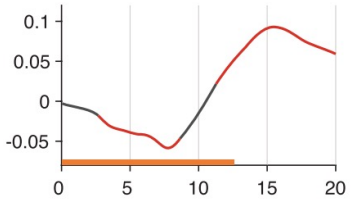
Data visualization - Tips for data visualization

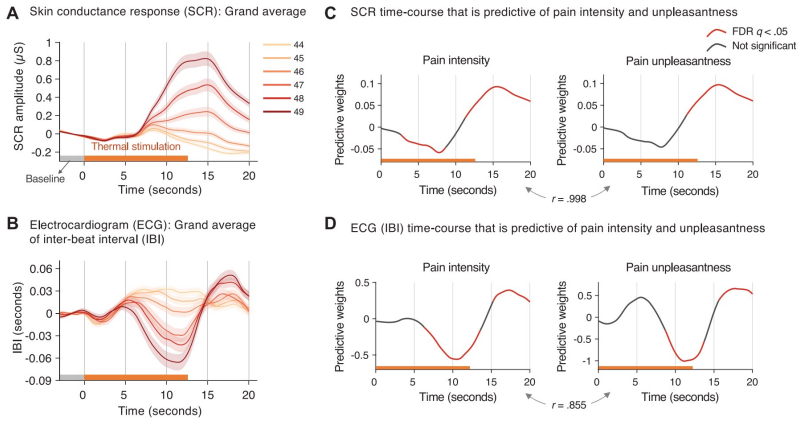
Choong-Wan Woo
Director of the Cocoan Lab

Different stages of data visualization for scientific publication

One plot



A collection of plots (a figure for publication)



Early phase of a project

Late phase

Different tips for different phases

Data visualization - Tips for data visualization (1)

Tips for drawing one plot

Choong-Wan Woo
Director of the Cocoan Lab

Tips for drawing one plot: Basic ingredients

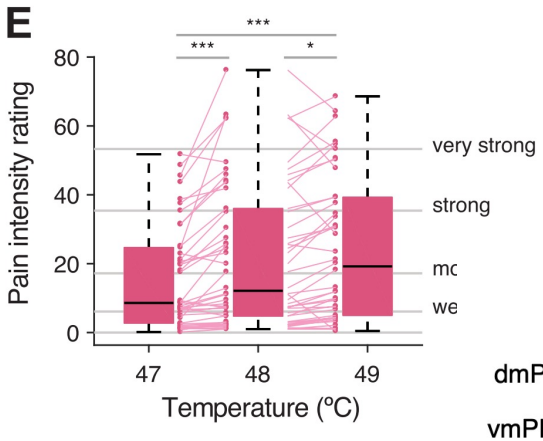
Know the **basic** ingredients of plots first: **Dots, lines, and Areas**

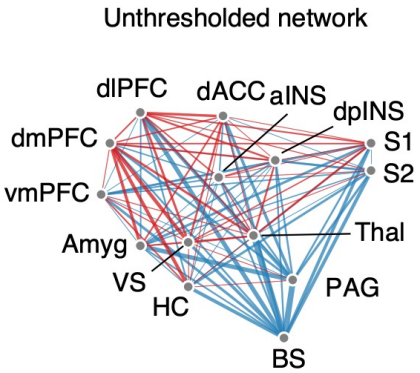
For dots, scatterplot

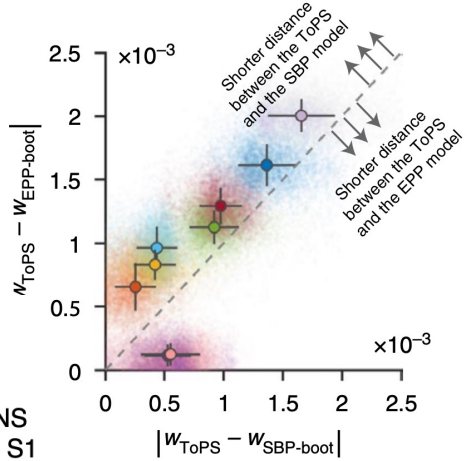
For lines, line plot

For Areas, bar and box plots

If you know how to draw dots, lines, and boxes in a software (e.g., matlab, R, python etc.), you can draw any types of plots







Figs from Matthewson, Woo, 2019; Lee et al., 2021

Tips for drawing one plot: Efficient use of ink and space

Be effective and efficient

$$data/ink\ ratio = \frac{amount\ of\ ink\ used\ on\ data}{total\ amount\ of\ ink}$$

Simpler is better for the most of cases

but show raw information as much as possible

(Tension between raw and processed)

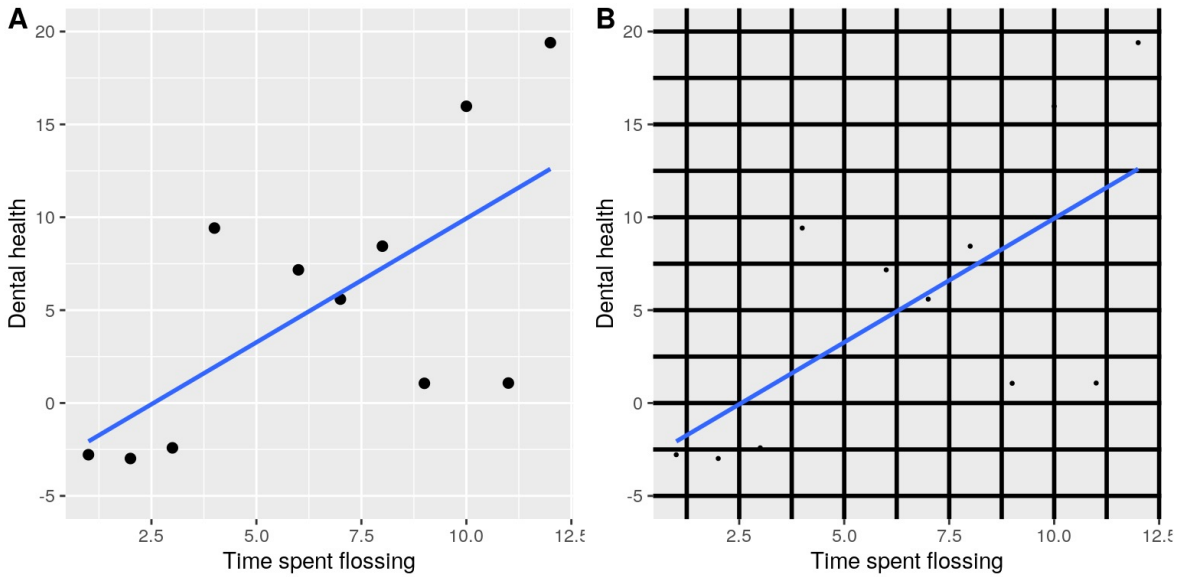


Figure 4.5: An example of the same data plotted with two different data/ink ratios.

From Poldrack, 2021, Statistical Thinking for the 21st Century

Tips for drawing one plot: Avoid chartjunk

Be effective and efficient

$$data/ink\ ratio = \frac{amount\ of\ ink\ used\ on\ data}{total\ amount\ of\ ink}$$

Avoid chartjunk

Created using Excel

Figure 4.6: An example of chart junk.

From Poldrack, 2021, Statistical Thinking for the 21st Century

Tips for drawing one plot: Do not distort the data (1)

Do not distort the data

The distortion could happen in both positive (e.g., exaggerating) and negative (e.g., obscuring) directions

Try to be fair to all sides, to data, researchers, and readers!

Include zero?

"In general, my inclination for line plots and scatterplots is to use all of the space in the graph, unless the zero point is truly important to highlight."

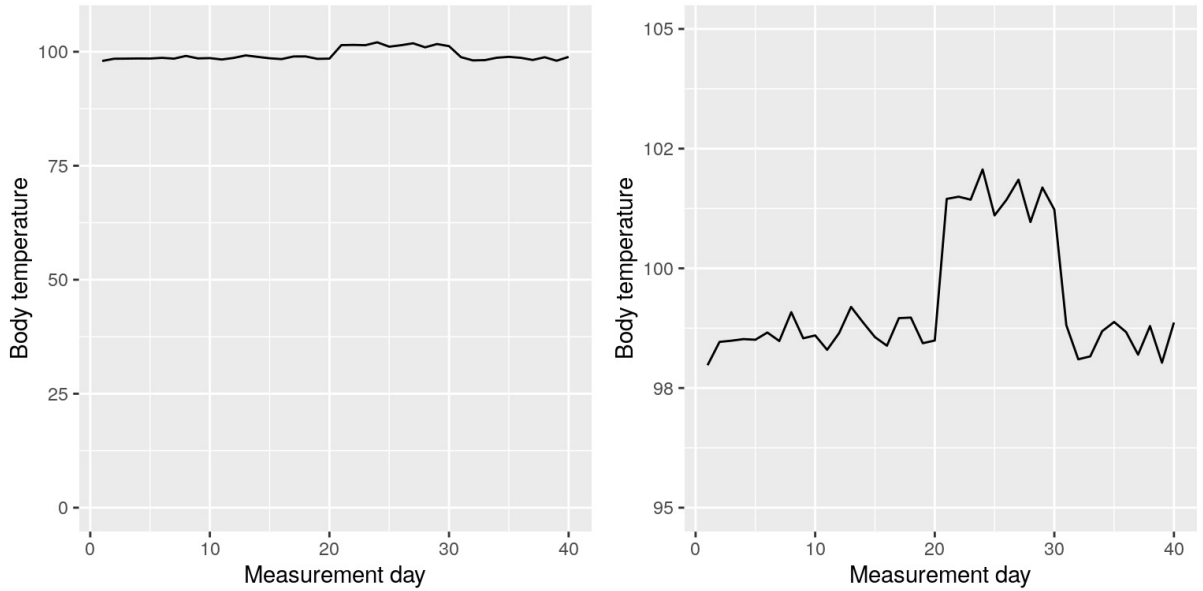


Figure 4.8: Body temperature over time, plotted with or without the zero point in the Y axis.

From Poldrack, 2021, Statistical Thinking for the 21st Century

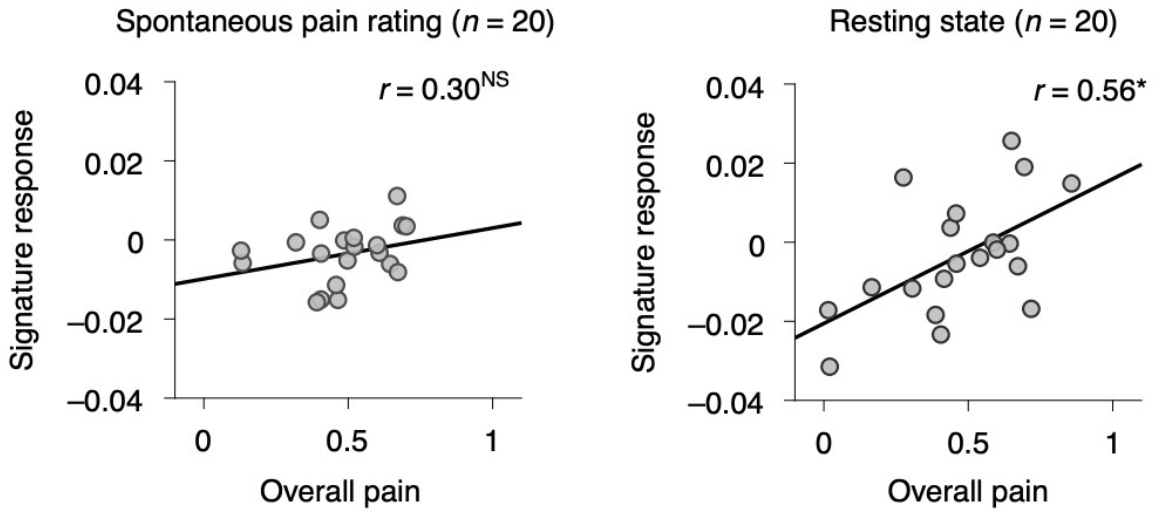
Tips for drawing one plot: do not distort the data (2)

Do not distort the data

The distortion could happen in both positive (e.g., exaggerating) and negative (e.g., obscuring) directions

Try to be fair to all sides, to data, researchers, and readers!

Use the same scale
when you show two plots side by side!!!



From Poldrack, 2021, Statistical Thinking for the 21st Century

Tips for drawing one plot: do not distort the data (3)

Do not distort the data

The distortion could happen in both positive (e.g., exaggerating) and negative (e.g., obscuring) directions

Try to be fair to all sides, to data, researchers, and readers!

The lie factor:

The degree to which physical differences in a visualization correspond to the magnitude of the differences in the data

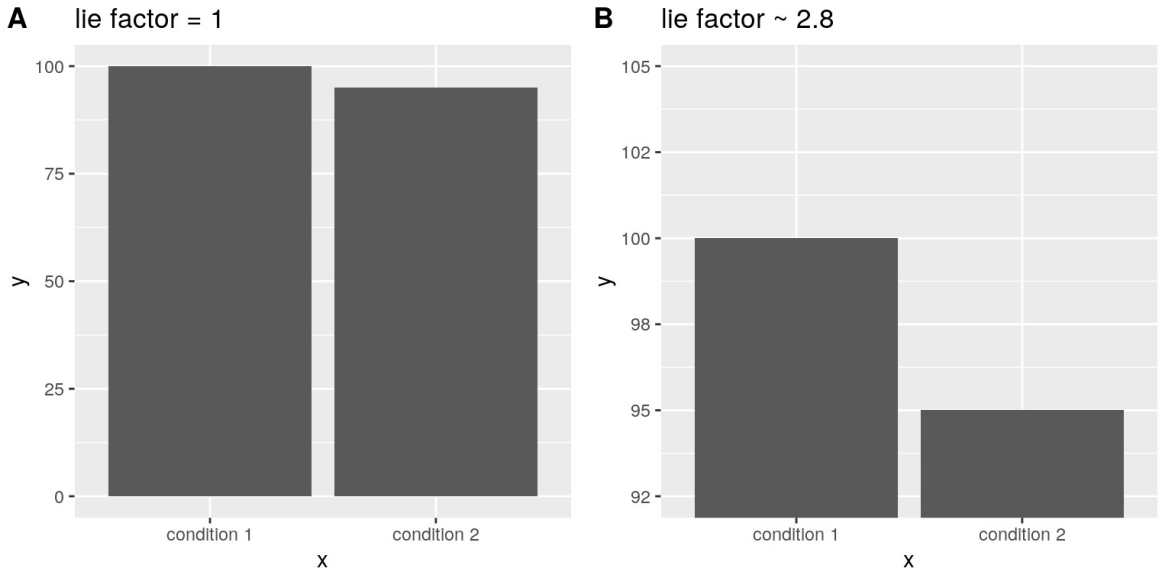


Figure 4.9: Two bar charts with associated lie factors.

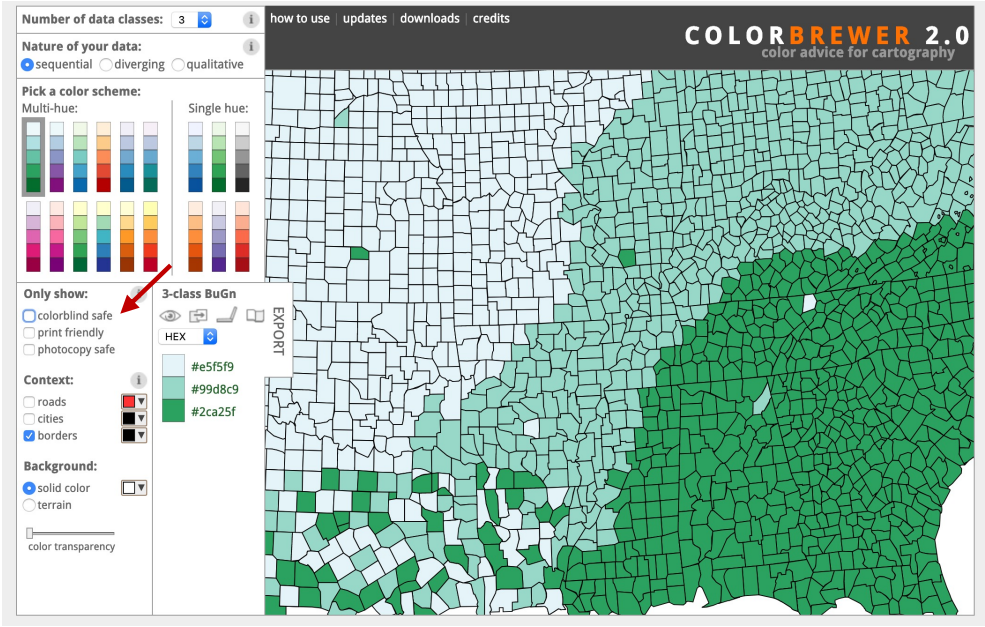
From Poldrack, 2021, Statistical Thinking for the 21st Century

Tips for drawing one plot: colorblind-friendly colors

https://github.com/cocoanlab/cocoanCORE/blob/master/Visualization/viz_data/colormap_blind_wani.mat

Accommodating human limitations

Use colorblind-friendly palett



Colorblind-friendly colors

Color	Color name	RGB (1–255)	CMYK (%)	P	D
	Black	0, 0, 0	0, 0, 0, 100		
	Orange	230, 159, 0	0, 50, 100, 0		
	Sky blue	86, 180, 233	80, 0, 0, 0		
	Bluish green	0, 158, 115	97, 0, 75, 0		
	Yellow	240, 228, 66	10, 5, 90, 0		
	Blue	0, 114, 178	100, 50, 0, 0		
	Vermillion	213, 94, 0	0, 80, 100, 0		
	Reddish purple	204, 121, 167	10, 70, 0, 0		

Figure 2 | Colors optimized for color-blind individuals. P and D indicate simulated colors as seen by individuals with protanopia and deuteranopia, respectively.

Wong et al., 2011

Colorblind-friendly palette in ggplot2

http://www.cookbook-r.com/Graphs/Colors_(ggplot2)/#a-colorblind-friendly-palette

Colorbrewer2

http://colorbrewer2.org/

From Poldrack, 2021, Statistical Thinking for the 21st Century

Tips for drawing one plot: monochrome printer

Accommodating human limitations

If someone print out the paper with a monochrome printer

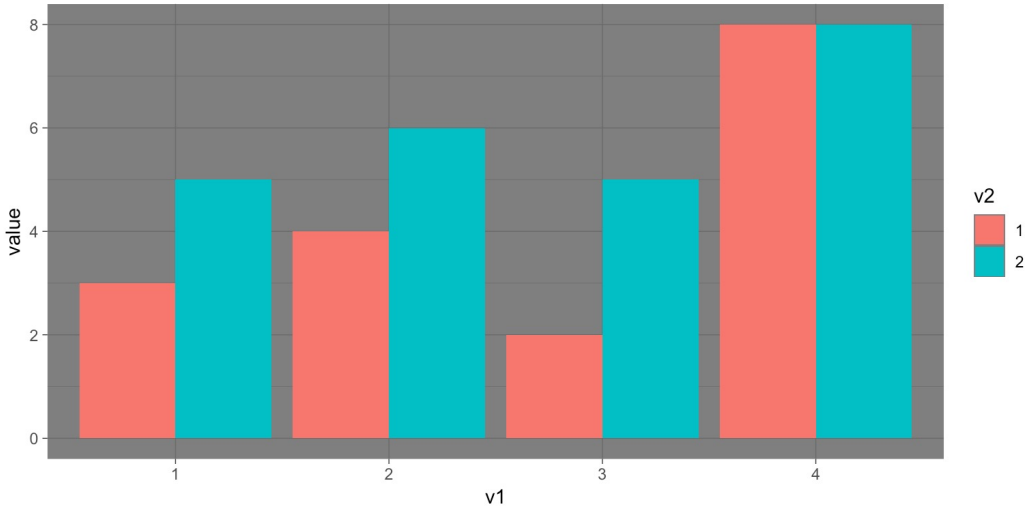


Figure 6.13: Example of a bad figure that relies solely on color contrast.

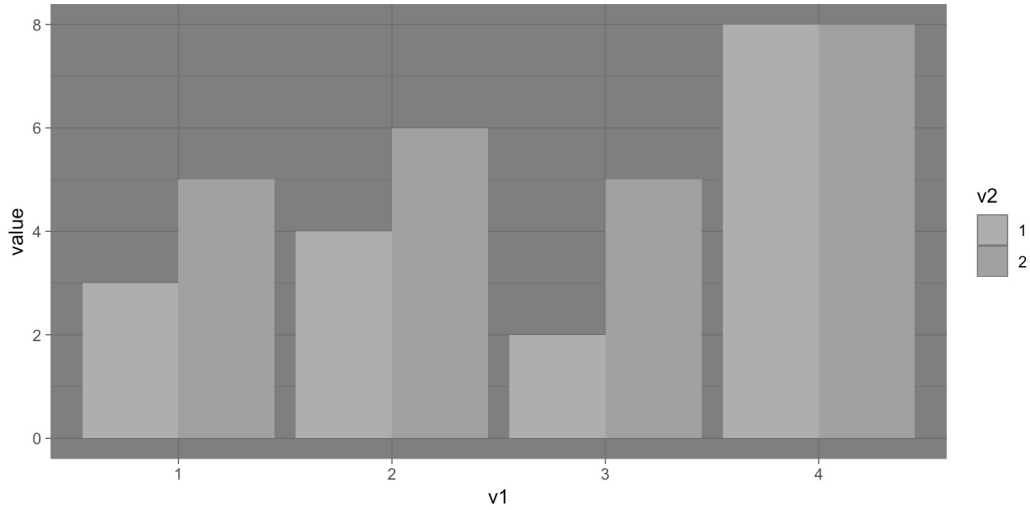
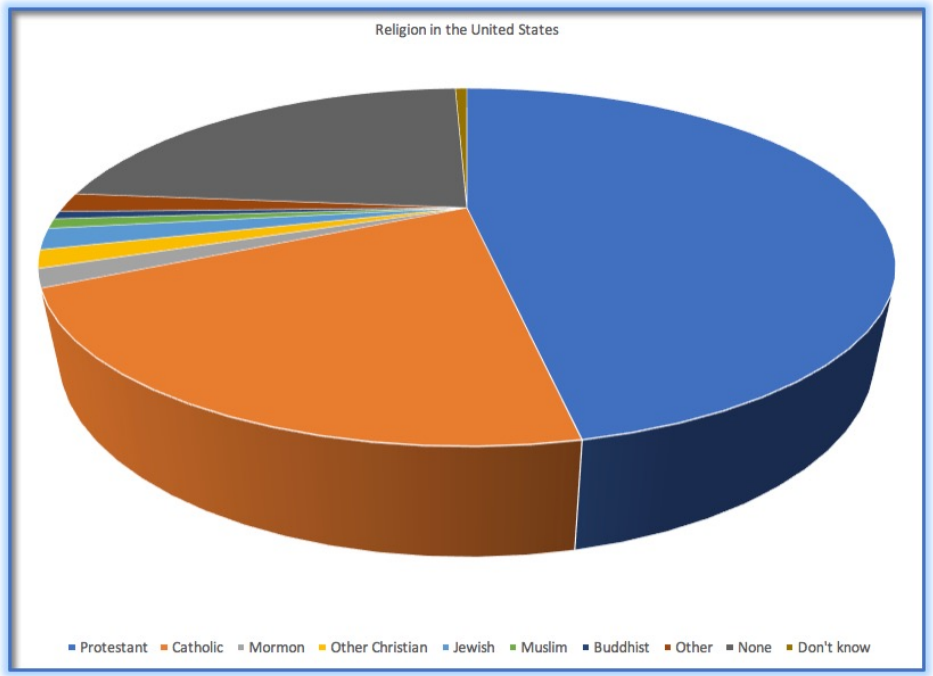


Figure 6.13: Example of a bad figure that relies solely on color contrast.

From Poldrack, 2021, Statistical Thinking for the 21st Century

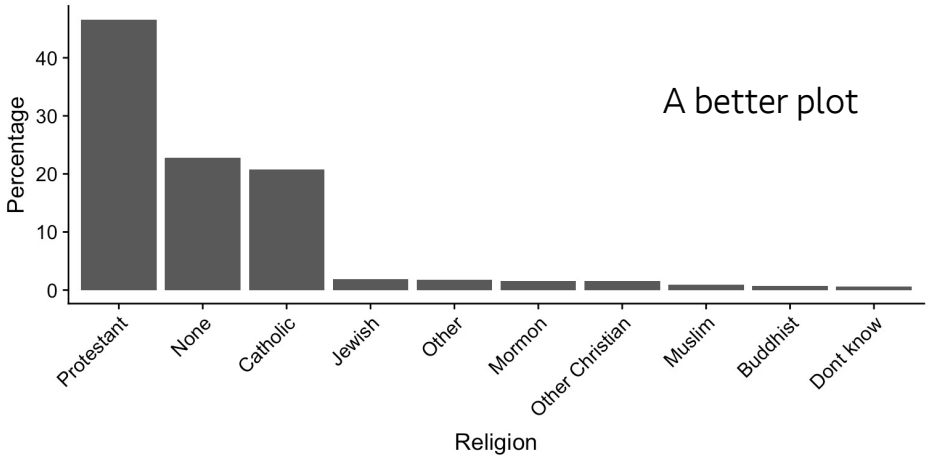
Tips for drawing one plot: area principle

Accommodating human limitations



Created using Excel

"Area principle": the area occupied by a part of the graph should correspond to the magnitude of the value it represents.



From Poldrack, 2021, Statistical Thinking for the 21st Century

Tips for drawing one plot: self-explanatory

Accommodating human limitations

Each plot should be self-explanatory (do not rely on readers' working memory capacity)

- Include legends
- Always, always, include titles of axes, subplots, etc. (even for a project meeting with me)
- Avoid acronyms if possible (but in some cases, using acronyms is useful)
- Assume a person who knows nothing about your project will read your plots

Data visualization - Tips for data visualization (2)

Tips for drawing a figure for publication

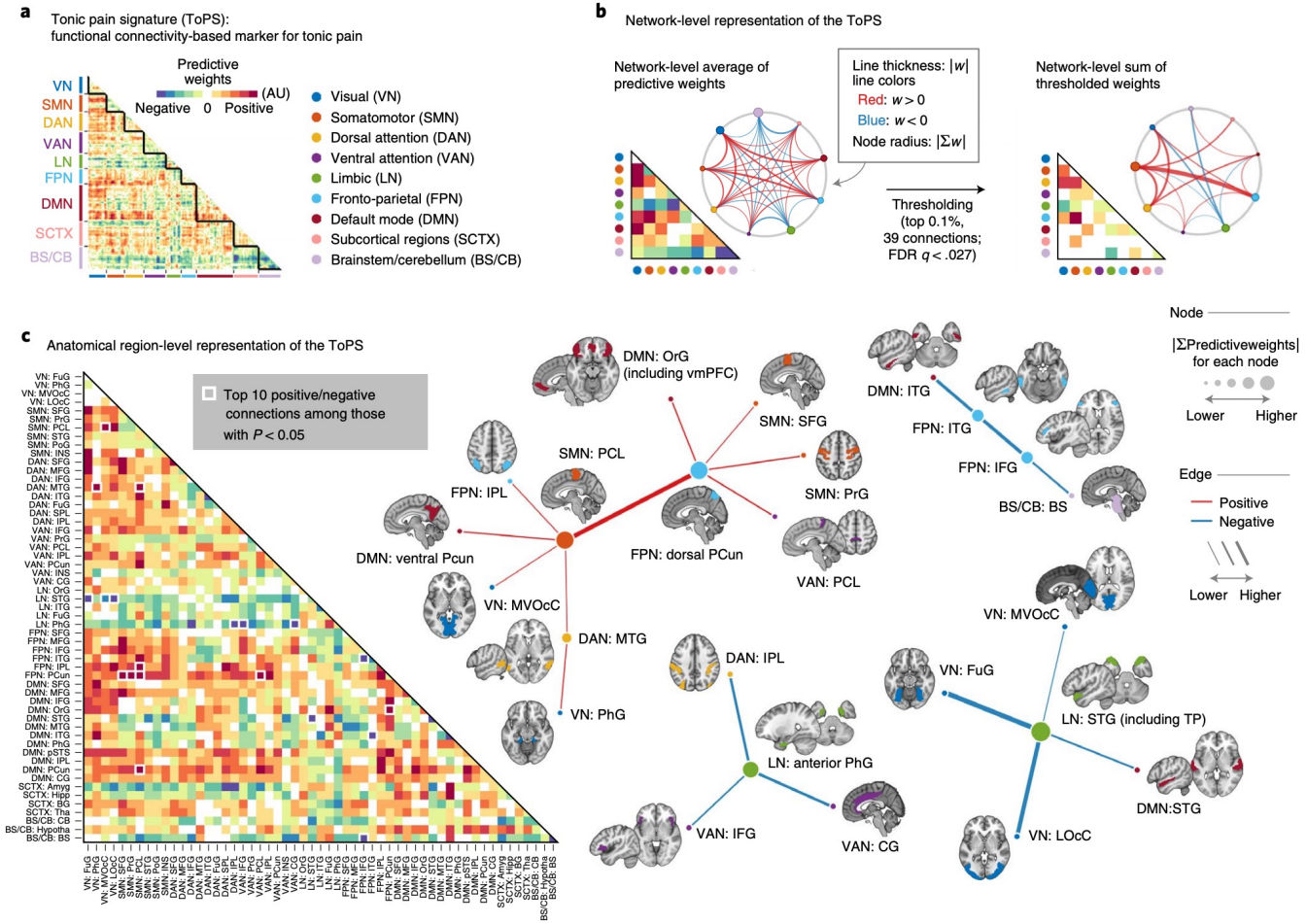
Choong-Wan Woo
Director of the Cocoan Lab

Tips for drawing a figure for publication: a collection of plots

Example:

A figure for publication

= A collection of plots



Tips for drawing a figure for publication: figures are important

Figures are important!!!

- to convey the message clearly and effectively
  (many readers will see only the abstract and figures of your paper)
- to provide a good impression to reviewers
  (reviewers = the first readers of the paper)

Tips for drawing a figure for publication: considerations

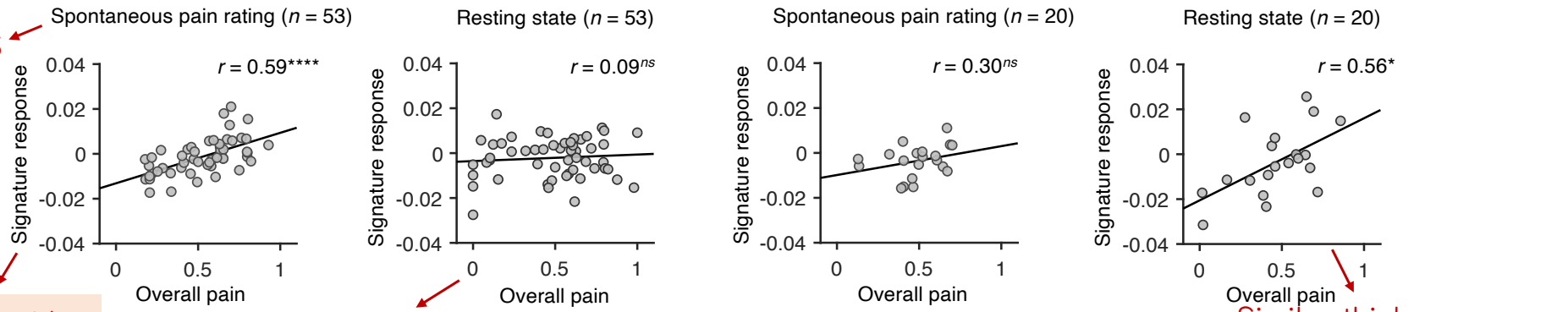
Important considerations:

- Effective use of the space
- Layout – Two-column figure? One-column figure? Etc. How would it look like in a letter size paper?
- "Relative" sizes are the key: Line width, font size (not too small)

Tips for drawing a figure for publication: a real example (1)

Title: font 9

subtitle: font 8

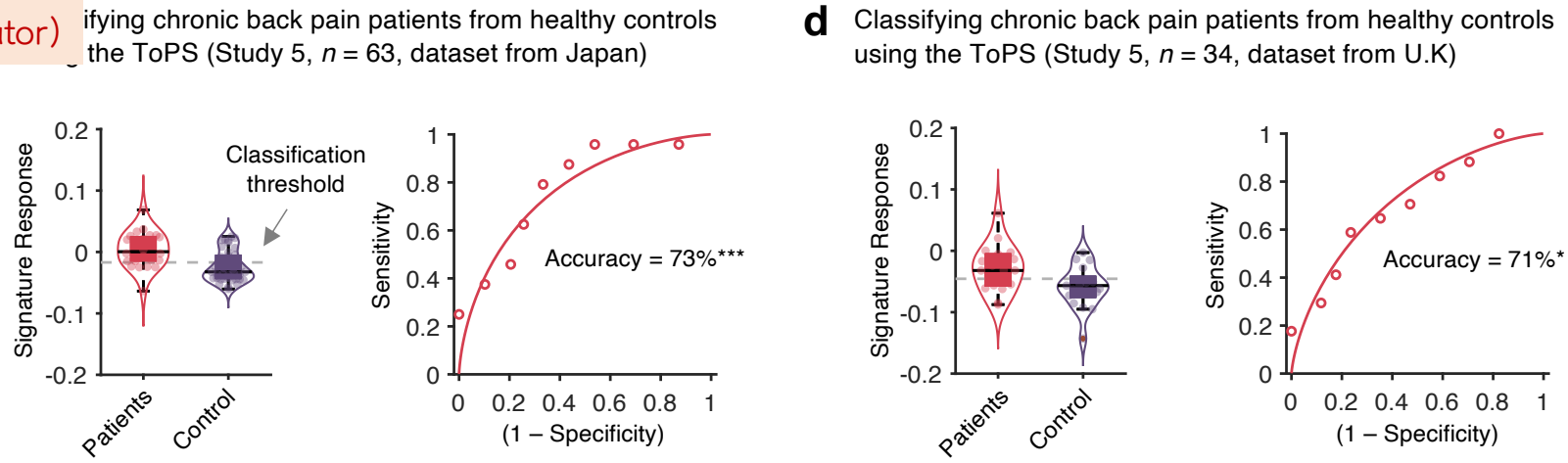


Make the axis title in figure program (e.g., ppt/illustrator)

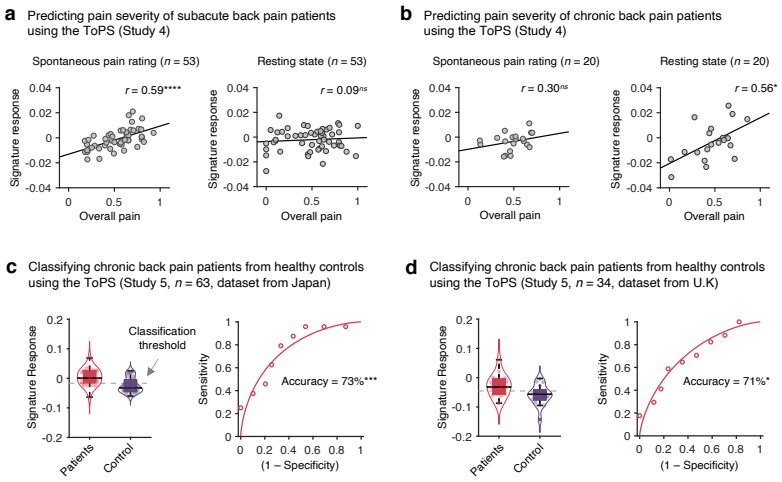
Tick label font ≈ 7 or 8

No smaller than font size 7

Similar thickness between lines and fonts



Tips for drawing a figure for publication: a real example (2)



The journal will edit the figures, and thus they usually require us to submit them as vector images

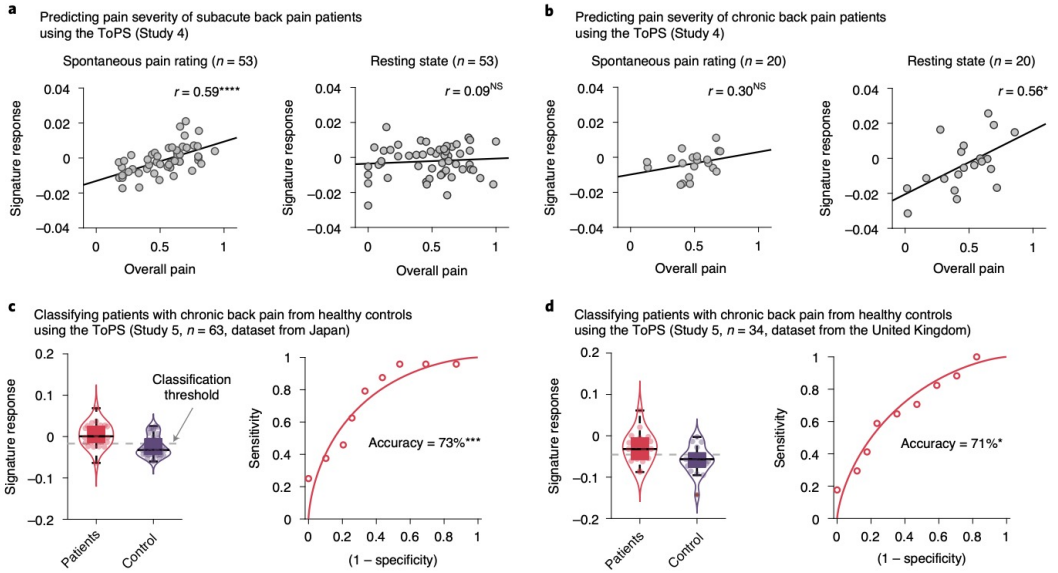


Fig. 3 | Testing the ToPS on the clinical pain data. a, **b**, We tested the ToPS on a publicly available clinical pain dataset[5,24–26] (Study 4) to evaluate how much the model can explain clinical pain severity of (**a**) patients with subacute back pain (SBP; $n = 53$) and (**b**) patients with chronic back pain (CBP; $n = 20$). The plots show the relationships between the actual pain scores (visual analog scale) versus signature response (arbitrary unit). Each dot represents an individual participant, and the line represents the regression line. The exact P values and degrees of freedom (d.f.) were (**a**) $P = 3.91 \times 10^{-6}$ (left) and 0.528 (right), d.f. = 51; (**b**) $P = 0.197$ (left) and 0.011 (right), d.f. = 18; two-tailed, one-sample t-test. **c**, **d**, We further tested the ToPS on two publicly available datasets[26] to evaluate how well the model can classify the patients with CBP from healthy control participants. One dataset (**c**) was obtained from Japan ($n = 63$), which included 24 patients and 39 healthy participants. The other dataset (**d**) was obtained from the United Kingdom ($n = 34$), which included 17 patients and 17 healthy participants. The exact P values were $P = 0.0003$ for **c** and $P = 0.024$ for **d**, two-tailed, binomial tests. $^{*}P < 0.05$, $^{***}P < 0.001$ and $^{****}P < 0.0001$. NS, not significant.

with CBP, differences in the correlation coefficients between the task types were not significant ($z = 0.93$ and $P = 0.353$), indicating that the task-dependent difference in prediction was minimal

accuracy with the optimal threshold was 71% ($P = 0.024$, binomial test), with 65% sensitivity, 76% specificity and AUC = 0.74.

Overall, the ToPS model 1) predicted variation in tonic pain over

Tips for drawing a figure for publication: make a vector image

Making a vector image

- Save the plot (in MatLab or R or Python, etc.) as pdf (or eps)
- Work on the figure in Powerpoint or illustrator (in our lab, we usually use Powerpoint)
- If you used illustrator, there is no problem.
- If you worked on the figures using Powerpoint, save the figure as PDF. And read the images into a software that can change it into a vector image. We usually use the Affinity Designer (that our iMac Pro has or it's really cheap (e.g., $24.99 when I checked now) and really good).
- Check the figures and export it as any formats required by the target journal.

Let's try together (in the next video)

Data visualization - Tips for data visualization (3)

Making a vector image

Choong-Wan Woo
Director of the Cocoan Lab

Data visualization - Tips for data visualization (4)

Tips for data visualization for early and late phases

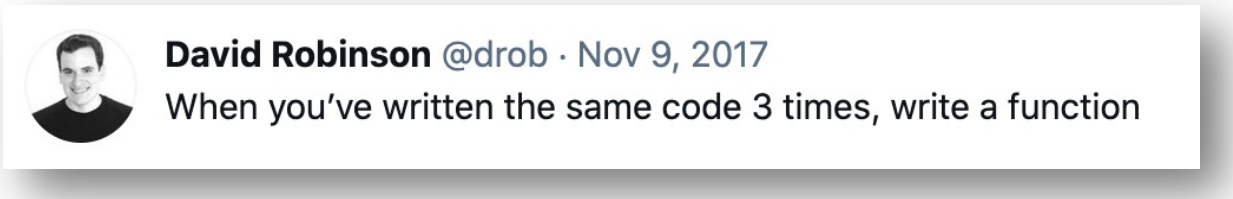
Choong-Wan Woo
Director of the Cocoan Lab

Tips for data visualization for early phases

In the early phase,

- Plot raw data (distribution) – histogram, scatter plot, violin plot, etc.
- Aesthetically pleasing visuals are not important at this stage
- But this doesn't mean that omitting tick labels, axis titles, legends is okay. The plots should always be self-explanatory.
- If you make ppts for the project meeting, you should save all those in slack or in the sync folder for the archive purpose. Oftentimes, you will want to go back and check what you did in the past. Then, having the figures in one place is really helpful.
- But what if you cannot read the plots??? It becomes fighting between present-you and past-you.
- Please please please make self-explanatory plots. And add the main take-home message (interpretation or observation of data analysis) in each slide if possible.

- Making functions is often helpful



Tips for data visualization for late phases

In the late phase

- The main goal of making figures is the story-telling
- Each figure should have a clear story to tell
- Ask yourself – "what's the point of this figure"

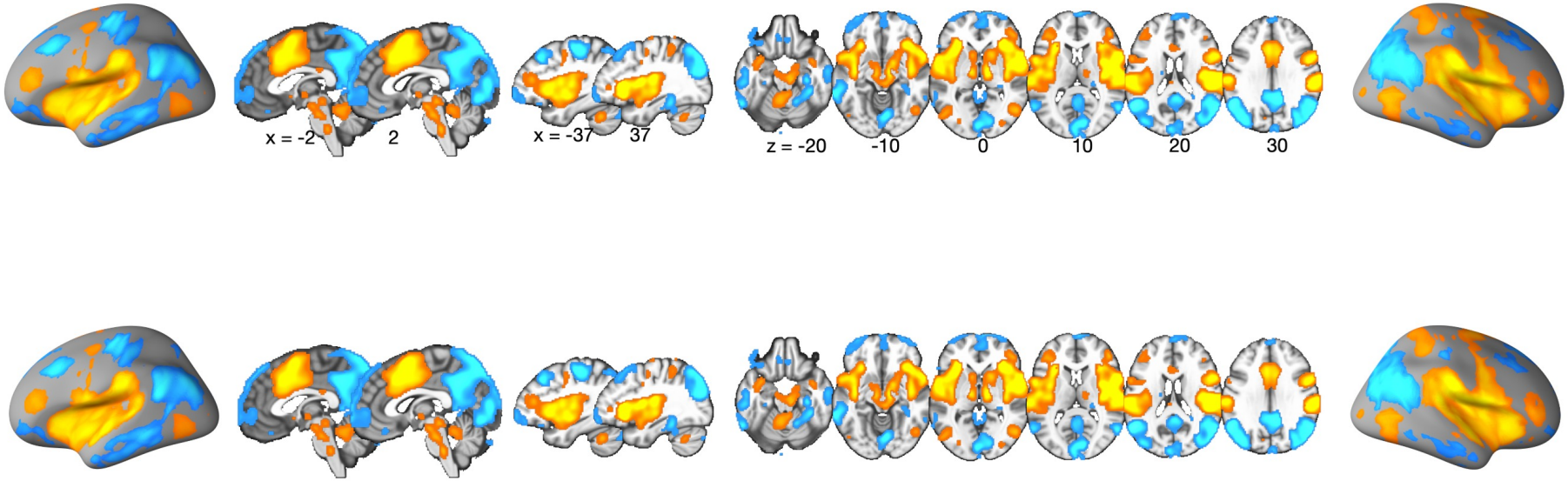
- As you are experienced more and more, you might want to save the time for making figures. Then keep thinking about the story flow early on and make nice figures.

Data visualization - Some useful functions (Matlab)

Choong-Wan Woo
Director of the Cocoan Lab

brain_activations_display.m

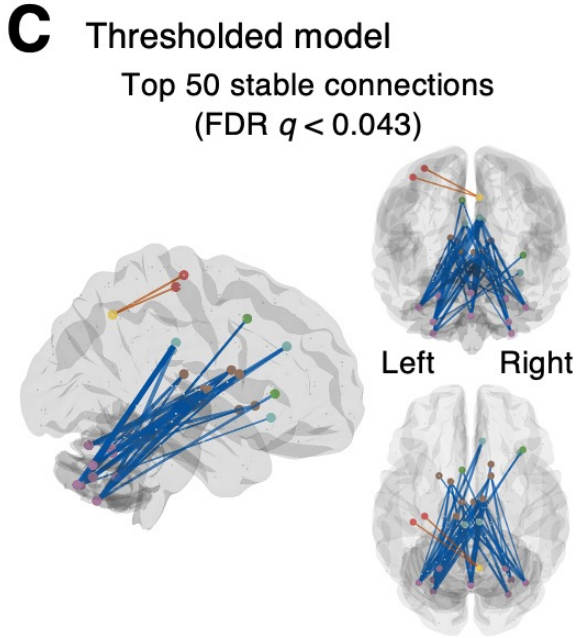
https://github.com/cocoanlab/cocoanCORE/blob/master/Visualization/brain_activations_display.m



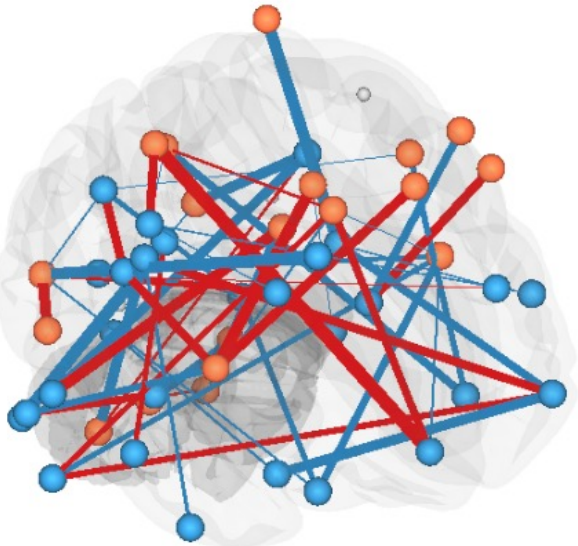
glass_brain_network.m

https://github.com/cocoanlab/cocoanCORE/blob/master/Visualization/glass_brain_network.m

Paper (Lee et al., submitted)



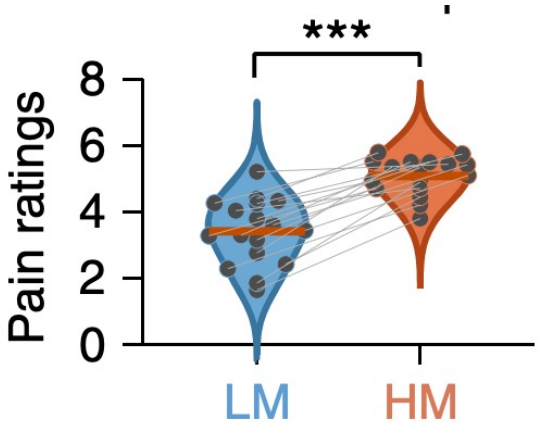
From our code



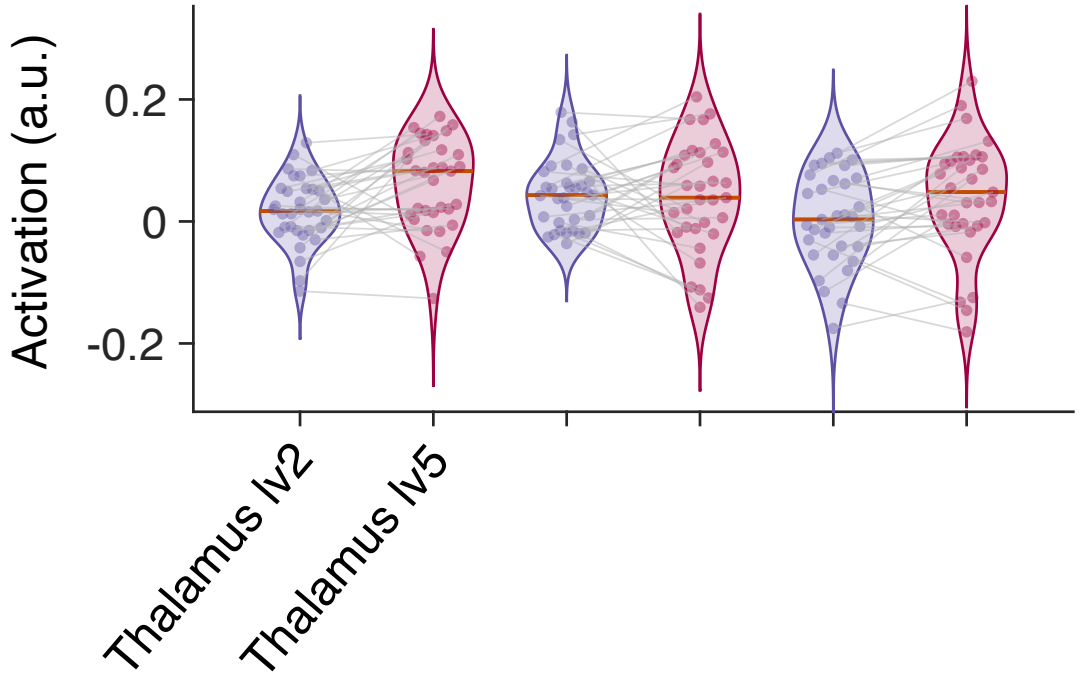
boxplot_wani_2016.m

https://github.com/cocoanlab/cocoanCORE/blob/master/Visualization/boxplot_wani_2016.m

Paper (Woo et al., 2017)



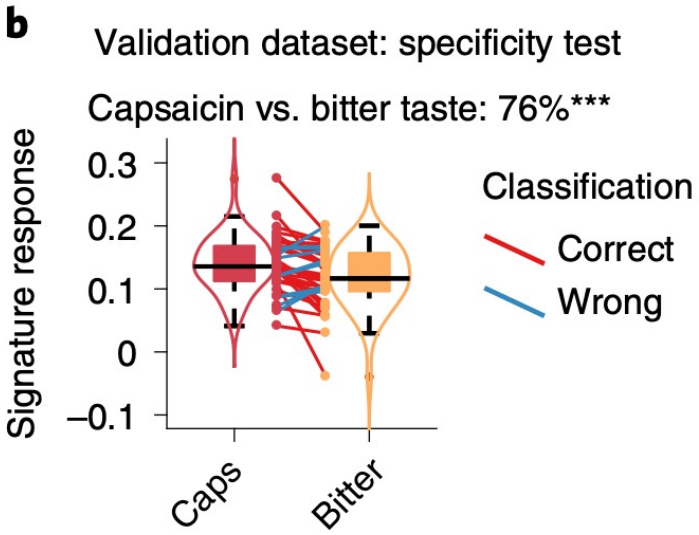
From our code



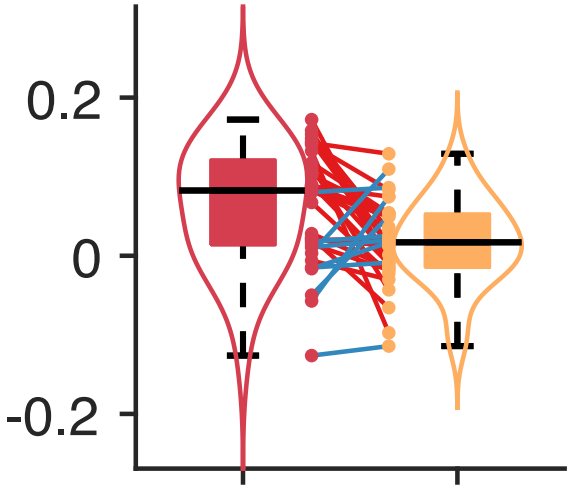
plot_specificity_box.m

https://github.com/cocoanlab/cocoanCORE/blob/master/Visualization/plot_specificity_box.m

Paper (Lee et al., 2021)



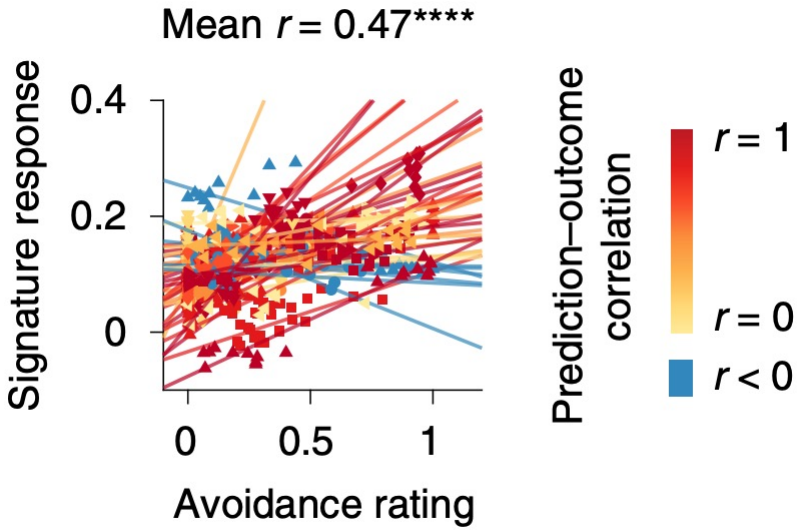
From our code



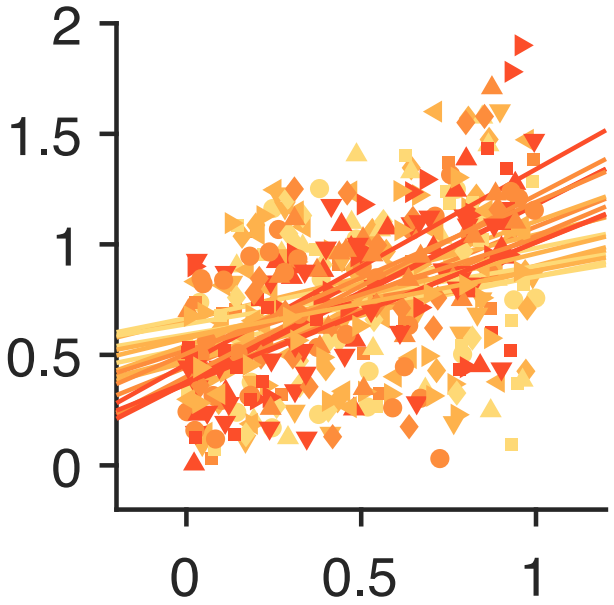
plot_y_yfit.m

https://github.com/cocoanlab/cocoanCORE/blob/master/Visualization/plot_y_yfit.m

Paper (Lee et al., 2021)



From our code



Data visualization - MRIcroGL

Choong-Wan Woo
Director of the Cocoan Lab

MRIcroGL

Download: https://www.nitrc.org/projects/mricrogl