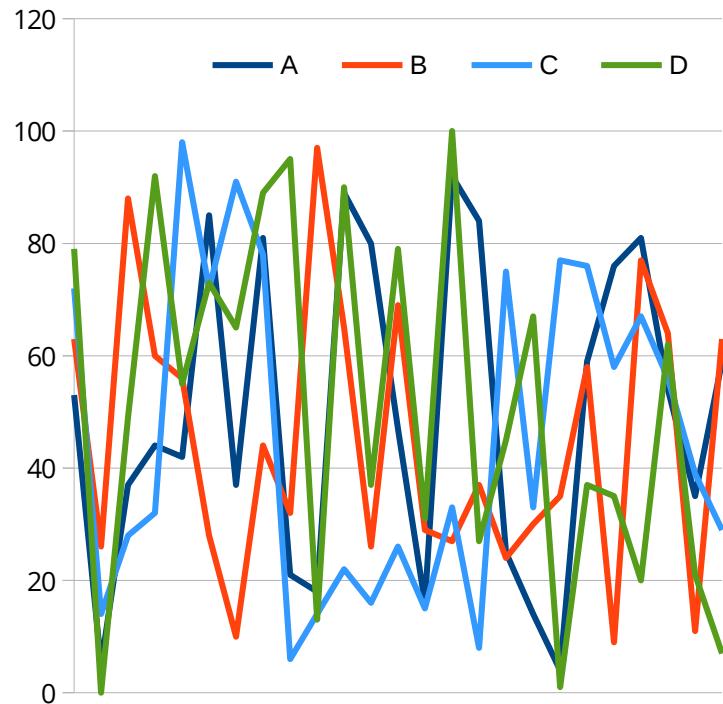
# Filter, link, embed



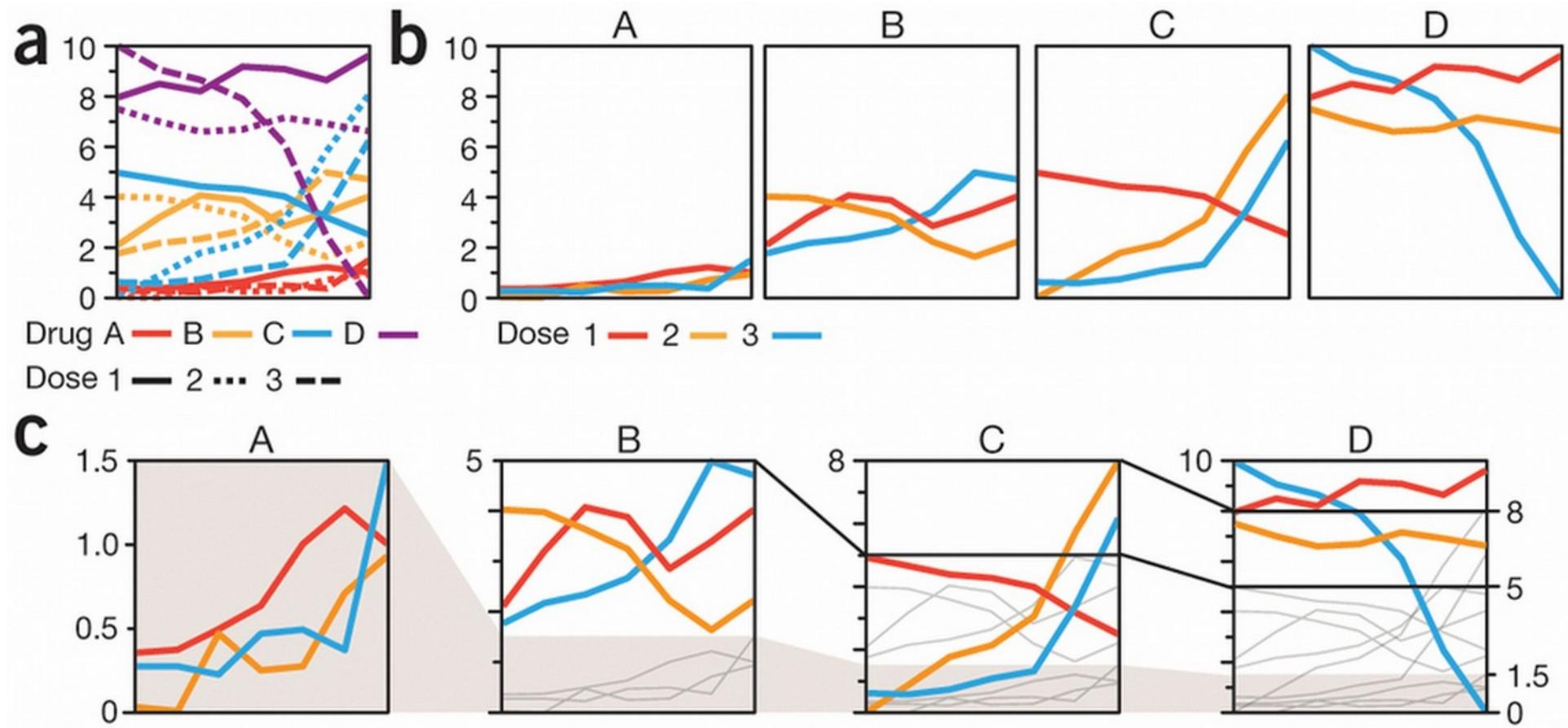
# Small multiples



# Small multiples



# Small multiples

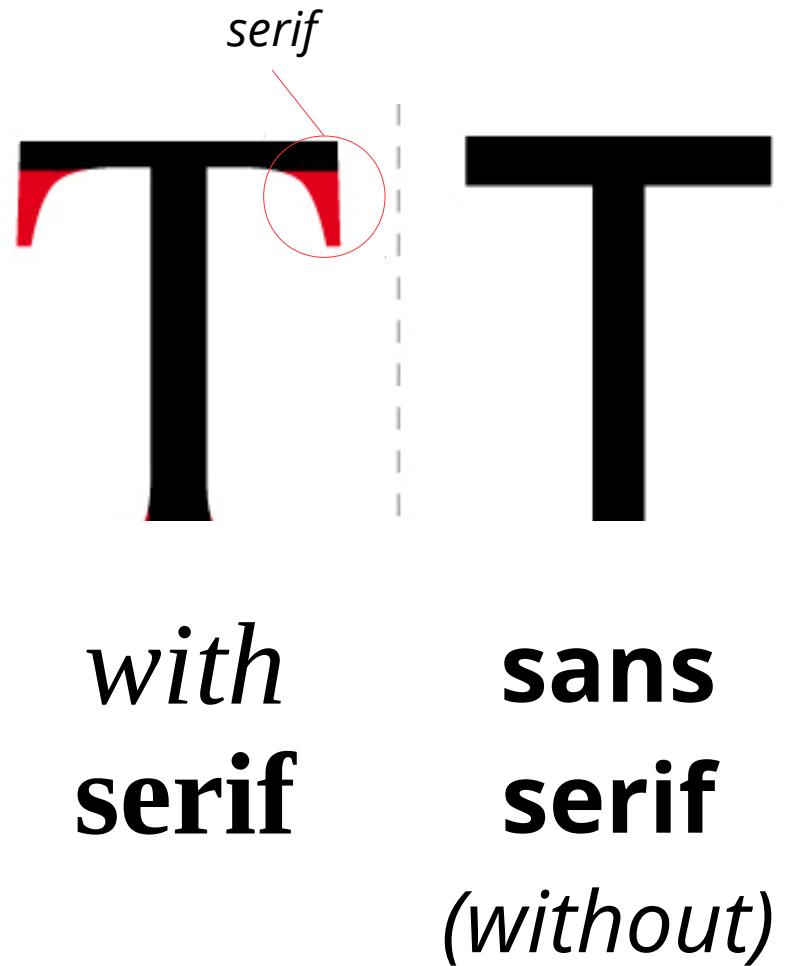


# Typography (fonts)

- All the elements need to be labelled
- The essential criteria for choosing fonts is **readability**:
  - **Scalability** (readable at small sizes)
  - **Contrast** with the background
- Fonts convey a personality, mood or attitude (some more than others)

# Typography

- **Serif** for large blocks of text,  
**sans-serif** for titles, labels  
and annotating figures
  - Sans-serif is easier to read at smaller sizes
- **Sizing:** the size of fonts is given in points, and it's the size of an imaginary block of metal that is used in printing.
  - In practice, the only way to know exactly how well your font will be read is to print it.



# Typography

- **Monospace** is good for code, or for text intended to be aligned from line to line (e.g. pseudo-tables)  
e.g. **m** vs **m** ; **i** vs **i**
- **Casing:**
  - UPPERCASE,
  - lowercase,
  - Sentence case,
  - Title Case.
- Check the journal guidelines for font types

**Monospace font keeps the alignments tidy.**  
**(this is monospace! )**

**Monospace font keeps the alignments tidy.**  
**(not monospace font)**

*Each of the lines above has 20 characters.*

# Typography: Guidelines



UNIVERSITY OF  
CAMBRIDGE



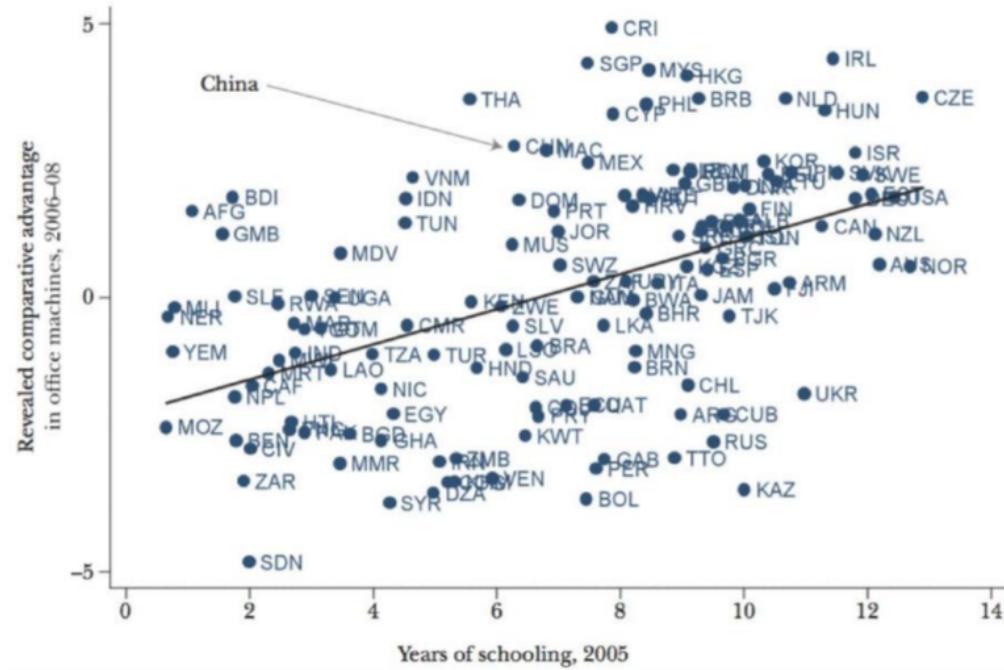
UNIVERSITY OF  
CAMBRIDGE

Avoid **aspect-ratio distortions**: changing font height or size.

- The same applies to images and circular objects
- Scale axes using comparable units

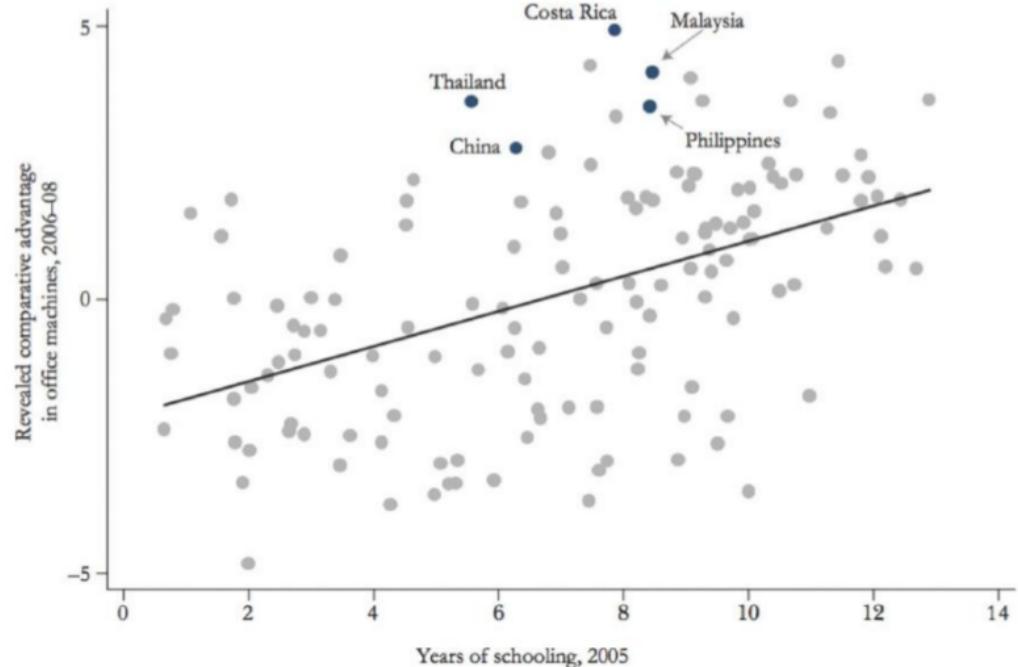
# Typography: Guidelines

## **Education and Exports of Office Machines**



# Minimise text; keep it simple

## **Education and Exports of Office Machines**



# Typography: Typesetting

- Is the arrangement (spacing) of characters in words, lines or paragraphs
  - **Tracking:** space between characters
  - **Leading:** line height
  - **Paragraph alignment:** left, justified, etc.
- Important considerations where figures have many annotations, and in axis and figure titles.

# Typography: Guidelines

- **Avoid colour** in text, particularly in figures (to maximise contrast)
  - **Do not tilt** text, always horizontal (or vertical)
  - Check **scalability**: text should be readable after resizing
-  Typeset in blocks of text that are **solid shapes**
-  Avoid typeset in blocks of text that are not **solid shapes**

# Typography: Heed the numbers in your font

1	I	1	1
2	2	2	2
7	7	7	7
6	6	6	6

- Each font has different styles of numbers
- Make sure that the font you choose distinguishes them well (e.g. I in *Gill Sans*) and is legible at small sizes

# Typography:

## Think your words carefully

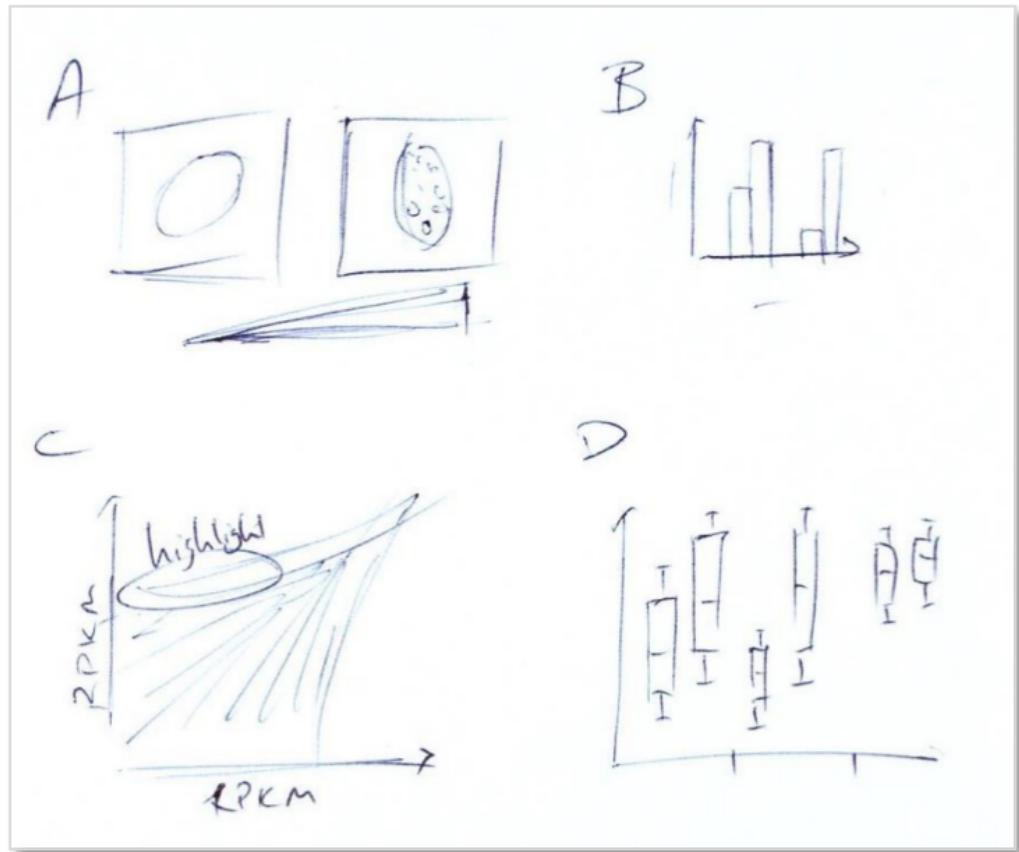
- Avoid wordiness... it's a figure!
- Choose words that “precisely convey what you mean”
- Avoid contractions and spell out whenever possible

# Composition and layout

- Draft
- Grid and alignments
- Balance and hierarchy

# Composition and layout

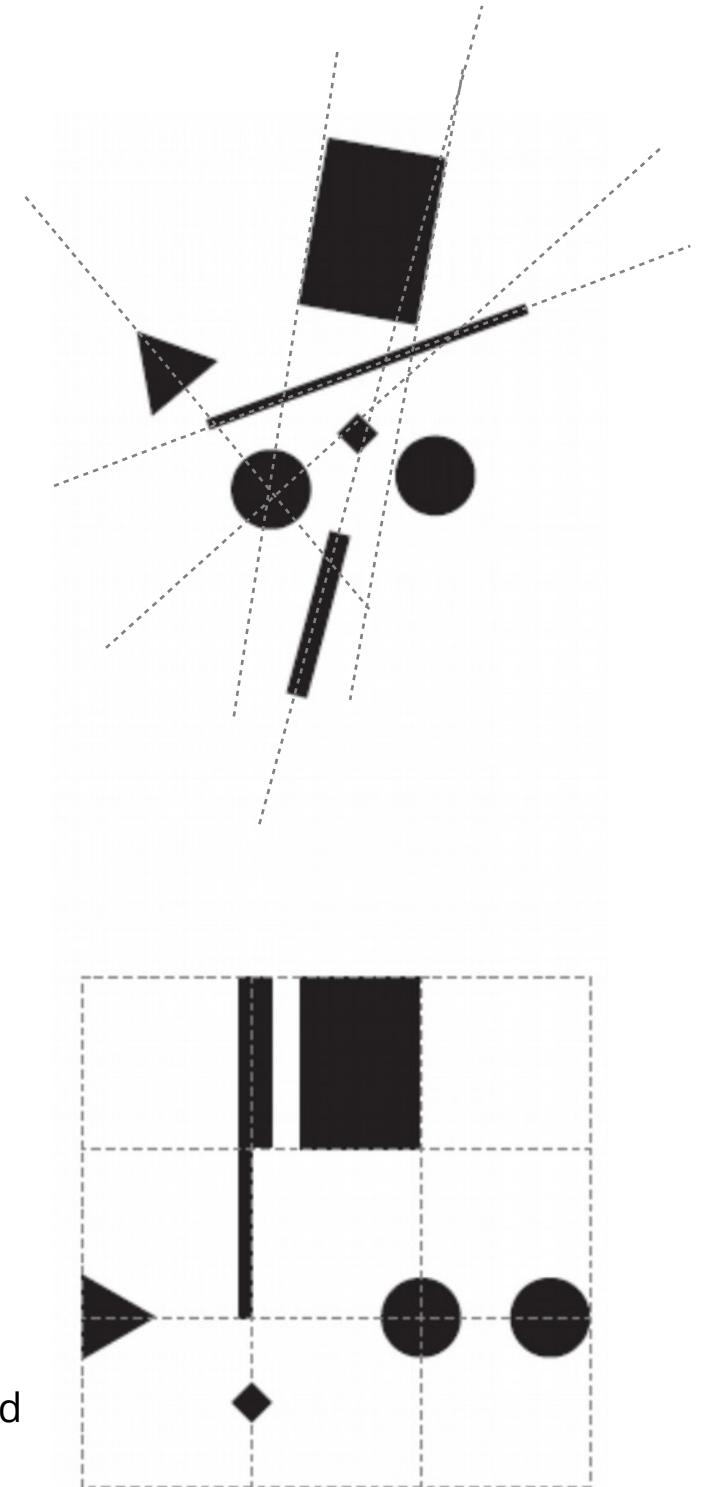
- Have an idea of what your final figure will look like
  - What message are you trying to convey?
  - How does each figure contribute to that message?
  - Identify what is essential (Supporting Information)
- Outlines can reduce time spent moving or resizing images



# Grids

- Grids are the invisible structure behind a composition that makes it look balanced
- Every alignment (of a box, column, text line and text margin) creates a **visual line** in the grid
- Conversely, a composition where elements are aligned to a grid creates a sense of balance

Grids can help to organize the spaces around and in-between elements. *Roland et al 2011*



# Alignments



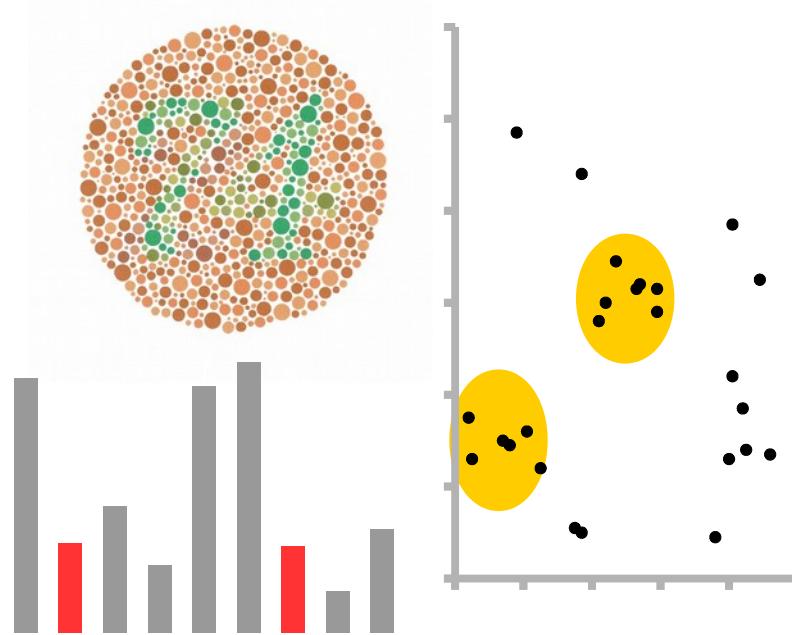
# Alignments



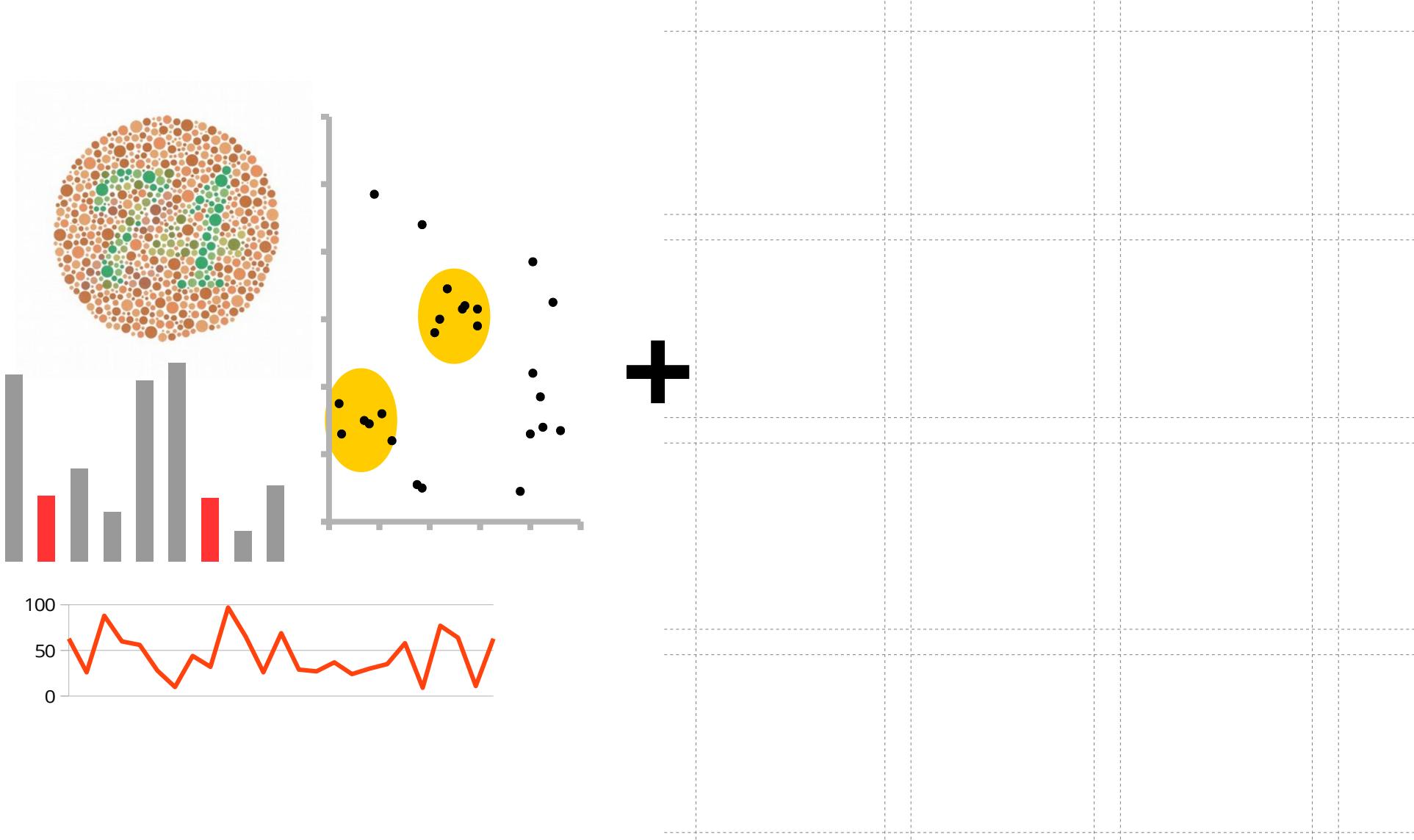
**Use tools to align objects, don't do it by eye!**

Most programmes have tools for automatic alignment and to distribute objects with equal space.

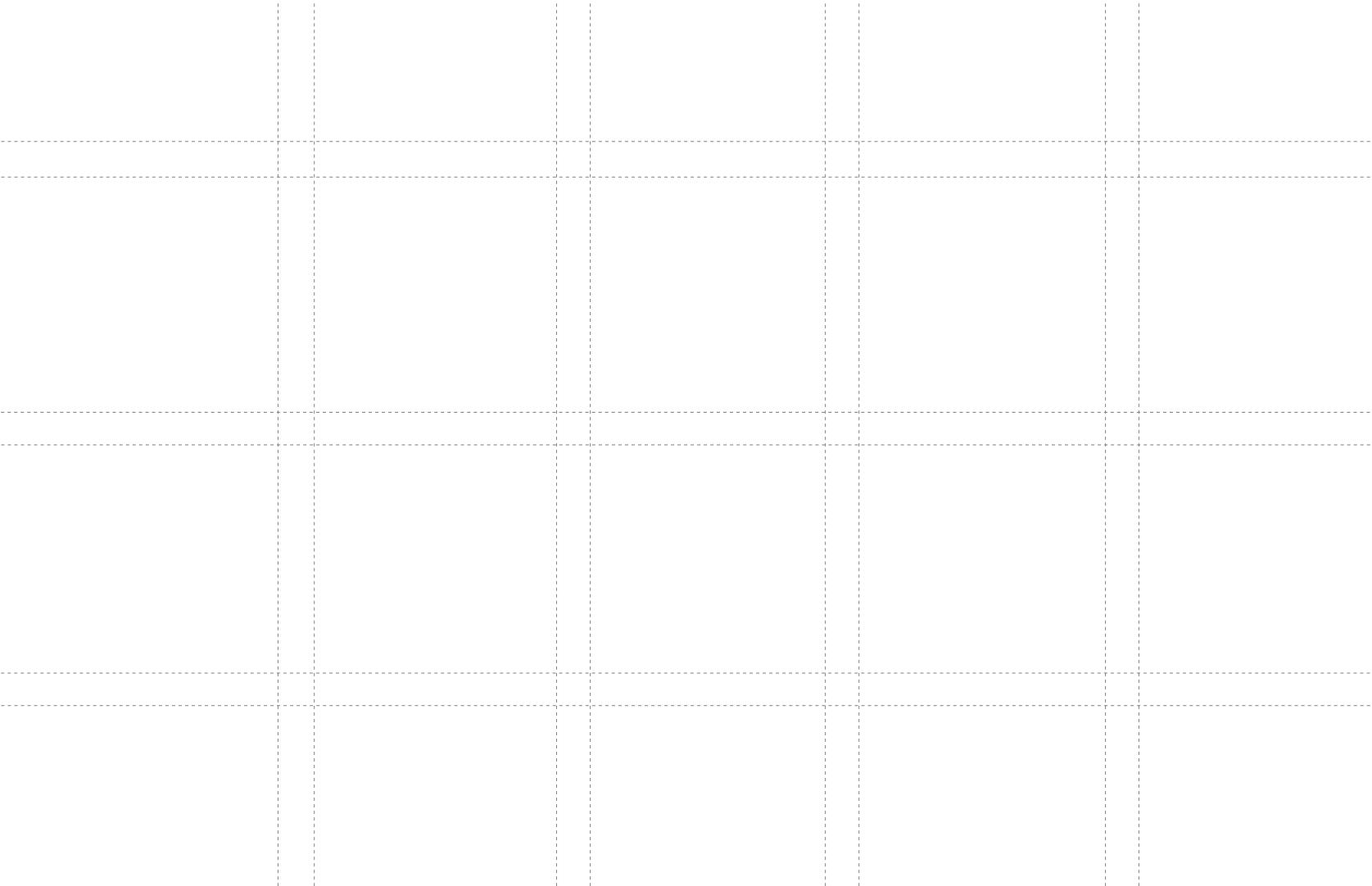
# Using grids



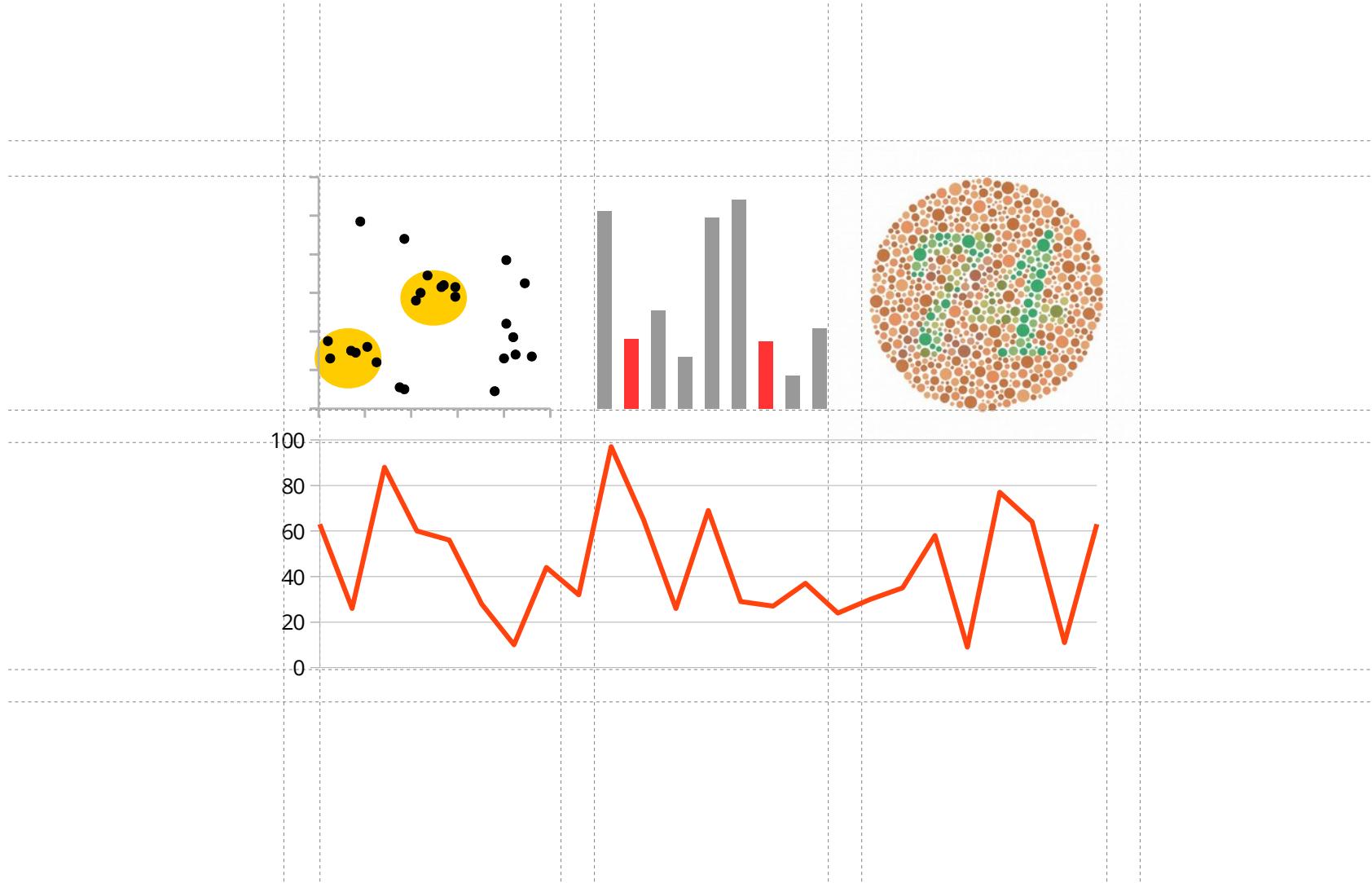
# Using grids



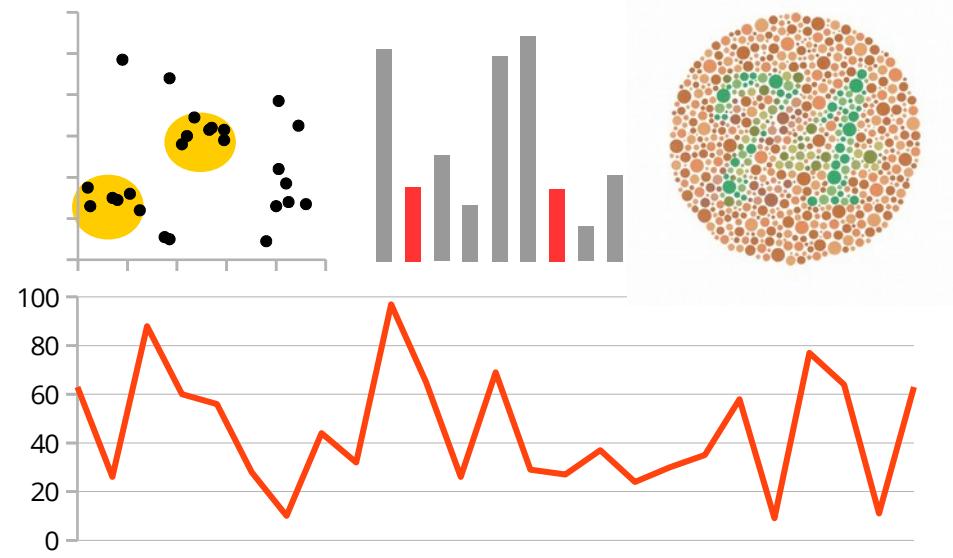
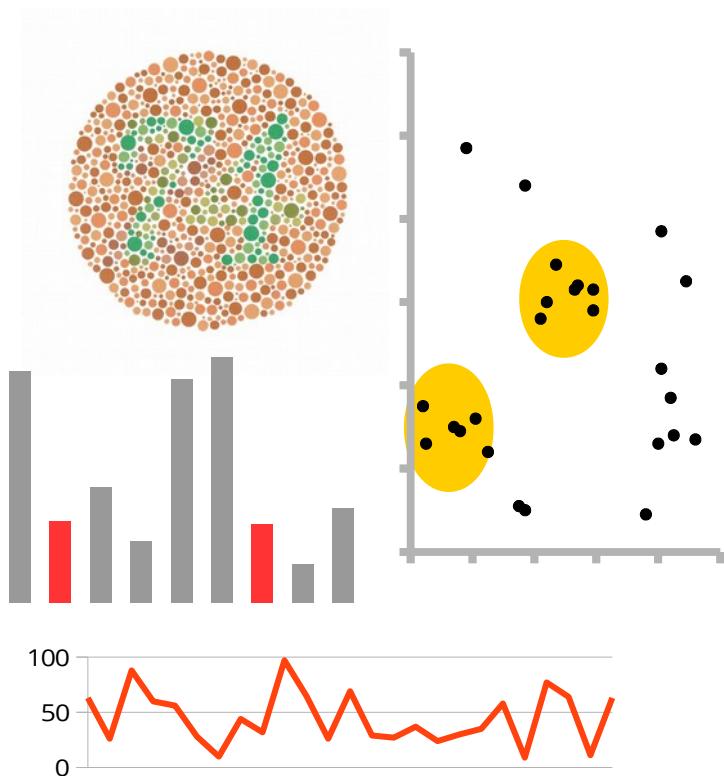
# Using grids

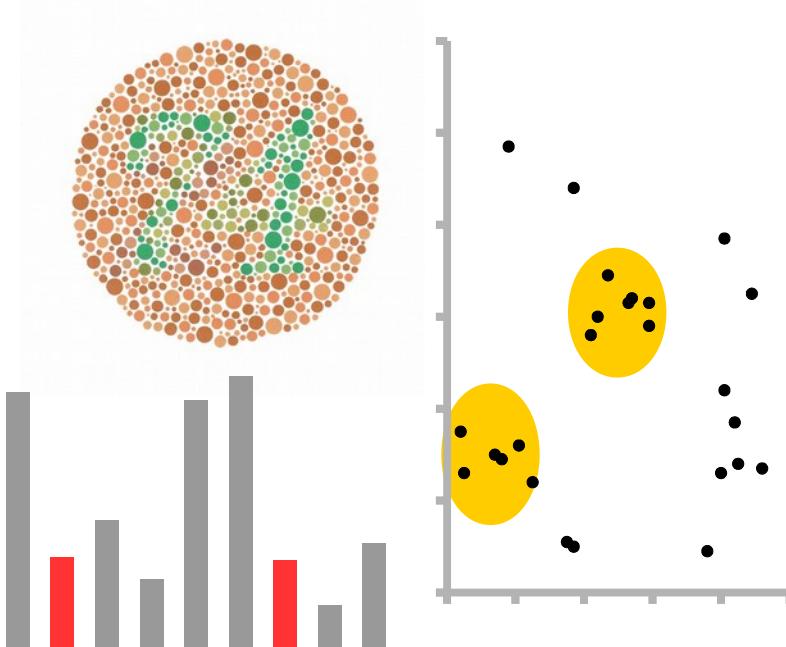


# Using grids

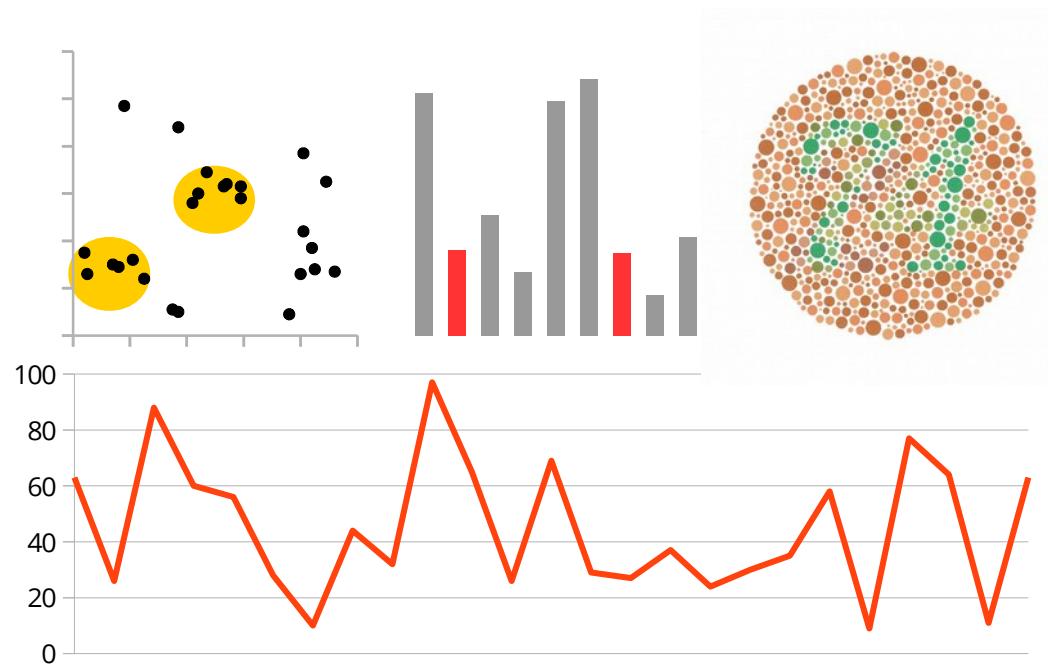


# Using grids





# Using grids



# Visual balance and hierarchy

The composition of a graphic object and the **emphasis** on each element will determine what is the **hierarchy between elements**, and how the eye will **flow** and where it will **focus**

Keep a balance between **white space**, text and figures

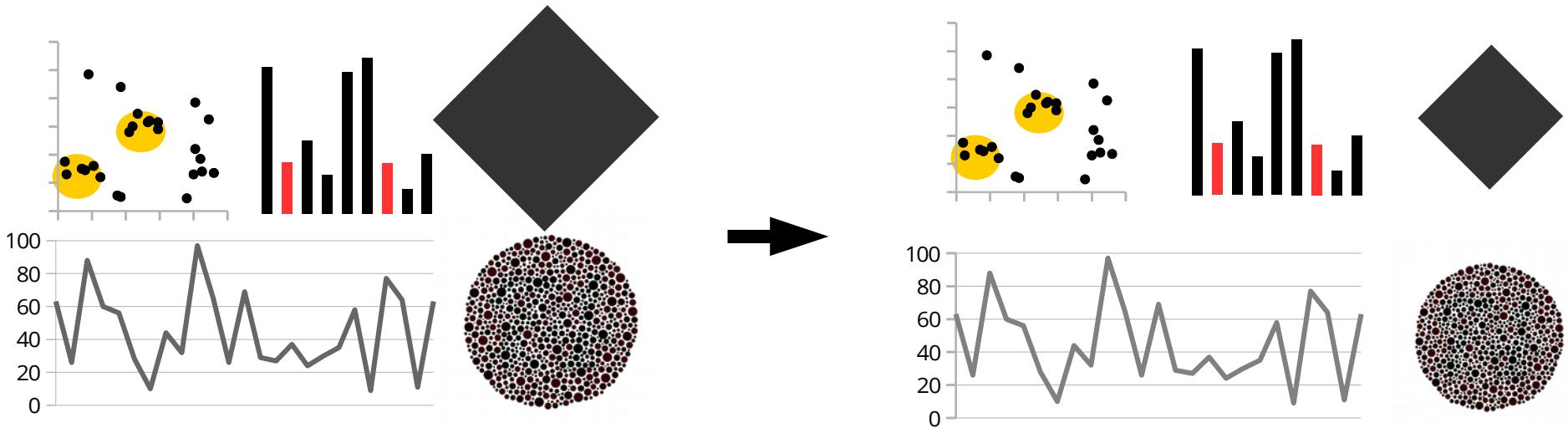
## **Visual weight/ emphasis:**

- How much an object on the page attracts and retains the attention of your viewer
- Depends on size, colour, position, etc.
- Should match the relevance of the information

These are some questions you can make to assess visual balance and flow: *Is there a clear (and justified) hierarchy or arrangement between elements? Can adjustments be made to make more relevant connections? Does the place feel cluttered/scattered?* (Krause, 2004)

# Visual weight and balance

**Visual weight:** A measure of how much an object on the page attracts and retains the attention of your viewer



In the left figure, the black diamond and, to a lesser extent, the circle stand out (*is this our intention?*).

There is also little separation between the charts, which makes the figure look cluttered.

A black and white photograph of a long, straight road stretching into the distance under a cloudy sky. The road is marked with dashed white lines and leads towards a flat horizon.

YOU FOCUSED ON THIS FIRST

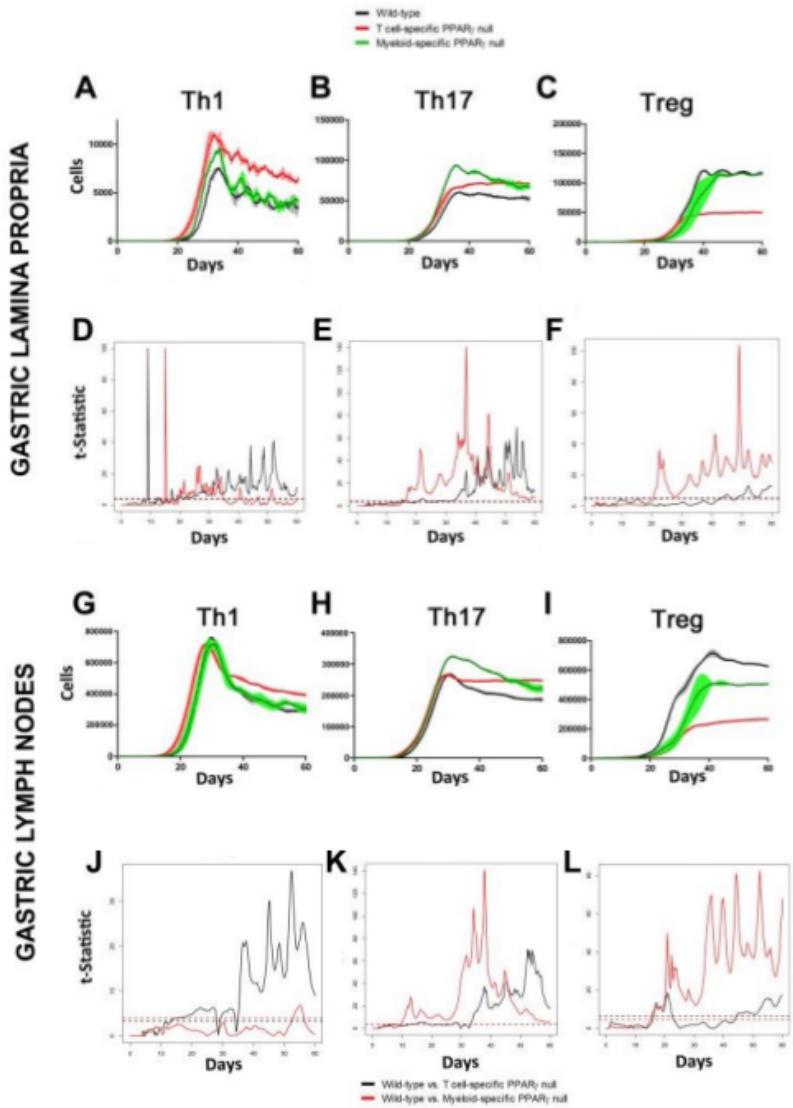
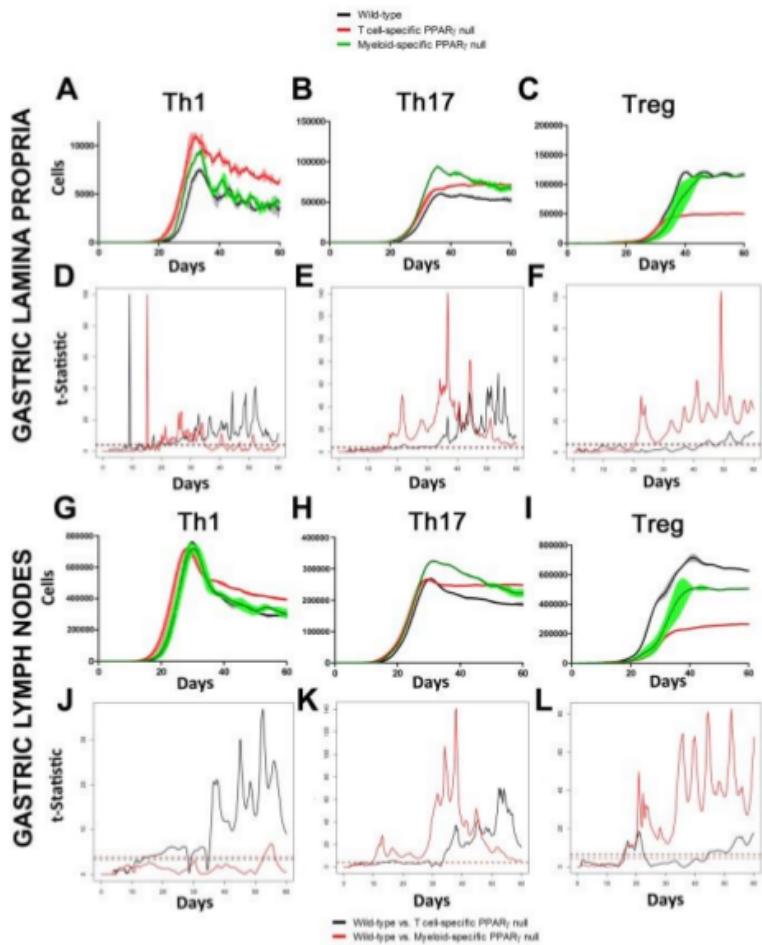
YOU FOCUSED ON THIS SECOND

# Visual weight and balance

**Visual weight:** A measure of how much an object on the page attracts and retains the attention of your viewer

Can help to guide the viewers eye through the figure

# Use of white space



# General tips

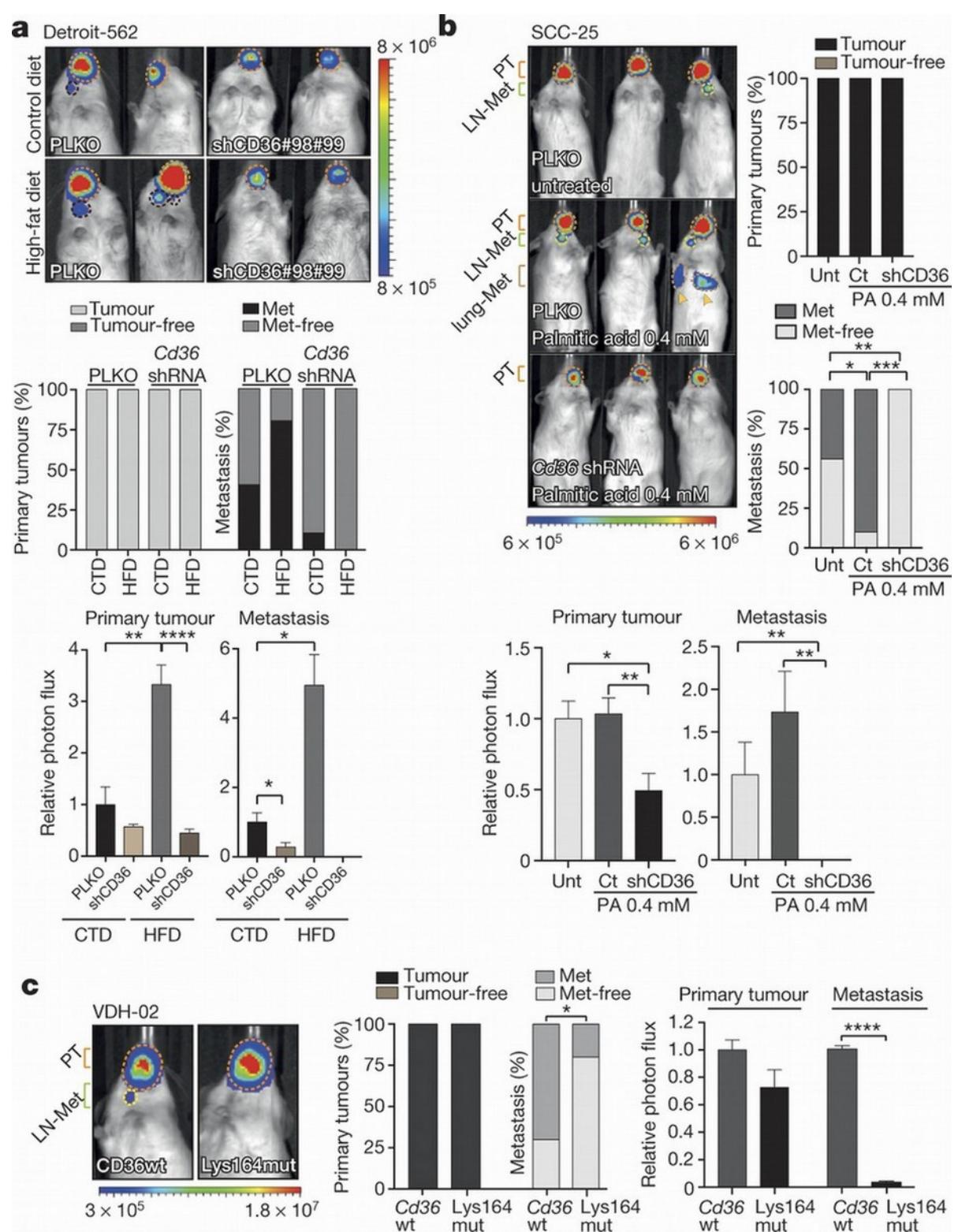
## Don't-s:

- Don't distort the data
- No unnecessary figures or elements: *do we really need a figure? or a table would suffice?*
- Don't rely absolutely on colour
- No 3D: in most cases it distorts perception

## Do-s:

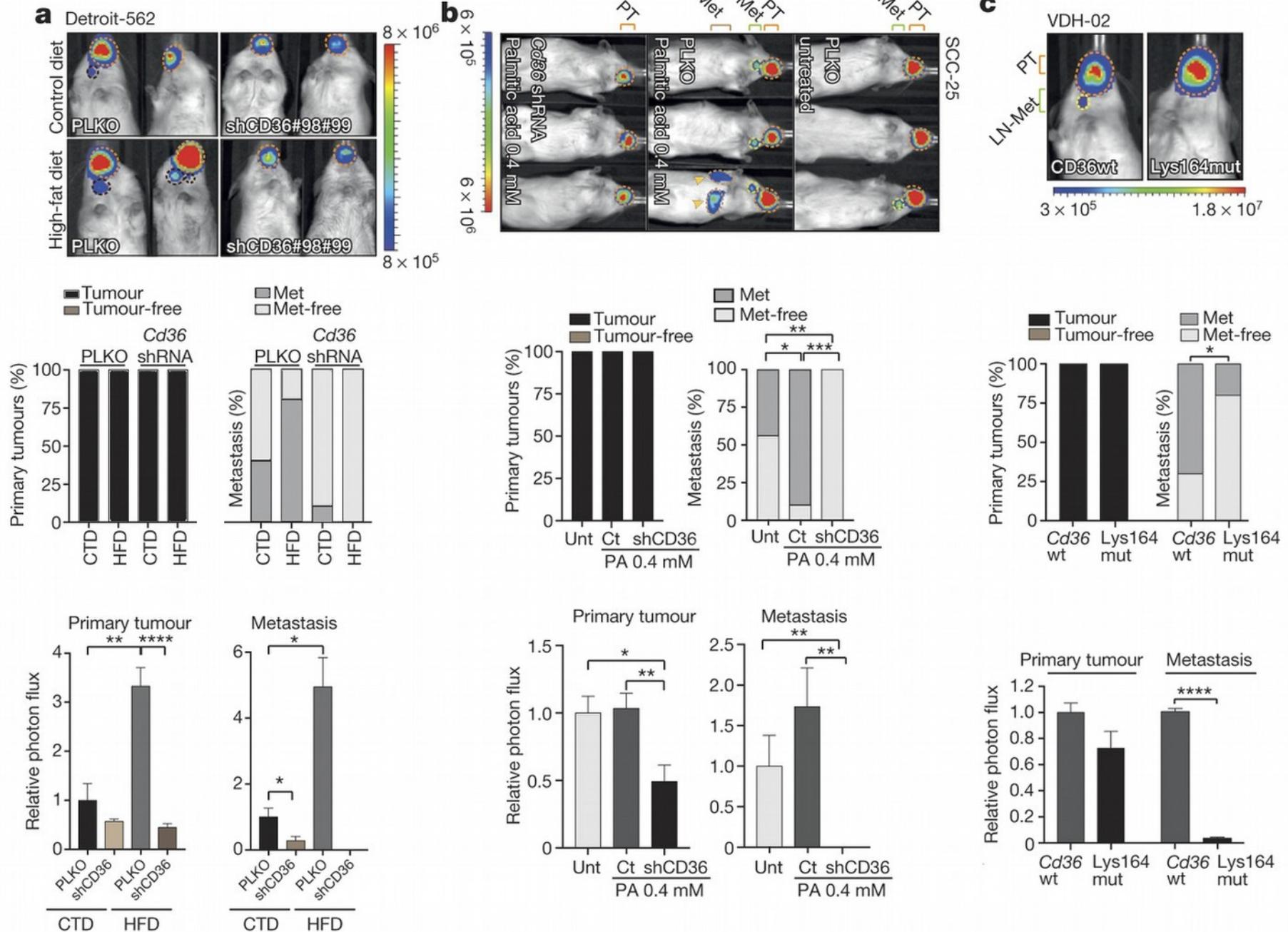
- One point per figure
- Summarise to clarify
- Have a clear purpose/ message
- Link to accompanying text and statistics

# Can you find ten ways to improve this figure?

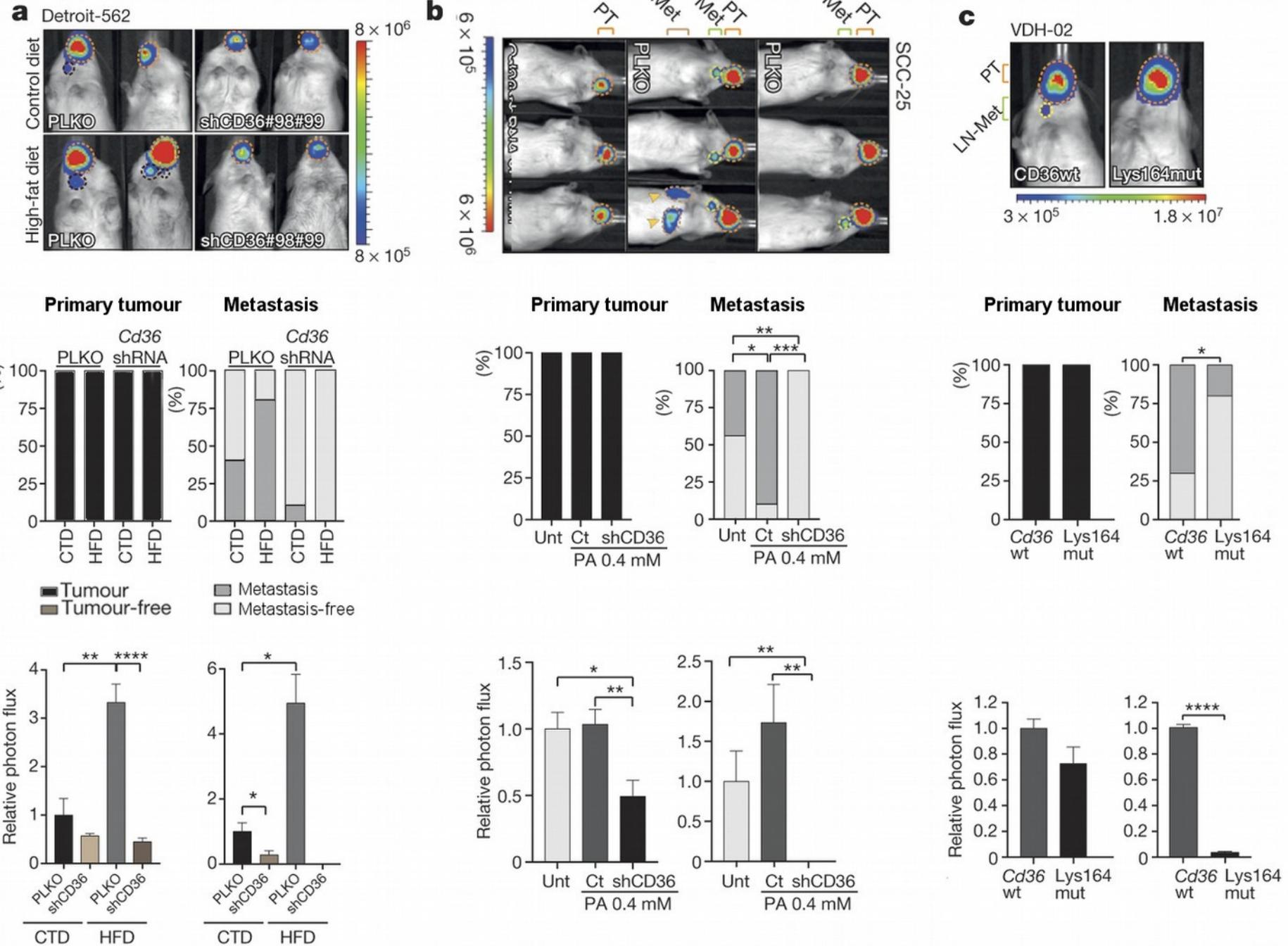


Pascual et al. Targeting metastasis-initiating cells through the fatty acid receptor CD36.  
Nature. 7 December 2016

# Work in progress...



# Work in progress...

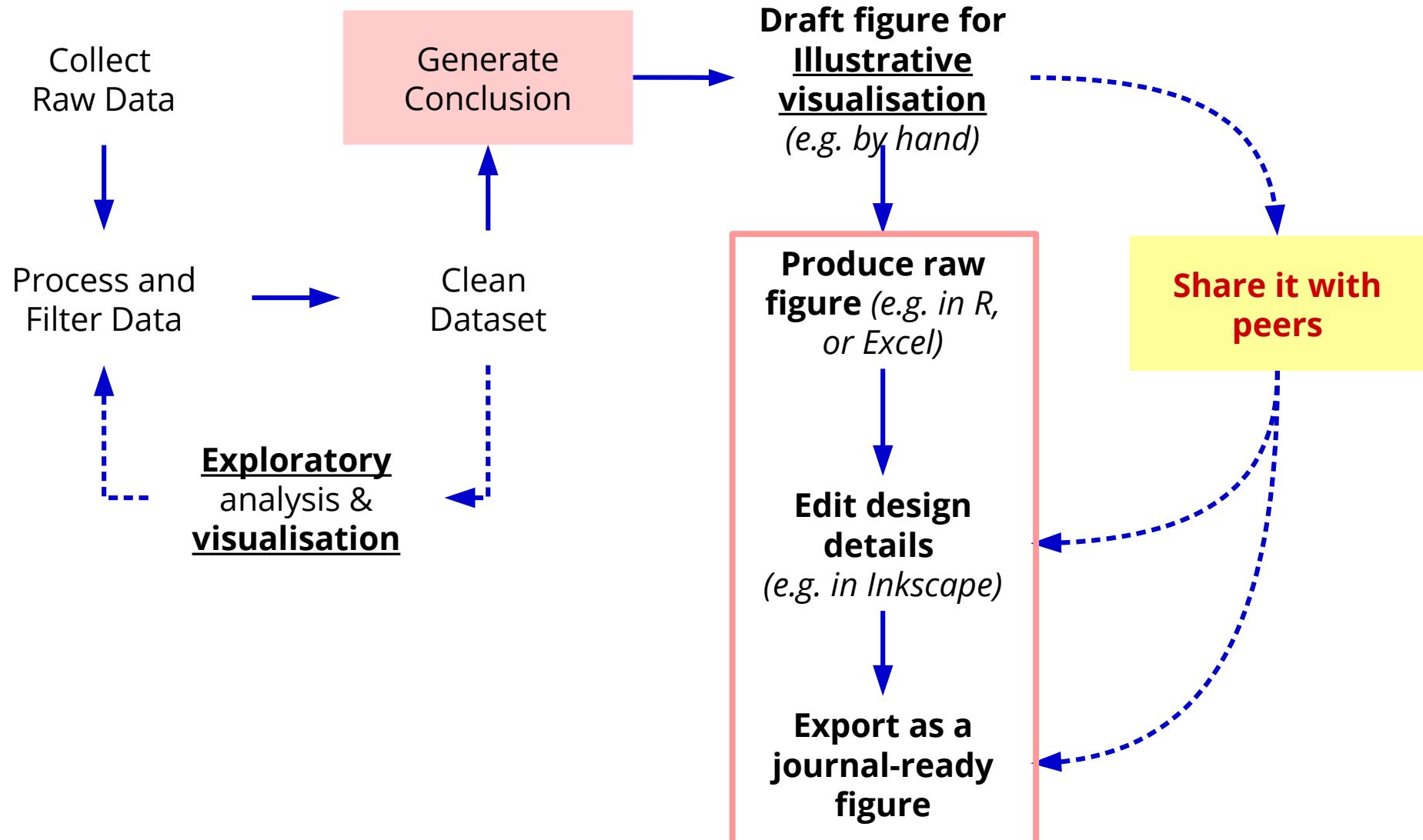


# Checklist

## Is your figure effective?

- The figure is **self contained**: understandable without additional information
- Every element is **labelled** or explained in the caption, including x and y units
- x and y axis: **scales** show appropriate variation of the data, or are comparable
- Readability** and **contrast** are appropriate
- Every use of **colour** has a reason
- The figure works in **grayscale** (except for very complex figures)
- If there are **groupings**, they help understand the message without manipulating
- There are no channel **inconsistencies** within the figure
- It is as **simple** as possible: i.e. no decorations, every piece that could be eliminated without losing information has been eliminated
- Has been **validated** with other people...

# Data Visualisation Process



# Validation

- Always try to validate plots you create
- You have seen your data too often to get an unbiased view
- Show the plot to someone not familiar with the data
  - What does this plot tell you?
  - Is this the message you wanted to convey?
  - If they pick multiple points, do they choose the most important one first?

# Not covered in this session

## Diagrams

- Definition
- Workflow:
  - Clarify the purpose: essential elements to depict and their relation
  - Draft the structure of the diagram by hand and share and discuss it
- Use grids and think carefully about the label choice and position
- Types: Venn diagrams (composition of datasets), flowcharts (for decision making processes), tree diagrams, timelines, networks, pathways, procedural diagrams
- Remember: the key *"is not the quality of the diagram or drawing, but the clarity of the information"* Carter p128

## Photos

- Avoid unethical manipulation (deleting noise, etc.), even if it doesn't change the results
- Crop to emphasize important bits
- Rule of thirds
- Use good quality images (sufficient resolution and colour/ brightness settings)
- Format differences: JPEG, TIFF, GIF, PNG
- Resolution
- Cropping and image composition
- Image size and proportions
- In context: contrast and relation with surrounding content
- Check license for use

# Some useful resources

- Short papers:
  - **Roland**i et al 2011. A Brief Guide to Designing Effective Figures for the Scientific Paper. *Advanced Materials* 23
  - **Rougier** et al 2014. Ten Simple Rules for Better Figures. *Plos Computational Biology* 10:9
- Design for scientists/ data:
  - **Carter**. 2013. Designing science presentations – *not just for figures, very clear*
  - **Munzner**. 2014. Visualization, analysis and design
    - *from a computer-graphics perspective*
  - **Tufte**. 2001. The visual display of quantitative information
    - *from a theory-of-design perspective*
  - **Meirelles**. 2013. Design for information
    - *advanced information visualizations (maps, time-space, flows)*
- Graphic design more generally:
  - **Krause**. 2004. Design basics index – *very concise and to the point*
  - **Samara**. 2014. Design elements: a graphic design manual – *reference book*
- *Nature Points of View*:  
<http://blogs.nature.com/methagora/2013/07/data-visualization-points-of-view.html>

If you need additional references, help or want to collaborate:  
[aiora.zabala@gmail.com](mailto:aiora.zabala@gmail.com), <http://aiorazabala.net/portfolio/>