Scientific Figure Design

v2024-05

Simon Andrews, Anne Segonds-Pichon, Boo Virk, Jo Montgomery

simon.andrews@babraham.ac.uk
jo.montgomery@babraham.ac.uk
The volume of the cells increased upon treatment

<table>
<thead>
<tr>
<th>Control</th>
<th>Treated</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8245</td>
<td>1.3232</td>
</tr>
<tr>
<td>1.0136</td>
<td>2.5644</td>
</tr>
<tr>
<td>1.3224</td>
<td>1.4899</td>
</tr>
<tr>
<td>1.0128</td>
<td>1.512</td>
</tr>
<tr>
<td>0.9644</td>
<td>2.6002</td>
</tr>
<tr>
<td>0.9668</td>
<td>2.1132</td>
</tr>
<tr>
<td>1.2296</td>
<td>13.228</td>
</tr>
<tr>
<td>1.0532</td>
<td>1.7566</td>
</tr>
</tbody>
</table>
The volume of the cells increased upon treatment

<table>
<thead>
<tr>
<th>Control</th>
<th>Treated</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8245</td>
<td>1.3232</td>
</tr>
<tr>
<td>1.0136</td>
<td>2.5644</td>
</tr>
<tr>
<td>1.3224</td>
<td>1.4899</td>
</tr>
<tr>
<td>1.0128</td>
<td>1.512</td>
</tr>
<tr>
<td>0.9644</td>
<td>2.6002</td>
</tr>
<tr>
<td>0.9668</td>
<td>2.1132</td>
</tr>
<tr>
<td>1.2296</td>
<td>1.3228</td>
</tr>
<tr>
<td>1.0532</td>
<td>1.7566</td>
</tr>
</tbody>
</table>
What this course covers...

• Theory of data visualisation
  – Why do some figures work better than others?
  – Applying theory to common plot types

• Ethical data representation

• Incorporating principles of graphic design

• Practical figure editing and compositing in Inkscape
What this course doesn’t cover...

- How to draw graphs in specific programs

http://www.bioinformatics.babraham.ac.uk/training.html
Consider the requirements for a figure

1. Collect Raw Data
2. Process and Filter Data
3. Clean Dataset

- Exploratory Analysis
- Generate Conclusion

- Reference Figures
- Illustrative Figures
- Exploratory Figures
Exploratory figures

- Quick!
- Complete
- Interactive
Reference figures

• Complete
• Flexible
Illustrative figures

- Simple
- Easy to understand
- Well Designed
What makes a good figure?

• Has a clear purpose and message
  – Helps to tell a story
  – Adds to the text, and links to it

• Is focused
  – Don’t confuse one message with another

• Is easy to interpret correctly
  – Good data visualisation
  – Good design

• Is an honest and true reflection of the data
The theory of data visualisation

Simon Andrews, Phil Ewels

simon.andrews@babraham.ac.uk
<table>
<thead>
<tr>
<th>Sample A</th>
<th>Sample B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>12</td>
<td>144</td>
</tr>
</tbody>
</table>
Different representations have common elements
Marks and Channels

• Marks
  – Geometric primitives
    • Lines
    • Points
    • Areas
  – Used to represent data sets

• Channels
  – Graphical appearance of a mark
    • Colour
    • Length
    • Position
    • Angle
  – Used to encode data
Figures are a combination of marks and channels

**Mark = Circle segment**  
1 Channel = Angle

**Mark = Diamond shape**  
2 Channels = X position, Y position

**Mark = Circle**  
4 Channels:  
- X position  
- Y position  
- Area  
- Colour

**Mark = Rectangle**  
2 Channels = X Position, Length of longest side
Golden Rules

• Expressiveness
  – Match the properties of the data and channel

• Effectiveness
  – Encode the most important information with the most effective channel
Types of channel

• Quantitative
  – Position on scale
  – Length
  – Angle
  – Area
  – Colour (saturation)
  – Colour (lightness)

• Qualitative
  – Spatial Grouping
  – Colour (hue)
  – Shape

• Quantitative
  – Weight
  – Length
  – Height
  – Expression
  – Time
  – Density

• Qualitative
  – Treatment
  – Genotype
  – Batch
Golden Rules

• Expressiveness
  – Match the properties of the data and channel

• Effectiveness
  – Encode the most important information with the most effective channel
Matching the data and channel

- Cell adhesion
- Positive regulation of cell proliferation
- Positive regulation of cell migration
- Response to cAMP
- Melanin biosynthetic process
Representing Colour

• Only channel to appear in both Qualitative and Quantitative

• Technical representations of colour
  – Red + Green + Blue (RGB)
  – Cyan + Magenta + Yellow + Black (CMYK)

• Perceptual representation of colour
  – Hue + Saturation + Lightness (HSL)
HSL Representation

• Hue = Shade of colour = Qualitative
• Saturation = Amount of colour = Quantitative
• Lightness = Amount of white = Quantitative

• Humans have no innate quantitative perception of hue but we have learned some (cold – hot, rainbow etc.)

• Our perception of hue is not linear
Types of colour channel

• Quantitative
  – Colour (saturation)
  – Colour (lightness)

• Qualitative
  – Colour (hue)

In a single plot you should modify only ONE colour parameter
Golden Rules

• Expressiveness
  – Match the properties of the data and channel

• Effectiveness
  – Encode the most important information with the most effective channel
Effectiveness of quantitative channels

4.5X

1.8X

7X

2X

16X
Quantitation Perception

Steven’s Psychophysical Power Law: $S = I^N$
Golden Rules

• Effectiveness
  – Encode the most important information with the most effective channel

• Expressiveness
  – Match the properties of the data and channel
Most Quantitative Representations

Good quantitation

- Bar chart
- Stacked bar chart with common start
- Stacked bar chart with different starts
- Pie charts
- Bubble plots (circular area)
- Rectangular area
- Colour (luminance)
- Colour (saturation)

Poor quantitation
Effectiveness of Qualitative Channels

• If you encode categorical data are the differences between categories easy for the user to perceive correctly?
Colour Discrimination

• How many colours can you discriminate?
Colour Discrimination

• How many colours can you discriminate?
Colour Discrimination
Qualitative Discrimination

- How many (fillable) shapes can you discriminate?

- Can combine shape with colour, but you need to maintain similar fillable areas
Qualitative Discrimination

- You can combine shape with colour, but you need to maintain similar fillable areas
Separability

Adding channels can adversely affect the effectiveness of existing channels

- Position + Hue (Color)
  - Fully separable
    - There is no confusion between the two channels

- Size + Hue (Color)
  - Some interference
    - Larger points are easier to discriminate than smaller ones

- Width + Height
  - Some/significant interference
    - We tend to focus on the area of the shape rather than the height/width separately

- Red + Green
  - Major interference
    - Humans are very bad at separating combined colours
Separability

Higher rates on Go trials

Higher rates on NoGo trials

$r = 0.69$, $p = 3.96 \times 10^{-7}$
Other visual cues

How can you modify your plot to improve its ease of interpretation, without changing the basic data representation?
Pop-out

• Sometimes you want to draw people's attention to parts of the plot

• We can use colours or shapes to trigger a 'popout' reaction

• An implicit rather than explicit cue
Popout
(find the red circle)
Popout

Speed of identification is independent of the number of distracting points
Popout

Colour pops out more than shape
Popout
Mixing channels removes the effect
(Find the red circle)
Other visual clues
Grouping
Other visual clues

• Is a monkey heavier than a dog?
Other visual clues

• Is a monkey heavier than animal X?
Containment / Linking

Wild Type

Mutant
Containment / Linking
How do you know if your figure is working?
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?
Exercise

You will be given a series of (not very good) plots to validate. Try to think what message the plot is trying to convey and whether it is doing so effectively.

Work out how you would choose to represent the data if you don’t like the way it’s presented now.
Making effective use of common plot types

Anne Segonds-Pichon
Simon Andrews
Phil Ewels

simon.andrews@babraham.ac.uk
Types of plot
Things you can illustrate

- Distribution
- Comparison
- Relationship
- Composition
Distributions
Representing Distributions
Single Samples

Histograms

Density Plots
Representing Distributions
Single Samples - Bandwidth
Representing Distributions
Single Samples – Discontinuous data

Plotting Integer Data

1.5

1.8

2
Representing Distributions
Multiple Samples
Comparisons
Comparisons
Error Bars

- Standard Error of Mean (SEM)
- How accurately is the mean calculated
- Gets smaller with increased data
- Good when comparing means

- Standard Deviation (SD)
- How well does the mean summarise the data
- No systematic change with increased data
- Good when comparing variability
Setting a suitable baseline
Relationships
Relationships – Line Graphs
Relationships - Scatterplots
Composition
Pie Charts

Total = 62

Total = 62
Stacked Bar Charts
Heatmaps
Making Heatmaps Effective

- Cluster rows and columns
- Median centre rows
- Diverging symmetrical colour scheme (colourblind friendly)
- Clear annotation
Ethics of data representation

Simon Andrews, Anne Segonds-Pichon

simon.andrews@babraham.ac.uk
What is an Ethical data visualisation?

• 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
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?
Advertising and politics are built on unethical data representation.

https://venngage.com/blog/misleading-graphs/
Not exploring the data well enough

One experiment, multiple measures, two conditions.
Not exploring the data well enough

Five experiments, single measures, control plus 3 treatments
Choosing the wrong axis/scale

Salaries offered vs date
Choosing the y-axis/scale

Inappropriate use of a log scale
Choosing the y-axis/scale

- **Logarithmic axis** should only be used for:
  - Lognormally spaced values
  - Lognormal data
Image Manipulation

• ‘Playing’ too much with contrast

“Adjusting the contrast/brightness of a digital image is common practice and is not considered improper if the adjustment is applied to the whole image.

Adjusting the contrast/brightness of only part of an image is improper, however, and this practice can usually be spotted by someone scrutinizing a file.”
Image Manipulation can be detected

https://forbetterscience.com
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?
Practical Design Theory

Boo Virk
Simon Andrews

simon.andrews@babraham.ac.uk
Why does good design matter?

• Good design makes a great first impression

• Good design makes for effective communication

• Good design keeps the reader engaged
Planning

• Always look at the guidelines for the journal you're submitting to
  – https://www.sciencemag.org/authors/instructions-preparing-initial-manuscript
  – https://www.nature.com/nature/for-authors/formatting-guide
  – https://www.cell.com/figureguidelines

• Huge variation in the amount of detail they provide

• Getting things right from the start saves huge amounts of time
General Figure Guidelines

• Use distinct colors with comparable visibility and consider colorblind individuals by avoiding the use of red and green for contrast. Recoloring primary data, such as fluorescence images, to color-safe combinations such as green and magenta, turquoise and red, yellow and blue or other accessible color palettes is strongly encouraged. Use of the rainbow color scale should be avoided.

• Use solid color for filling objects and avoid hatch patterns.

• Avoid background shading.

• Figures divided into parts should be labeled with a lower-case, boldface 'a', 'b', etc in the top left-hand corner. Labeling of axes, keys and so on should be in 'sentence case' (first word capitalized only) with no full stop. Units must have a space between the number and the unit, and follow the nomenclature common to your field.

• Commas should be used to separate thousands.

• Unusual units or abbreviations should be spelled out in full, or defined in the legend.

https://mts-ncomms.nature.com/cgi-bin/main.plex?form_type=display_auth_instructions
Plan out your panels

• Plan your panels before starting to draw final figures

• Plan to be consistent
  – Multiple figures of the same type
  – Common colour/shape schemes
  – Common fonts and sizing
  – Common abbreviations and units
  – Common naming of samples / conditions
Fig. 2: Caspase-1, caspase-11 and RIPK3 promote lethality in Casp8<sup>G362A/C362A</sup>Mkl<sup>−/−</sup> mice.

Kaplan–Meier curves of mouse survival. P values were calculated by two-sided Gehan–Breslow–Wilcoxon test. The number of mice differs from the list in Table 1, as some of the mice in the graph had a Casp8<sup>G362A/C362A</sup> parent.

Source data.
Alignment: We are sensitive to aligned edges, even when they are separated.
Use a grid to help align disparate parts of a figure
Don't make figures too crowded
Don't make figures too crowded
Don't cram too much information onto one figure
Don’t invent your own colour schemes
If possible try to consider colour blind readers

Affects 1:12 men and 1:200 women worldwide

“If a submitted manuscript happens to go to three male reviewers of Northern European descent, the chance that at least one will be colour blind is 22 percent.”
See how well your figure works for colour blind people

- Gradients are easy to change
- Categorical colours are very limited
- Basic interpretability in black and white is ideal

https://www.color-blindness.com/coblis-color-blindness-simulator/
Try to consider colour blind readers.
Only use plain colours as fills

- Use a standard colour scheme
- Optimise for colour blind people if possible
- Keep colours plain
When overlaying information, make sure you have sufficient contrast

- Poor contrast
- Good contrast

- Poor contrast
- Good contrast
Add overlays to increase contrast

Poor contrast

Good contrast
Keep text and fonts simple

• All fonts for figures should use sans serif fonts

  sans-serif    serif

• All text in figures should be black or white*

  Wild type  Wild type
  Knockout   Knockout

* Some journals insist on coloured text. They're wrong, but you can't fight the system
Contrast and text
Keep text horizontal

- Numbers are small, text is big
- All graphs still work when rotated 90°
Labelling and annotation

- Each axis is labelled
- Axis scales are appropriate
- Quantitative axes have units
- Colour scheme is explained
- Point shapes are explained

You need enough annotation that the figure is understandable on its own.
Labelling and annotation
Make sure all text is legible
at the final printed size

6 point font is the smallest you can comfortably read
(just over 2mm height on paper)
When resizing be aware of what can and cannot have its aspect ratio changed

• Things that always need to maintain their aspect ratios:
  – Images
  – Text
  – Circular objects
  – Axes with comparable units
Checklist

• Consistent use of
  – Figure types
  – Colours / Shapes
  – Fonts and Sizes
  – Names

• Colour
  – Uses a standard scheme
  – Colourblind friendly (if possible)

• All figures are correctly annotated
  – Axes labelled with names and units
  – Colours and Shapes explained

• Text
  – Sans serif font
  – Large enough to be legible
  – Ideally in black or white
  – Sufficient contrast to be legible