

# Data visualization for inference in tomographic brain imaging

Cyril R. Pernet<sup>1</sup> and Christopher R. Madan<sup>2</sup>

<sup>1</sup>Centre for Clinical Brain Sciences, Edinburgh Imaging, University of Edinburgh, Edinburgh, United Kingdom

<sup>2</sup>School of Psychology, University of Nottingham, Nottingham, United Kingdom

## ABSTRACT

Tomographic brain imaging has a rich iconography. Whilst figures are prepared for scientific communication (i.e., directed to other researchers) they also often end-up on magazine and journal covers (i.e., directed to a lay audience). Scientific figures should however not be just glossy illustrations of what is in the text. One of the primary roles of figures is to carry information that cannot be easily explained in words or summarized in tables ([Rougier et al., 2014](#)). Poor scientific figures are figures that not only fail to convey additional information, but also figures that convey or induce incorrect information, especially for non-specialists. Here we provide a guideline on which visual information to display and in which context, to improve information content and minimize false inference. We first discuss the use of slices versus renders and in which situations they should be used. We next reiterate the need for unthresholded statistical maps (Jernigan et al., 2003) along with (i) the highlighting of significant areas on such maps (ii) the necessity to plot results in all regions of interest, and (iii) the choice of colour scales. Together, these measures provide additional contextual information and should prevent readers natural tendency to falsely infer differences in activations or absence of activations. Additional recommendations are also given to convey information about hemispheric asymmetry and effect sizes.

## Author contributions statement

CP and CM contributed equally to the concept and writing up of the article. CP created the new colour maps and analysed the fMRI data.

## Introduction

Tomographic imaging (i.e., magnetic resonance imaging [MRI], positron emission tomography [PET], X-ray computed tomography [CT]) offers a unique window in understanding structure-function relationships in the living brain. Nowadays, it is standard practice to acquire entire brain volumes and perform statistical analyses at each voxel, an approach known as statistical parametric mapping (Friston et al., 1991). Because it is impossible to report the statistical results in every voxel, summary tables and figures are of importance. The choice made by authors to create such figures should however not be driven by aesthetic considerations alone but also driven by the message to convey (Rougier et al., 2014). This is not to say that beautiful figures should not be used, as more appealing figures might in fact help in remembering results (Borkin et al., 2013; Madan, 2015b). At the intersection of psychology, computer vision, graphic design, and statistics, there is a field of research that looks at how to represent information and what features are beneficial or harmful in figure designs. For instance, Cleveland and McGill (1983), showed that changing the axis scaling in scatter plots can alter inference on associations between variables. Siegrist (1996) show that using perspective in pie charts, lead to falsely infer magnitude differences because the slices that are closer to the reader appear to be larger than those in the back. In general, there are recommendations for plotting the data rather than summary statistics as those summary values can be obtained with very different distributions which can preclude the use of some statistical tests (see e.g. Anscombe, 1973, for correlations or Weissgerber et al., 2015, for bar graphs). Here we discuss information provided in figures when presenting tomographic data results. Many proposals have already been made by others, what we offer is a principled way chose among those proposals and apply them.

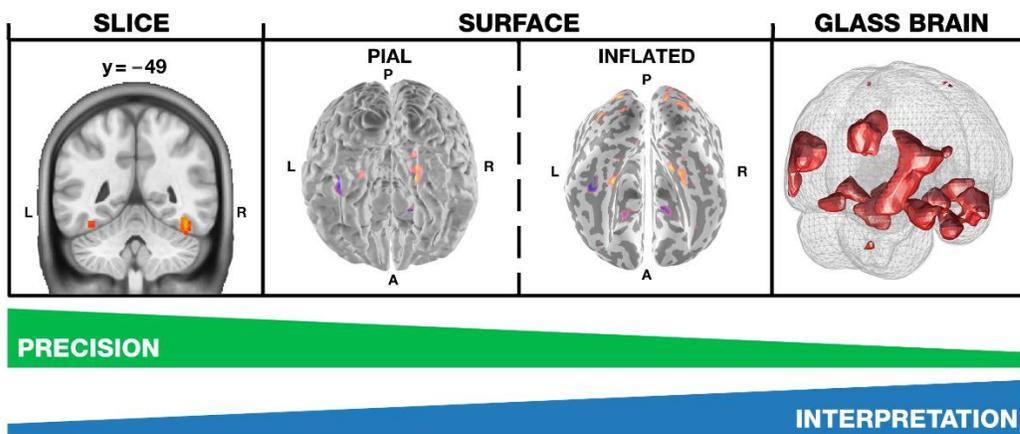
A review of articles using tomographic techniques published between January 2016 and June 2018 in the European Journal of Neuroscience ( $N = 30$  - [https://github.com/CPernet/MRI\\_FaceData\\_Wakeman-Henson/blob/master/DataViz/EJN\\_paper\\_review.csv](https://github.com/CPernet/MRI_FaceData_Wakeman-Henson/blob/master/DataViz/EJN_paper_review.csv)) shows that four broad types of messages are obtained from statistical parametric maps: (1) demonstrating an effect over the whole brain or in specific tissues (e.g., grey matter); (2) showing the anatomical location of an effect; (3) revealing an hemispheric asymmetry for a condition or stimulus; (4) demonstrating the involvement of a set of areas or networks in a given task or between groups or conditions. By associating the message with the design, it appears clearly that slices are preferred to display the anatomical location of an effect (5.3 slice/render ratio), while displaying a set of areas and networks results are less clearly associated with a design (2.5 slice/render ratio) although still dominated by displaying slices. The description of the anatomical location of an effect was described 19 times, with 16 figures using slices and 3 using renders. The involvement of a set of areas was described 22 times with 15 figures using slices only, and 6 with renders (3 of them being both renders and slices). Among all 30 studies, only 2 showed the raw statistical data (i.e., unthresholded map), and 13 (i.e., 43%) did not produce any data plots associated with the data. Among the 17 studies plotting data associated to maps, 3 showed inconsistency in that mapping (i.e., do a plot for one effect but not another) and 10 plotted significant results only, i.e., only 23% of all studies plotted results independent of statistical significance.

In the following, we propose graphical designs for each type of message, adopting the three principles proposed by (Cleveland & McGill, 1985). First, for readers to appreciate where an effect is, slices or renders should be used depending on the message (principle 1: detection). Second, to improve inference, the assemblage of visual information must be performed to create a unified representation of the results (principle 2). Third, to convey information about the strength of effects, accurate colour scales and plots must be used (principle 3). Proposals are illustrated using data from (Wakeman & Henson, 2015) in which 16 participants view famous, unfamiliar and scrambled faces. Each image was repeated twice (immediately in 50% of cases and 5-10 stimuli apart for the other 50%) and subjects pressed one of two keys with either their left or right index finger indicating how symmetric they regarded each image relative to the average. Here, only the main effects of face recognition (famous faces + unfamiliar faces > scrambled faces) is investigated, independently of repetition levels. Resources necessary to process raw data and generate the figures in this article are available at [https://github.com/CPernet/MRI\\_FaceData\\_Wakeman-Henson](https://github.com/CPernet/MRI_FaceData_Wakeman-Henson) and <https://www.github.com/cMadan/MRIdataviz>. All figures are also available under CCBY licence at data share <https://doi.org/10.7488/ds/2516> (Pernet & Madan, 2019).

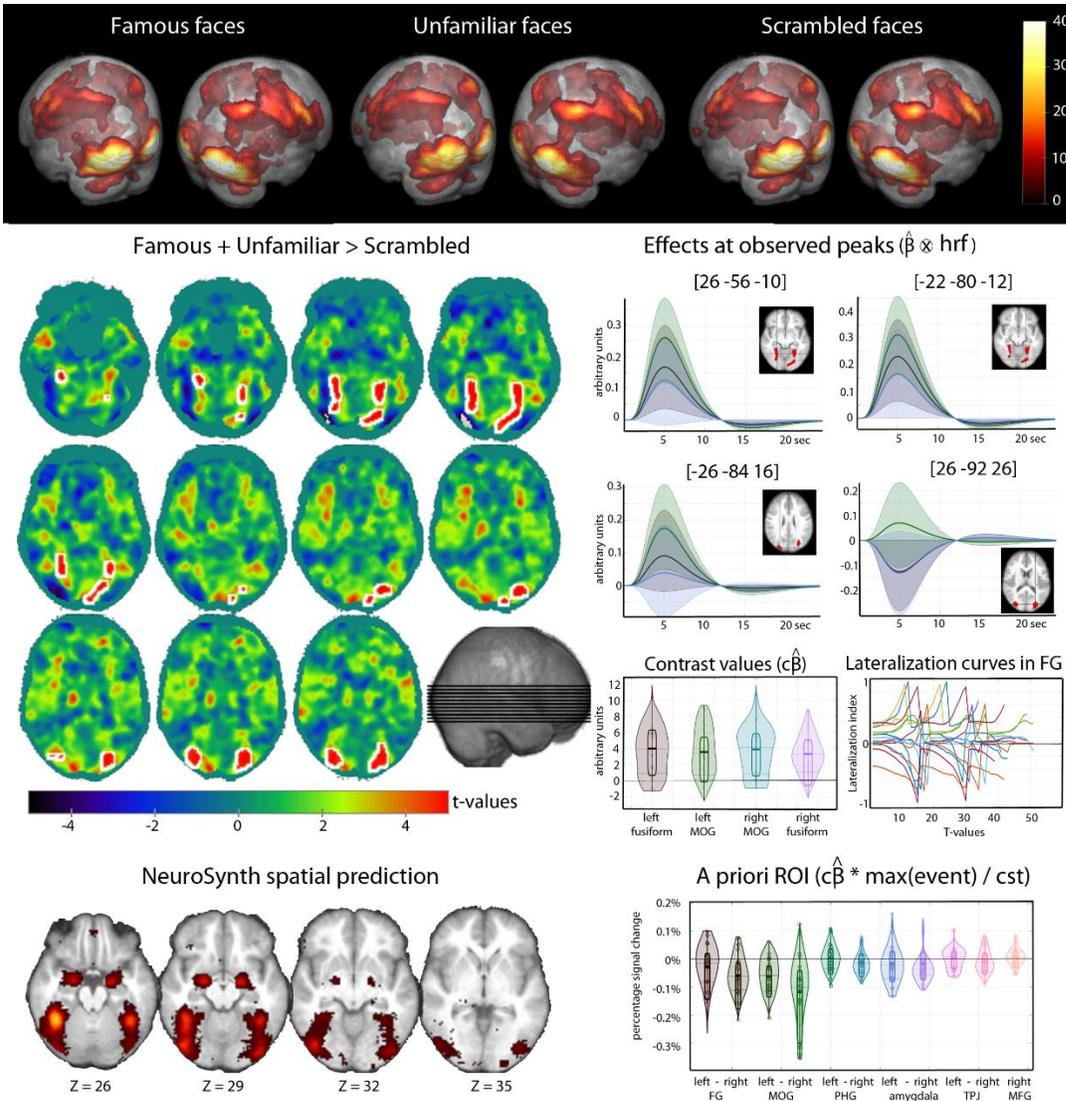
## Show me the blobs

The first principle, detection, refers to the ability for readers to detect where effects are. Presenting an SPM using multiple views is, therefore, better than any single view approach (Madan, 2015a; Ruisoto et al., 2012). We can, however, distinguish two general cases that will drive the design: presenting networks/sets of areas involved in a task vs. illustrating the precise anatomical involvement of an area. In the first case, readers must be able to detect all areas, and in the second case, they must be able to detect the spatially circumscribed area under scrutiny. Our mini review shows that slices are typically shown even when the message is about sets of areas, thus failing to show the distribution of activity throughout the brain. Figure 1 illustrates this using the thresholded map for the contrast famous faces + unfamiliar faces > scrambled faces. In the slice view, we can clearly see bilateral fusiform activity. The surface view also shows the extent of this activity along the fusiform gyrus, particularly on the inflated surface. This surface view provides, in addition, some indication of the distribution of activity. Considerations should, therefore, be taken to decide if a 'regular' grey matter (pial) surface or an inflated surface better conveys the cluster extent. Indeed, the presentation of multiple image display techniques ("fused images") has been shown to aid in data interpretation (e.g. increase in location agreement among clinicians) and comprehension (e.g. relating lesions to an activation pattern – see Stokking et al., 2003, for a review). The use of inflated or pial surfaces can be particularly relevant if some clusters are sufficiently within a sulcus to be not visible on a pial surface, such as the occipital clusters in Figure 1. The glass brain representation gives the most complete depiction of "active" areas but makes it difficult to localize the precise location of the activity. When the message is about networks or the involvement of many areas, we thus recommend using a glass brain (Madan, 2015a), preferably shown from two viewing angles with a slight offset to aid in the interpretation of overlapping clusters and the perception of depth. If space is available this may be complemented with

slices when subcortical structures are involved as only presenting a glass brain view makes it difficult for the reader to determine the depth of the activation cluster. To illustrate the anatomical location of an effect, slices and (orthogonal) cross-sectional views are recommended. For deep structures, additional three-dimensional representations may also be useful (Ruisoto et al., 2012) to provide information about the volume of activation relative to anatomical structures. This principle is illustrated in Figure 2 with all areas significantly activated by stimuli (simple effects) displayed on a render, thus creating a representation of the overall pattern of activation for this task. In contrast to these large effects, localized effects were observed for the contrast of interest famous faces + unfamiliar faces > scrambled faces and are thus displayed on slices.



**Figure 1.** Slice, surface, and glass brain representations of the intact faces > scrambled pictures contrast. Slice view was generated using bspmview. Surface views were generated using mni2fs (Price, 2015). Glass brain views were generated following the procedure described in Madan (2015a). An unthresholded statistical map of the contrast can be viewed at <https://neurovault.org/images/68963/>.



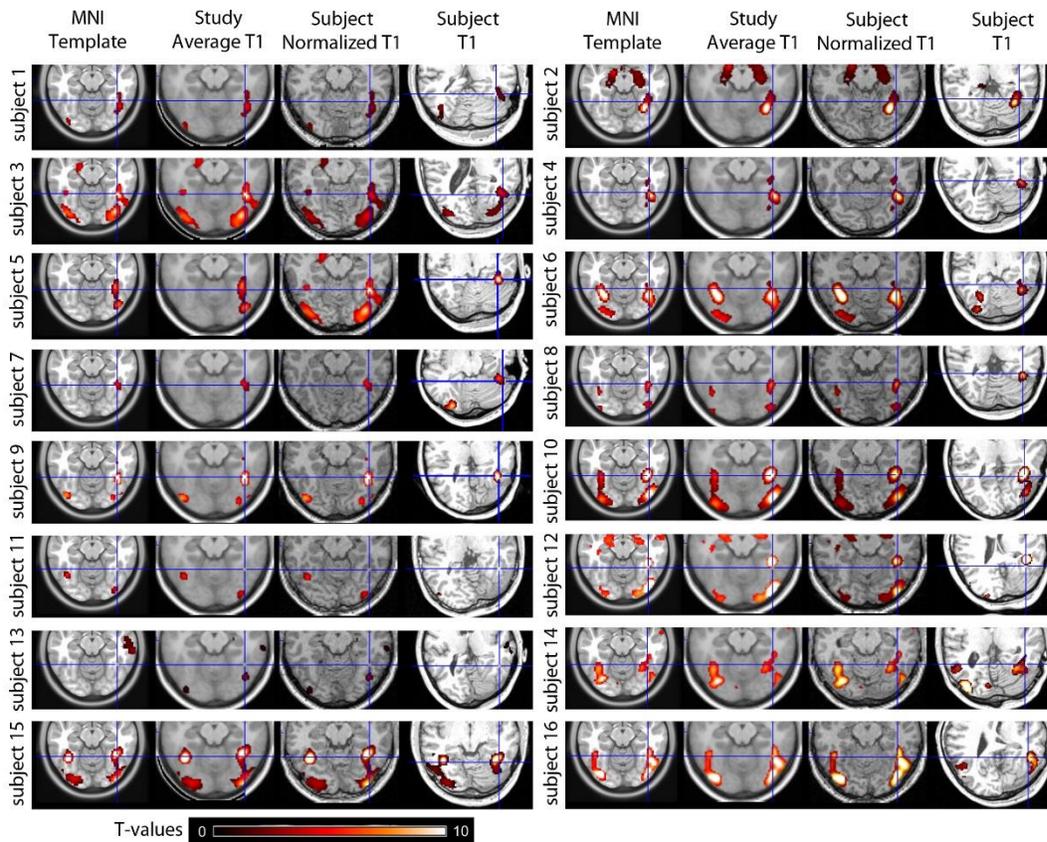
**Figure 2.** Random effect results from [Wakeman and Henson \(2015\)](#). At the top is shown areas involved in each condition and the contrast of interest famous faces + unfamiliar faces > scrambled faces is shown in the middle. Maps show the T-values with significant areas outlined in white (cluster FDR<0.5 CDT=0.001). Responses observed at the peak coordinate for each of the four clusters are shown per condition (reconstructed hemodynamic response: famous faces in brown, unfamiliar faces in blue and scrambled faces in green; shaded areas show the bootstrapped 95% confidence intervals) along with the scatter plots and kernel density estimates of the resulting contrast. Lateralization curves in the fusiform gyri (i.e. lateralization index computed on maps with increasing threshold) for each subject are also shown for that contrast. At the bottom is shown the association map for the term faces from NeuroSynth meta-analysis, with the percentage signal change (contrast maximum(event)/constant, see [Pernet, 2014](#)) in those *a priori* regions of interests (data scatter with kernel density estimates in shaded colours, with rectangles showing the means and 95% Bayesian credible intervals). Data behind all the plots are available from the online repository.

For group studies, we recommend using the average of normalized participants' T1 volumes as an underlay to more accurately portray the anatomical locations of effects relative to structures (and incidentally show the alignment of anatomical structures among subjects). For instance, results from [Wakeman and Henson \(2015\)](#) shown in Figures 2 and 4 are using such average. By using the average T1, one can appreciate better the amount of smoothness in registration and true brain coverage. When this approach is impractical, e.g., the study involves a between groups analysis where anatomical differences are expected (such as young vs. older adults), we recommend using the ICBM 2009c nonlinear asymmetric structural volume ([Fonov et al., 2011](#)) ('*mni\_icbm152\_t1\_tal\_nlin\_asym\_09c*') or the study template, if one was created. When considering individual participant's activation (fMRI, PET), results must be presented on their own structural images, and never on a 'standard' brain as it leads to inaccurate reporting of the anatomy ([Devlin & Poldrack, 2007](#)). As shown on Figure 3, for some participants differences in activation locations between the subject anatomy and the template are small (e.g. participant 15) and for others, the anatomy is very different from the template (e.g. participant 3).

It should finally be noted that figures do not have to be static. No doubt science communication is moving away from paper and Portable Document Format (pdf), and we encourage the community to embrace interactive figures using visualization software such as Papaya ([Mango Team, 2016](#)), NiftyView ([Deng, 2016](#)), BrainBrowser ([Sherif et al., 2015](#)), BrainNetBrowser ([Xia et al., 2013](#)), PySurfer ([Waskom et al., 2016](#)), or Pycortex ([Gao et al., 2015](#)). NeuroVault ([Gorgolewski et al., 2015](#)) also provides a useful demonstration of such visualizations where raw statistical maps can be seen and exchanged (see our results from Figure 3 at: <https://neurovault.org/collections/4319/>).

## **Beyond blobs**

The second principle for graphic design is the assemblage of visual information. The goal is to provide readers with a visual summary of the different information available to help with inference and interpretation. As discussed in [Poldrack et al. \(2017\)](#), claims of absence of effect and selective activations as well as usage of reverse inference are common in neuroimaging, but are often wrong because they require additional quantitative testing. We contend that these errors partly relate to which information and how information is displayed in figures and that better figure designs can help with inference.



**Figure 3.** Single participant results from [Wakeman and Henson \(2015\)](#) displayed for each subject in MNI space (MNI template, average study T1, normalized T1) and in native (back-reconstructed SPM) space, illustrating the importance of using adequate anatomical underlays when looking at SPMs. For instance, when the subject anatomy is similar to the template results appear at the same location (subject 15) but large displacements can occur because of spatial normalization (subject 3). Looking at the fusiform face area (marked by the blue cross), each subject (except 11 and 13) show a right fusiform activation although the locations seen on individual's brain indicate that some subjects (e.g. 1, 2, 5, 6, 7) have maximum activations located in a more inferior part of the gyrus than what is expected from the group result.

### Absence of effect

Absence of statistical significance is not an absence of effect ([Killeen, 2005](#)) and in the absence of significance one only fails to reject the null hypothesis ([Pernet, 2017](#); [Lakens, 2017](#)). Whilst the error is common in behavioural sciences, it becomes the norm when describing results from statistical parametric maps: if there is no activation (above a statistical threshold signal), one typically infers that there is no effect. Unless using equivalence testing or Bayesian statistics, it is however impossible to infer that a given experiment or comparison did not lead to activation in a given region. For instance, results from [Wakeman and Henson \(2015\)](#) show a significant activation for faces compared to scrambled faces in the medial fusiform

gyri (Figures 2 and 4) but that does not indicate that other regions are not also activated in response to face stimuli, or even more activated by intact faces than scrambled faces. In fact, strong but non-significant effects can also be seen more laterally.

Reporting all effects, no matter the level of significance is the most effective way to convince readers of the results. Showing raw (unthresholded) statistical maps is thus a step in that direction (Jernigan et al., 2003). We recommend here to go even further and plot and test effects for all areas expected to be *a priori* activated, given the experimental hypotheses. Thanks to the vast literature on face perception, we can generate spatial predictions using meta-analysis engines such as NeuroSynth (Yarkoni et al., 2011). Here, we can predict bilateral activations in the posterior middle occipital gyri, lateral fusiform, parahippocampal regions, amygdalae, temporo-parietal junctions and right inferior frontal gyrus (reverse inference map thresholded with a minimal extent of 20 voxels from <http://www.neurosynth.org/analyses/terms/face/>). Since our whole brain analysis did not reveal significant differences between intact faces and scrambled faces in these regions, one might infer that there is no effect. Statistical testing in these *a priori* ROIs, however, shows that this would be the wrong inference to make, as differences can be observed. As shown in Figure 2, the Bayesian bootstrap of the mean reveals stronger activations for faces than scrambled faces in the right (lateral) fusiform gyrus and left and right middle occipital gyri (i.e. highest density intervals of the difference did not include 0, see Table 1). Because of expectations (i.e., hypotheses) about where effects should be localized, and that many studies found effects or differences at these locations, reporting results using such priors is worthwhile as this allows comparing results across studies and reduce false inference. A practical aspect to consider when using *a priori* ROIs is how to generate them. When possible, using NeuroSynth or GingerALE (Eickhoff et al., 2009) is recommended as these meta-analysis tools can generate unbiased ROI. For investigators interested in checking across the whole brain where are "in limbo" areas (above baseline but also under the statistical threshold of significance), sandwich estimator can be used (de Hollander et al., 2014) testing if the difference relative to significant areas is itself significant (Gelman & Stern, 2006).

Region	[ x, y, z ]	Mean	95% CI
left fusiform	-40 -50 -18	-0.0286	[-0.1616, 0.0141]
right fusiform	45 -50 -22	-0.0589	[-0.1407, -0.0170]
left middle occipital	-38 -84 -10	-0.0620	[-0.1513, -0.0242]
right middle occipital	44 -78 -12	-0.1174	[-0.2771, -0.0388]
left parahippocampal	-26 0 -30	0.0046	[-0.0520, 0.0331]
right parahippocampal	34 -8 -32	-0.0109	[-0.0541, 0.0173]
left amygdala	-20 -6 -18	-0.0163	[-0.0966, 0.0311]
right amygdala	22 -6 -18	-0.0174	[-0.0772, 0.0225]
left temporo-parietal junction	-56 -54 8	0.0032	[-0.0396, 0.0266]
right temporo-parietal junction	54 -44 8	-0.0055	[-0.0525, 0.0197]
right middle frontal gyrus	42 14 22	0.0079	[-0.0377, 0.0293]

**Table 1.** Summary of choices and designs to use based on the messages to convey.

## Plotting effects

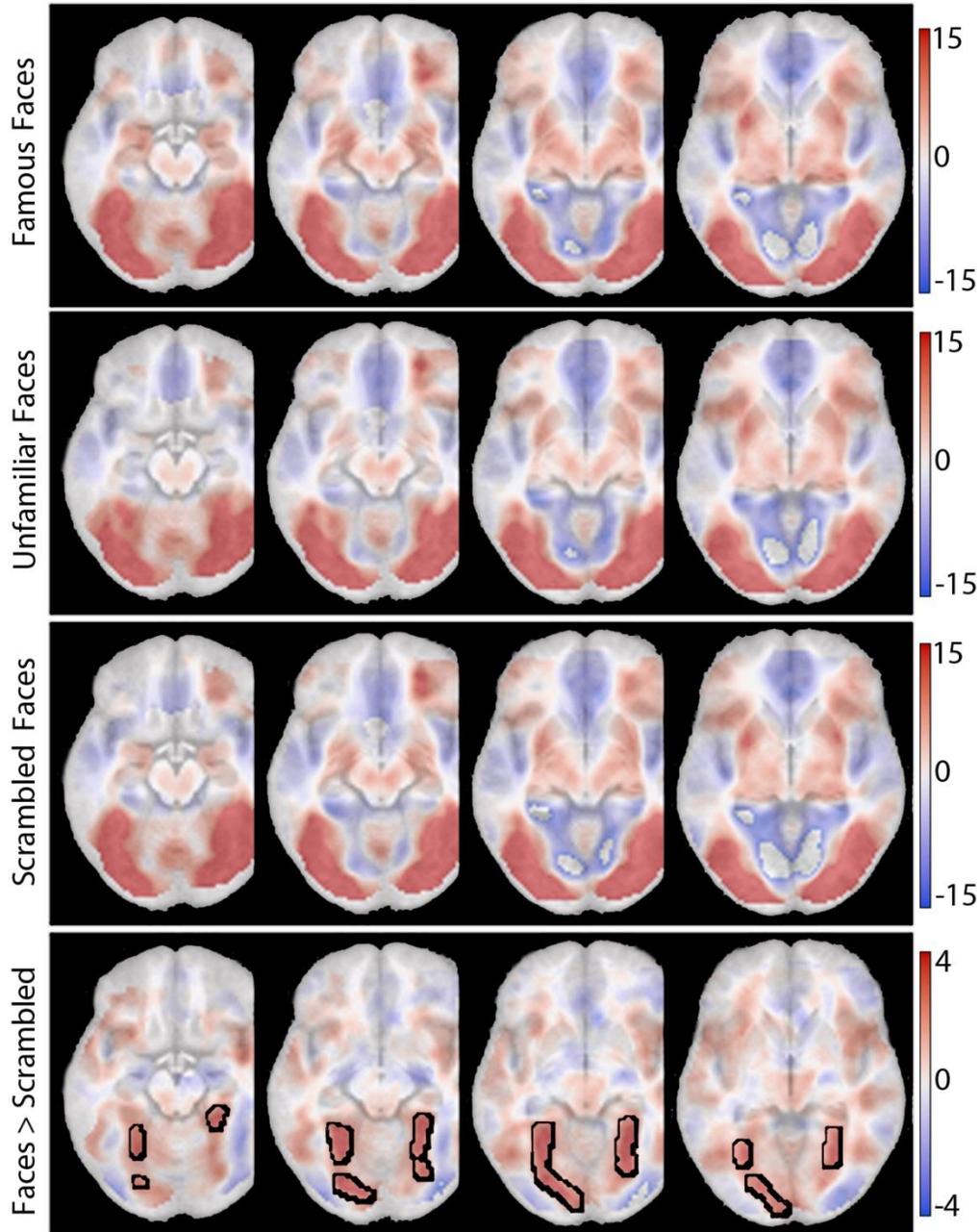
Because of the non-stationary spatial nature of baseline activity, summary statistics (average values, T or F values, etc) displayed on slices and renders can have widely different physiological interpretations, and it is, therefore, essential to plot results for all a priori ROI but also all regions seen as significant. It may, of course, be impractical to have all plots in the core of a publication when many results are observed, but it is easy to provide this information as supplementary material, along with csv files or raw data behind the plots.

Consider for instance a positive contrast (i.e. the mean value is bigger than 0). Such result can be obtained from three configurations: (i) all conditions are superior or equal to zero (e.g., left/right fusiform gyri), (ii) all conditions are inferior or equal to 0, (iii) conditions vary around baseline. The two first scenarios are 'easily' interpretable, while the last case is much harder to understand. This is well illustrated in Figure 2, with the contrast famous faces + unfamiliar faces > scrambled faces. The contrast values in significant areas have similar distributions, yet the right MOG has a completely different pattern of response. We, therefore, recommend to systematically show data points (e.g., beta estimates or percentage signal change) for each condition along with the summary statistics of contrasts (typically the mean, but not always). For these plots, showing means and standard deviations using bar graphs is inadequate ([Rousselet et al., 2016](#); [Weissgerber et al., 2015](#)), and box-plots or violin plots along with data scatter are recommended. Similarly, reconstructed hemodynamic responses can be plotted if they convey enough information about variance across subjects.

## Selective activations

The perception of a lack of effect in areas not significantly activated leads to incorrect inferences about the selectivity of significantly activated areas, an inferential error known as the 'imager fallacy' ([Henson, 2005](#)). [Engel and Burton \(2013\)](#) showed that over 80% of naive readers are making such error when looking at a thresholded SPM. What we detect as face-selective depends on both the task and the baseline used ([Stark & Squire, 2001](#)), here scrambled faces, and on the statistical threshold used. The issue of selectivity or specificity of activations in the brain has been hotly debated (see, for example, [Pernet et al., 2007](#)) but all agree that it requires statistical testing and cannot be inferred merely from showing a qualitative different activation pattern. What a qualitatively different pattern of activation between stimuli or conditions does allow one to infer (although it would need actual statistical testing of the interaction regions \* stimuli or conditions), is that information processing differ in at least one function (function-to-structure deduction as opposed to structure-to-function induction; [Henson, 2005](#)). We, therefore, recommended showing raw statistical maps ([Jernigan et al., 2003](#)), assembling results of simple effects and contrast of interests as shown in Figure 2. This allows addressing, at least visually, issues of the absence of effects, selectivity, and qualitative difference. In Figure 4, the raw maps of simple effects allow inferring that information processing was similar across all three conditions because we have similar activation pattern. Sharing such unthresholded statistical maps is also highly encouraged, using online repositories such as NeuroVault ([Gorgolewski et al., 2015](#)). The result from the contrast of interest is also shown unthresholded thus addressing

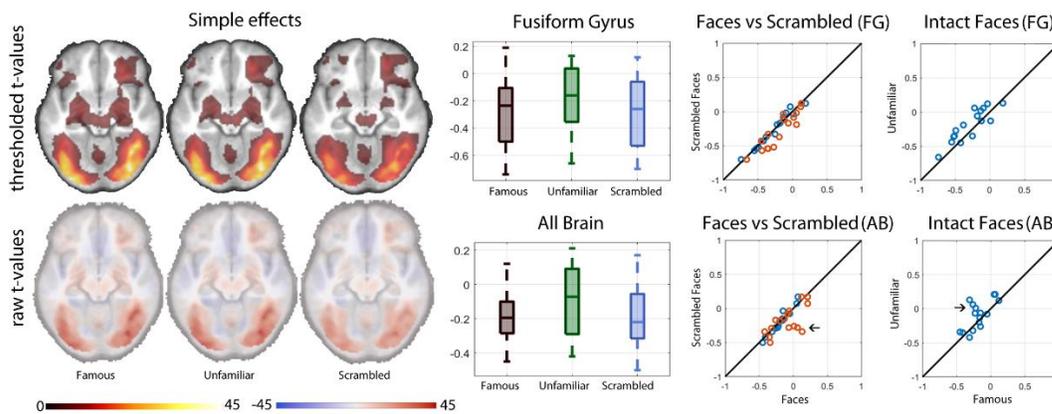
the issue of non-significant areas. It is however also important to point at where the evidence supports the existence of an effect, thus highlighting significant areas (Allen et al., 2012), here using contours (this can be achieved easily using tools such as nanslice; Wood, 2018).



**Figure 4.** Random effect results illustrating the selective response of the medial fusiform cortex. Activations, relative to baseline, are seen for each condition across the entire inferior occipital cortex while the contrast intact Faces > Scrambled ( $= 0.5 \cdot \text{Famous} + 0.5 \cdot \text{Unfamiliar} - 1 \cdot \text{Scrambled}$ ) leads to significant differences (cluster FDR  $\leq .005$ , CDT = .001, outlined in black) only in medial fusiform regions.

## Hemispheric asymmetry

The same way as spatial selectivity can wrongly be inferred from the absence of statistically significant results, hemispheric asymmetry is often inferred from thresholded maps. As for selectivity, it is recommended to statistically test for hemispheric differences going beyond the single level of activation by computing lateralization indexes based on bootstrapped lateralization curves (i.e., using the size and intensity of clusters across all thresholds as e.g. implemented in the LI toolbox, [Wilke & Lidzba, 2007](#)). Here, when testing for fusiform activation asymmetry, individual conditions/stimuli were right lateralized (95% CI famous faces [-0.55, -0.14], unfamiliar faces [-0.45, -0.06], scrambled faces [-0.54, -0.13]) while visually, maps did not allow to see this pattern (Figure 5). When considering the whole brain, only famous faces [-0.38, -0.08] and scrambled faces [-0.37, -0.06] show right lateralization (unfamiliar faces [-0.029, 0.022]). To best understand the pattern of lateralization across stimuli/conditions, we also recommend using paired observations on scatter plots rather than (or in addition to) box plots or violin plots ([Rousselet et al., 2016](#)).



**Figure 5.** Lateralization Indices (LI) for each condition obtained using the LI toolbox ([Wilke & Lidzba, 2007](#)). Box plots show the median LI and 1.5 times interquartile range. Scatter plots show LI for famous (blue) and unfamiliar faces (red) vs. scrambled faces, and the direct comparison famous vs. unfamiliar faces. Simple effects for each condition seen on slices (left) show no strong evidence of lateralization whilst lateralization indices tell otherwise. Compared to box plots, paired scatter plots allow to better understand relationships between items. In the fusiform gyri, all subjects lie along the diagonals, indicating very similar indices across stimuli and responses across subjects. For the whole brain, the box plots show the same pattern of results as for the fusiform gyri, although with a non-significant asymmetry for unfamiliar faces. Paired scatter plots also show that 4 subjects (indicated by the arrows on the scatter plots) have close to zero lateralization for unfamiliar faces while showing lateralization for familiar and scrambled faces, which is not seen for the fusiform gyri.

## Colouring maps and blobs

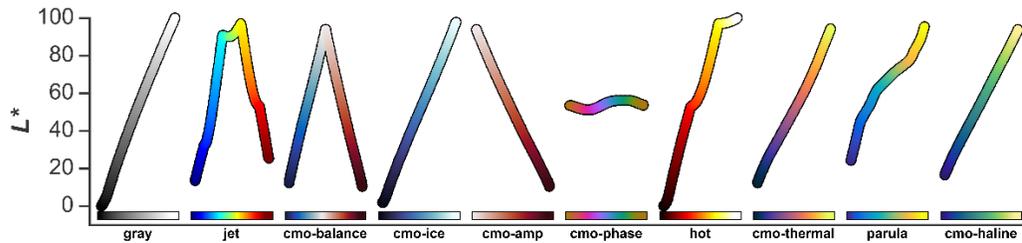
The third principle in graphical design is the use of 'accurate' colour scales. Although imaging researchers know to be cautious when conducting their analyses to account for relevant nuisance regressors and determining cluster thresholds,

they often pay little attention in selecting colours to visualize the results of their imaging study. When the message is about where active regions are, a single colour can be used. When information about the spatial distribution of statistical values is also of interest, colour palettes (scales) should be used. Such palettes must appropriately convey the underlying data and not introduce perceptual biases. This topic has been investigated at length within other fields of study including geography (Brewer et al., 2003; Thyng et al., 2016; Light & Bartlein, 2004) and astronomy (Green, 2011) and brain imagers should also be considering this issue.

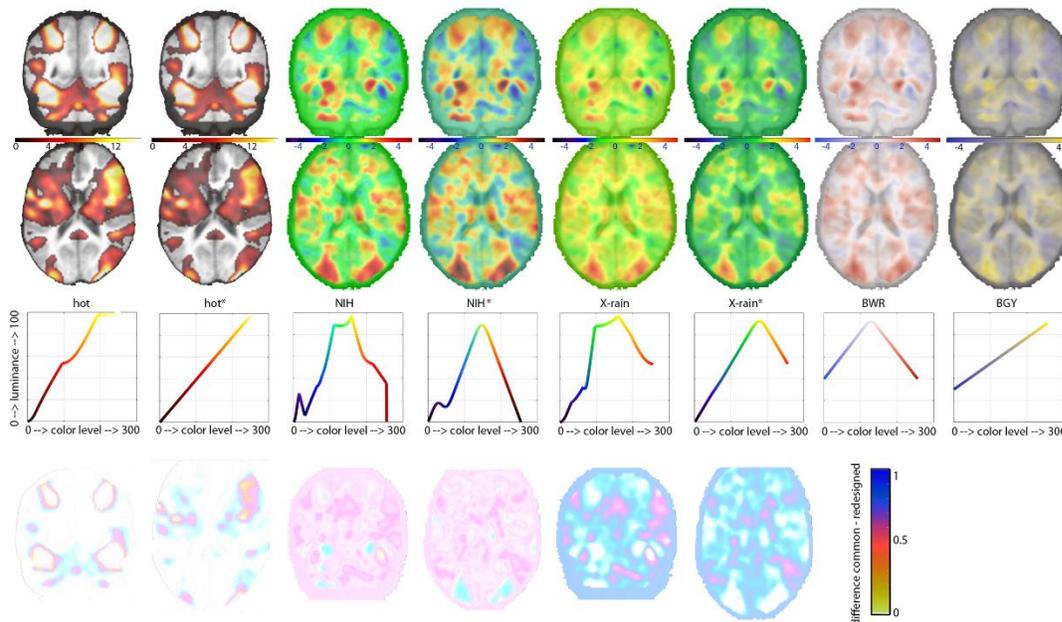
Colours in digital images are often generated as combinations of red, green, and blue (RGB) intensities. This is, however, *not* how colours are perceived by humans. An alternative colour space, CIELAB, has been developed to correspond to the human perceptual system. In 1976, the International Color Consortium (a.k.a. Commission Internationale de l'Eclairage, CIE) defined a colour space perceptually uniform that relies on luminance ( $L^*$ ), red-green ( $A^*$ , ~550–700 nm wavelength), and blue-yellow ( $B^*$ , ~400–550 nm wavelength) colours (International Color Consortium, 2006). The critical factor is that changes in luminance are better perceived than hue to reflect changes in magnitude (Cleveland & McGill, 1985) and that averaging RGB values does not linearly correspond to changes in luminance. 'Traditional' colour maps used in brain imaging are not informed by this and instead lead to distortions in how colour intensities are perceived and interpreted (Figure 6). Some colour maps have previously been developed with brain imaging in mind, e.g., Ridgway (2009), but it has saturation/luminance issues as most other maps. More recently, scientists have generally become more aware of this colour perception artefact with the change of default colour map in some software packages to veridic (Smith & van der Walt, 2015) or parula (Edens, 2014). The luminance function of many sequential, diverging, and rainbow colour maps are shown in Figure A1. For more detailed discussions of luminance effects in colour maps, see Borland and Taylor (2007) and Niccoli (2012).

Using the method described in Kovesi (2015), we have developed corrected colour maps now available as .mat (implemented by SPM via `spm_colourmap.m`), .csv, .cmap (implemented in FSLeYes 0.26.1), .lut (for MRICroN, and implemented as .clut in MRICroGL v12): [https://github.com/CPernet/brain\\_colours](https://github.com/CPernet/brain_colours), as demonstrated in Figure 7. At the bottom of the figure, is shown the difference between the common maps and the redesigned, corrected ones. On a linear scale such as hot, the corrected maps show less saturation within clusters, leading to better appreciate spatial variations, as best seen for the right frontal cluster. For diverging maps, the corrected maps show better the differences in space between positive and negative values, like here for negative values in the visual cortex or the right anterior fusiform gyrus. Linear colour scales (e.g., hot or BGY) are ideally suited for continuous positive or negative values (e.g., contrast T-maps) whilst diverging colour scales (e.g., NIH, BWR) are better suited for continuously negative to positive value maps (e.g., contrast F-maps), but it is essential to make them symmetric as to have the mid-luminance value reflecting the 0 value in the data. For this reason, we have added the CET-D1 and CET-D7 to the repository (referred to as Blue-White-Red (BWR) and Blue-Grey-Yellow (BGY) in 7), the latter one having the advantage of having no perceptual dead-spot at the centre. Cyclic colour maps are ideal to display information about angles as in retinotopic mapping or fibre orientation (see Table 2 for a summary of designs). Other colour maps such as rainbow or spectrum

should be avoided as they cycle through luminance. Finally, when using 3D renders, because lighting is used to give a three-dimensional aspect to the image, it interferes with the luminance of colour maps, and isoluminant colour maps (also available in the repository) or single colour should be used. Since inferential errors relate mostly to local effects (comparison of neighbouring regions, see above), this is, however, less problematic.



**Figure 6.** Overview of luminance function for common and recently improved colour maps. “cmo” refers to colour maps from the cmocean package (Thyng et al., 2016). See Figure A1 for additional luminance functions of common colour maps.



**Figure 7.** Same set of results shown using common and redesigned (marked with \*) colour maps. The first two rows show results on coronal and axial slices, with below the corresponding colour maps plotted as a function of luminance. For the hot, NIH and X-rain maps, the difference in presented SPMs between the old and redesigned maps is shown, highlighting where are the differences.

Although the focus here is on visualizing tomographic mapping, these principles apply generally to heat maps (also see Gehlenborg & Wong, 2012) and should be of interest to all brain imagers, with other possible applications for magneto- and

electroencephalography SPM, scalp maps and source reconstruction figures. Furthermore, it may be desirable to show categorical data rather than continuous, e.g., when showing several anatomical regions-of-interest or graphs of activations in different task conditions. Here we suggest using the distinctive colour sets proposed by [Brewer et al. \(2003\)](#), [Wong \(2011\)](#), or [Kelly \(1965\)](#). See Figure A2 for examples and details for these colours.

Principle	Design	Message
Detection	Render/Glass brain	networks, set of areas
	slices	localized effects
	render and slices	set of areas including subcortical structures
Assemblage	simple effects and contrasts	unified view of results
	unthresholded maps	location of effects and selectivity
Strength of Effect	linear colour maps	negative or positive contrast
	divergent colour maps	unthresholded maps
	cyclic colour maps	tonotopic mapping, fibre orientation

**Table 2.** Summary of choices and designs to use based on the messages to convey.

## Conclusion

SPM must be displayed in non-misleading ways. Our recommendations are simple to adopt and, we believe, should help in making further inference from the results. In general, use 3D renders and glass brains for sets of regions and networks; use slices for the precise anatomical location of an effect. Assemble visual information as to help with spatial inference, combining simple effects with contrast maps and use unthresholded maps highlighting significant areas ([Allen et al., 2012](#)). Carefully choose colour maps to reflect the magnitude of effects. Plot data for all *a priori* ROIs regardless of significance and plot data (simple effects and effects of interest) for each region declared significant during analyses.

## Acknowledgement

Thank you to Dr Guillaume Rousselet for comments on an early version of the manuscript.

## References

- Allen, E. A., Erhardt, E. B., & Calhoun, V. D. (2012). Data visualization in the neurosciences: overcoming the curse of dimensionality. *Neuron*, *74*, 603–608. doi: 10.1016/j.neuron.2012.05.001
- Anscombe, F. J. (1973). Graphs in statistical analysis. *American Statistician*, *27*, 17–21.
- Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, *19*, 2306–2315. doi: 10.1109/TVCG.2013.234
- Borland, D., & Taylor, R. (2007). Rainbow color map (still) considered harmful. *IEEE Computer Graphics and Applications*, *27*, 14–17. doi: 10.1109/mcg.2007.323435

- Brewer, C. A., Hatchard, G. W., & Harrower, M. A. (2003). ColorBrewer in print: A catalog of color schemes for maps. *Cartography and Geographic Information Science*, 30, 5–32. doi: 10.1559/152304003100010929
- Caress, D. W., & Chayes, D. N. (2009). *MB-System Seafloor Mapping Software*. Retrieved from <https://www.mbari.org/products/research-software/mb-system/>
- Cleveland, W. S., & McGill, R. (1983). A color-caused optical illusion on a statistical graph. *The American Statistician*, 6. Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science*, 229, 828–833. doi: 10.1126/science.229.4716.828
- de Hollander, G., Wagenmakers, E.-J., Waldorp, L., & Forstmann, B. (2014). An antidote to the imager's fallacy, or how to identify brain areas that are in limbo. *PLoS ONE*, 9, e115700. doi: 10.1371/journal.pone.0115700
- Deng, W. (2016). NiftyView: a zero-footprint web application for viewing DICOM and NIfTI files. *GigaScience*, 5, 8–9. doi: 10.1186/s13742-016-0147-0-g
- Devlin, J. T., & Poldrack, R. A. (2007). In praise of tedious anatomy. *NeuroImage*, 37, 1033–1041. doi: 10.1016/j.neuroimage.2006.09.055
- Edens, S. (2014). *Rainbow color map critiques: An overview and annotated bibliography* (Tech. Rep. No. 92238v00). The MathWorks Inc. Retrieved from [https://www.mathworks.com/tagteam/81137\\_92238v00\\_RainbowColorMap57312.pdf](https://www.mathworks.com/tagteam/81137_92238v00_RainbowColorMap57312.pdf)
- Eickhoff, S., Laird, A., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, 30, 2907–2926.
- Engel, S. A., & Burton, P. C. (2013). Confidence intervals for fMRI activation maps. *PLoS ONE*, 8. doi: 10.1371/journal.pone.0082419
- Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., & Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54, 313–327. doi: 10.1016/j.neuroimage.2010.07.033
- Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. J. (1991). Comparing functional (PET) images: The assessment of significant change. *Journal of Cerebral Blood Flow & Metabolism*, 11, 690–699. doi: 10.1038/jcbfm.1991.122
- Gao, J. S., Huth, A. G., Lescroart, M. D., & Gallant, J. L. (2015). Pycortex: an interactive surface visualizer for fMRI. *Frontiers in Neuroinformatics*, 9, 23. doi: 10.3389/fninf.2015.00023
- Gehlenborg, N., & Wong, B. (2012). Heat maps. *Nature Methods*, 9, 213–213. doi: 10.1038/nmeth.1902
- Gelman, A., & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. *The American Statistician*, 60, 328–331. doi: 10.1198/000313006X152649
- Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., . . . Margulies, D. S. (2015). NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics*, 9, 8. doi: 10.3389/fninf.2015.00008
- Green, D. A. (2011). A colour scheme for the display of astronomical intensity images. *Bulletin of the Astronomical Society of India*, 39, 289–295.
- Haxby, W. F., Karner, G. D., LaBrecque, J. L., & Weissel, J. K. (1983). Digital images of combined oceanic and continental data sets and their use in tectonic studies. *Eos*, 64, 995–1004. doi: 10.1029/eo064i052p00995
- Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *Quarterly Journal of Experimental Psychology*, 58, 193–233.
- International Color Consortium. (2006). Architecture, profile format, and data structure.

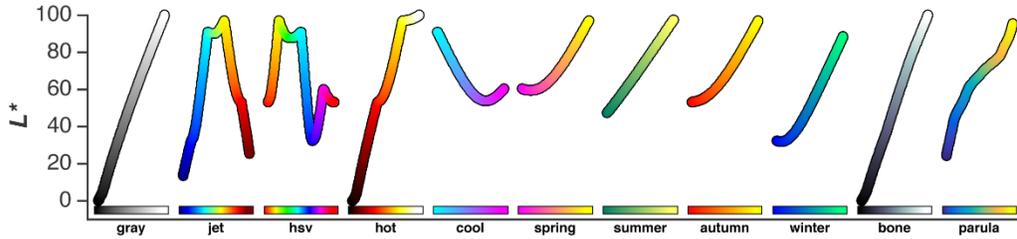
- Specification ICC.1:2004-10. Jernigan, T. L., Gamst, A. C., Fennema-Notestine, C., & Ostergaard, A. L. (2003). More “mapping” in brain mapping: Statistical comparison of effects. *Human Brain Mapping*, 19, 90–95. doi: 10.1002/hbm.10108
- Kelly, K. L. (1965). Twenty-two colors of maximum contrast. *Color Engineering*, 3, 26–27.
- Killeen, P. R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science*, 16, 345–353. doi: 10.1111/j.0956-7976.2005.01538.x
- Kovesi, P. (2015). Good colour maps: How to design them. *arXiv*. Retrieved from <http://arxiv.org/abs/1509.03700>
- Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8, 355-362. doi: 10.1177/1948550617697177
- Light, A., & Bartlein, P. J. (2004). The end of the rainbow? Color schemes for improved data graphics. *Eos*, 85, 385. doi: 10.1029/2004eo400002
- Madan, C. R. (2015a). Creating 3D visualizations of MRI data: A brief guide. *F1000Research*, 4, 466. doi: 10.12688/f1000research.6838.1
- Madan, C. R. (2015b). Every scientist is a memory researcher: Suggestions for making research more memorable. *F1000Research*, 4, 19. doi: 10.12688/f1000research.6053.1
- Mango Team. (2016). *Papaya*. Retrieved from <https://github.com/rii-mango/Papaya>
- Niccoli, M. (2012). *The rainbow is dead. . . long live the rainbow!* Retrieved from <https://mycarta.wordpress.com/2012/05/29/the-rainbow-is-dead-long-live-the-rainbow-series-outline/>
- Pernet, C. R. (2014). Misconceptions in the use of the general linear model applied to functional MRI: a tutorial for junior neuro-imagers. *Frontiers in Neuroscience*, 8, 1. doi: 10.3389/fnins.2014.00001
- Pernet, C. R. (2017). Null hypothesis significance testing: a guide to commonly misunderstood concepts and recommendations for good practice. *F1000Research*, 4, 621. doi: 10.12688/f1000research.6963.5
- Pernet, Cyril; Madan, Christopher. (2019). Best practice for fMRI displays, plots and colour maps, [image]. University of Edinburgh, Centre for Clinical Brain Sciences & Edinburgh Imaging. <https://doi.org/10.7488/ds/2516>.
- Pernet, C. R., Schyns, P. G., & Demonet, J.-F. (2007). Specific, selective or preferential: Comments on category specificity in neuroimaging. *NeuroImage*, 35, 991–997. doi: 10.1016/j.neuroimage.2007.01.017
- Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., . . . Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, 18, 115–126. doi: 10.1038/nrn.2016.167
- Price, D. (2015). *MNI2FS: Surface Rendering of MNI Space Volumes for MATLAB*. Retrieved from <https://github.com/dprice80/mni2fs>
- Ridgway, G. (2009). *Bipolar colormap*. Retrieved from <https://www.mathworks.com/matlabcentral/fileexchange/26026-bipolar-colormap>
- Rougier, N. P., Droettboom, M., & Bourne, P. E. (2014). Ten simple rules for better figures. *PLOS Computational Biology*, 10, 1-7. doi: 10.1371/journal.pcbi.1003833
- Rousselet, G. A., Foxe, J. J., & Bolam, J. P. (2016). A few simple steps to improve the description of group results in neuroscience. *European Journal of Neuroscience*, 44, 2647–2651. doi: 10.1111/ejn.13400
- Ruisoto, P., Juanes, J. A., Contador, I., Mayoral, P., & Prats-Galino, A. (2012). Experimental evidence for improved neuroimaging interpretation using three-dimensional graphic models. *Anatomical Sciences Education*, 5, 132–137. doi: 10.1002/ase.1275
- Sherif, T., Kassis, N., Rousseau, M.-E., Adalat, R., & Evans, A. C. (2015). BrainBrowser: distributed web-based neurological data visualization. *Frontiers in Neuroinformatics*, 8, 89.

doi: 10.3389/fninf.2014.00089

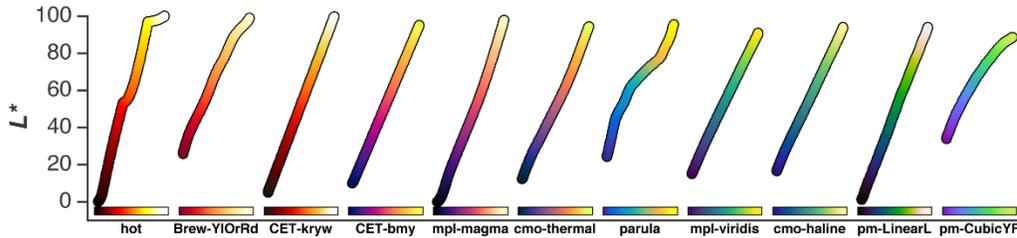
- Siegrist, M. (1996). The use or misuse of three-dimensional graphs to represent lower-dimensional data. *Behaviour & Information Technology*, 15, 96-100. doi: 10.1080/014492996120300
- Smith, N., & van der Walt, S. (2015). A better default colormap for Matplotlib. In *SciPy 2015*. Retrieved from [http:// bids.github.io/colormap/](http://bids.github.io/colormap/)
- Stark, C. E. L., & Squire, L. R. (2001). When zero is not zero: The problem of ambiguous baseline conditions in fmri. *Proceedings of the National Academy of Sciences USA*, 98, 12760–12766. doi: 10.1073/pnas.221462998
- Stokking, R., Zupal, I., & Viergever, M. A. (2003). Display of fused images: Methods, interpretation, and diagnostic improvements. *Seminars in Nuclear Medicine*, 33, 219–227. doi: 10.1053/snuc.2003.127311
- Thyng, K., Greene, C., Hetland, R., Zimmerle, H., & DiMarco, S. (2016). True colors of oceanography: Guidelines for effective and accurate colormap selection. *Oceanography*, 29, 9–13. doi: 10.5670/oceanog.2016.66
- Wakeman, D. G., & Henson, R. N. (2015). A multi-subject, multi-modal human neuroimaging dataset. *Scientific Data*, 2, 150001. doi: 10.1038/sdata.2015.1
- Waskom, M., Gramfort, A., Burns, S., Luessi, M., & Larson, E. (2016). *PySurfer: Cortical neuroimaging visualization in Python*. Retrieved from <https://github.com/nipy/PySurfer>
- Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLOS Biology*, 13, 1–10. doi: 10.1371/journal.pbio.1002128
- Wilke, M., & Lidzba, K. (2007). LI-tool: a new toolbox to assess lateralization in functional MR-data. *Journal of Neuroscience Methods*, 163, 128–136.
- Wong, B. (2011). Color blindness. *Nature Methods*, 8, 441–441. doi: 10.1038/nmeth.1618
- Wood, T. (2018). *Not Another Neuroimaging Slicer - nanslice*. Retrieved from <https://github.com/spinacist/nanslice>
- Xia, M., Wang, J., & He, Y. (2013). BrainNet viewer: A network visualization tool for human brain connectomics. *PLoS ONE*, 8, e68910. doi: 10.1371/journal.pone.0068910
- Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8, 665–670. doi: 10.1038/nmeth.1635

## Appendix

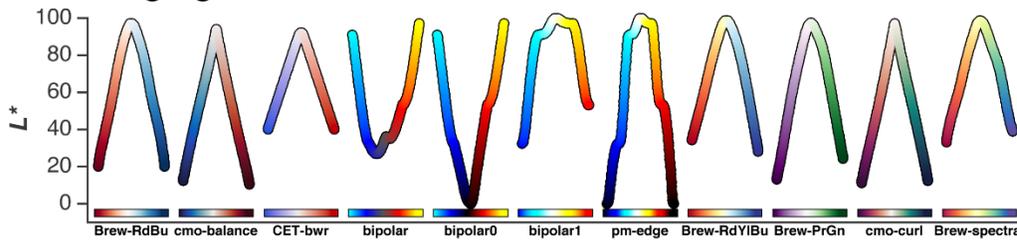
### A. Defaults



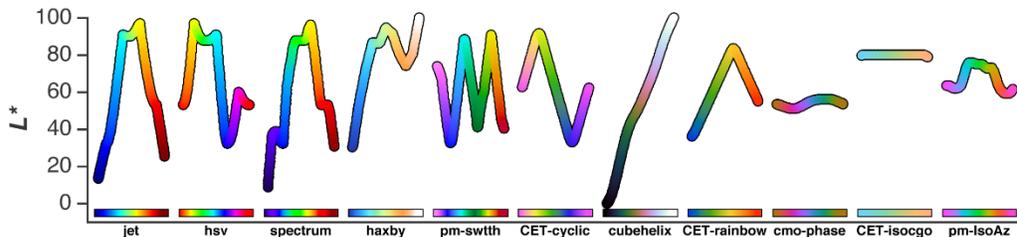
### B. Sequential



### C. Diverging

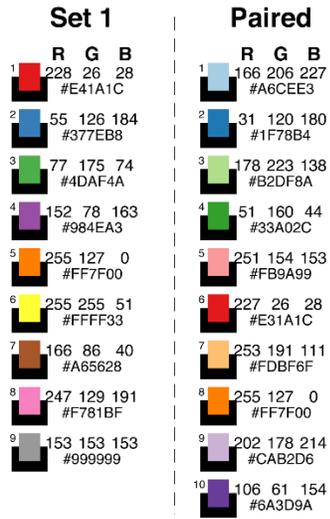


### D. Rainbow

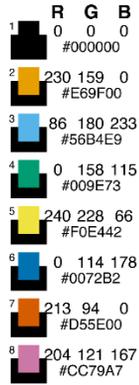


**Figure A1.** Luminance functions for many popular colour maps. For some colormaps, the sequence of colours was reversed to make comparisons among maps clearer. Colour maps are used from the following papers, respectively: “Brew” (Brewer et al., 2003); “CET” (Kovesi, 2015); “mpl” (Smith & van der Walt, 2015); “cmo” (Thyng et al., 2016); “pm” (Niccoli, 2012); parula (Edens, 2014); bipolar (Ridgway, 2009); haxby (Haxby et al., 1983; Caress & Chayes, 2009); cubehelix (Green, 2011); several of the default colour maps (e.g., spring, autumn, bone) are originally from MATLAB, but were later included in other software packages (such as in Matplotlib).

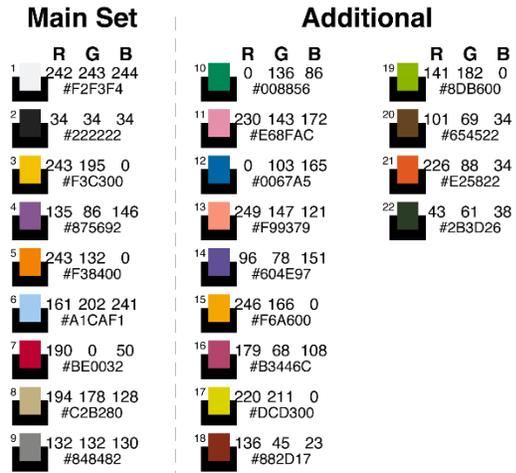
### Brewer et al. (2003)



### Wong (2011)



### Kelly (1965)



**Figure A2.** Suggested distinctive colours for categorical analyses. Colours selected from [Brewer et al.\(2003\)](#) ,[Wong \(2011\)](#), and [Kelly \(1965\)](#).