

Figure Design

Figures are the most engaging part of papers and presentations. Many people will skim your work and only take away anything from the figures, or will only read your work if the figures seem interesting. Figure design is just as, if not more, essential when communicating results to stakeholders and the media. As a result, it is important to make your figures memorable. As a rule of thumb, you should have one figure for each key result.

Unfortunately, making memorable figures which also communicate key points effectively is not easy. As Munzer (2014)¹ says,

The most fundamental reason that vis design is a difficult enterprise is that the vast majority of the possibilities in the design space will be ineffective for any specific usage context. In some cases, a possible design is a poor match with the properties of the human perceptual and cognitive systems. In other cases, the design would be comprehensible by a human in some other setting, but it's a bad match with the intended task. Only a very small number of possibilities are in the set of reasonable choices, and of those only an even smaller fraction are excellent choices. Randomly choosing possibilities is a bad idea because the odds of finding a very good solution are very low.

Tufte (1983)² characterizes graphical excellence as

[It] consists of complex ideas communicated with clarity, precision, and efficiency...[It] is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space...[It] requires telling the truth about the data.

Many graphics packages use default settings which may not be the result of ideal choices. Visualizing uncertainty is also challenging. This page presents some general guidelines on figure design, but these are not to be treated as dogma, as each case is unique, and breaking some of these rules may create more effective figures in certain cases. Ultimately, data visualization is contextual, and the only way to create effective figures is to understand the structure of your data to have insight into how to honestly guide viewers to relevant take-aways.

General Guidelines

1. **Be aware of what can go wrong:** Munzer (2014)¹ provides the following overview of what can make a visualization invalid or ineffective:
 - a. *Wrong domain:* you misunderstood the needs of the viewer;³
 - b. *Wrong abstraction:* you show viewers the wrong information;
 - c. *Wrong idiom or channel:* the way you show the information doesn't work.
 - d. *Wrong algorithm:* the code to generate the figure is too slow.⁴

Healy (2019)⁵ describes the following elements of bad figures: 1. *Bad taste:* figures which are too obsessed with looking flashy or are hard to read; 2. *Bad data:* figures based on data which are cherry picked or which represent the wrong measure for the question; 3. *Bad perception:* figures which inaccurately or misleadingly encode the data.

2. ***Use preattentive cues and features whenever possible without misleading:*** Human image processing and memory works in three stages. First, *preattentive* processing occurs in iconic memory. This is a fast recognition process, can work on many chunks at once, and requires little cognitive effort. Second, a few meaningful visual chunks are passed to short-term (or working) memory for *attentive* processing. Attentive processing can include searching or making conscious comparisons. However, information is retained here for a short amount of time. Finally, information which is repeatedly processed in working memory can be moved to long-term memory, from where they can be retrieved for additional or repeated attentive processing.

Effective use of features which can be processed preattentively can result in more complex patterns being extracted for attentive processing and less cognitive fatigue on the part of viewers. Preattentive features include differences in shape, size, color, movement, and positioning. Think about how to use these to make patterns "pop out" from a quick scan of the figure.

However, be aware that preattentive features can work against you if used carelessly or misleadingly (this is how many optical illusions work). Show your figures to several other people (and/or in the lab meetings) to get their feedback on how much cognitive effort is required to process the figure and to identify if you're communicating the right take-aways.

3. ***The effectiveness of different channels can vary:*** Channels are the ways in which information is encoded within a visualization, such as position, color, and shape. Summarizing several studies, Munzer (2014)¹ ranks channels for ordered attributes by their effectiveness as follows:
 - a. Position on a common scale

- b. Position on an unaligned scale
- c. Length (1D size)
- d. Tilt/angle
- e. Area (2D size)
- f. Depth (3D position)
- g. Color luminance (brightness/darkness)
- h. Color saturation (intensity)
- i. Curvature
- j. Volume (2D size)

Munzer (2014)¹ also ranks channels for categorical attributes:

- a. Spatial region
- b. Color hue
- c. Motion
- d. Shape

The effectiveness of channels can be reduced through interference, where interactions between multiple channels make it difficult to preattentively identify patterns. For example, differences in color saturation are more difficult to identify in small points or thin lines than in larger points or thicker lines.

4. ***Use points and lines thoughtfully:*** A common default is for lines to be used for all visual elements in some types of visualizations, such as time series plots. However, lines invite the viewer to interpolate between points. This is not necessarily appropriate for data points given observational uncertainties and nonlinearities, which mean they should be treated as discrete elements. Use points when individual elements should be treated discretely (or categorically) and lines or curves when you would like to suggest inter- or extrapolation, such as with continuous model output.

If multiple lines are used in a figure, try to rely on their color and position to encode information, rather than the size of areas between them, as area is a far less effective channel. Also, try to avoid relying on line width, both because of the problems with perceiving differences in area, and because small differences in line width are difficult to perceive, while width can only be increased so much before a line appears to be a polygon.

1. ***Use 3D judiciously:*** There are situations where 3D visualizations can be effective, such as when the key takeaway is the broad 3D structure of the data. Often it is more effective to use two spatial dimensions combined with

other channels (but not too many!), or multiple projections of the underlying 3D space onto 2D planes.

2. **Use multiple panels when comparisons across multiple visualizations are desired:** Using the eyes to switch between panels which are simultaneously visible (preattentively) is less cognitively demanding than requiring the viewer to remember previous plots (attentively).
3. **Align panels along common axes:** Position along a common scale is the most effective channel. When multiple panels are required and they have common axes, alignment along one of these axes allows the eye to compare features more easily.
4. **Avoid chartjunk -- when it is chartjunk:** Chartjunk refers to extraneous elements of a plot that aren't essential for communicating the relevant information. Tufte (1983)² argues strongly and convincingly against chartjunk. It complicates preattentive processing and increases cognitive load. However, what is chartjunk is contextual. In some cases, a background grid is chartjunk, but in others (such as multiple panels), the use of a grid or lines may help to align the eye and ease comparisons.
5. **Use color-blind palettes:** For example, avoid combining red and green in a colormap, as many individuals are red-green colorblind and will not be able to identify the intended differences. Some examples of color-blind palettes can be found at [ColorBrewer](#) and [Paul Tol's website](#).
6. **Experiment with different representations of uncertainty:** Uncertainty is difficult to communicate visually. Many different representations exist, including:
 - ranges (*e.g.* intervals or boxplots)
 - distributions (*e.g.* histograms or probability distribution functions)
 - cumulative distribution functions
 - traces of individual realizations.

Not all of these visualization types are relevant for all analyses. Each has its pros and cons. For example, ranges are relatively simple to interpret, but the type and extent of the range must be specified explicitly (*e.g.* 90% credible interval, or the range encoded by the box in a boxplot, as this can be non-standard). Distributions may be less intuitive but include additional information about where probability mass is located. Cumulative distribution functions are useful for comparing or locating quantiles. Try out different plot types and see what works the best for the message you're trying to communicate, and possibly combine different types. Also, try to analyze figures that you find effective.

1. Munzner, T. (2014). Visualization Analysis & Design. Boca Raton: CRC Press. ■■■■■ ■■■
2. Tufte, E. R. (1983). The Visual Display of Quantitative Information. Cheshire, Conn: Graphics Press. ■■■■■
3. Misunderstanding needs is less likely to be a problem when creating figures for papers, as you've designed the figure within the context of the storyline, but can easily be a problem when communicating results to stakeholders. ■■■
4. For our purposes, overly slow code is a problem for reproducible science. ■■■
5. Healy, K. (2019). Data Visualization: A Practical Introduction. Princeton, NJ: Princeton University Press. ■■■

Last update: June 26, 2023