

# The role of web-enabled dynamic, interactive visualizations in neuroscience

## Abstract

### The importance of visualization in science

Visualization is a fundamental tool for analysis and communication in science [Cleveland and McGill 1985]. Visualization serves two primary purposes:

First, we use visualization to quickly make multiple, simultaneous comparisons [Tukey 1993, Gelman 2012]. While it is easy to compare and reason about a few numbers, this becomes more difficult as the amount of data or the number of dimensions increases. Visualization eases the cognitive burden on our working memory by efficiently encoding properties of the data into features salient to the human visual perceptual system [Cleveland and McGill 1985, Card et al. 1999]. This involves exploiting easily processed, pre-attentive visual features such as color, line orientation, and line width as well as higher level perceptual grouping such as symmetry and proximity to reduce the search for information. [Healy et al. 1996, Card et al. 1999, Fetkete et al. 2008]. By doing so, we are able to more accurately and quickly answer questions about the data compared to tables [Spence and Lewandowsky 1991], solve problems related to the data, and make predictions about future data [more citations].

Second, we use visualization to assist in the understanding and checking of statistical assumptions – it helps qualify our knowledge and uncertainty about the data and the procedure(s) used to summarize the data. All statistical summaries rely on assumptions about the structure of the data (our implicit/explicit model of the data), but inspection of summaries alone cannot tell you about violations of those assumptions [Anscombe 1973]. Visualization complements the use of statistical summaries by revealing differences between the expected structure of the data and the observed data [Tukey 1972, 1979]. This is important, from the initial stages of analysis to publication, for revising our assumptions and models and for understanding and communicating where and how often our models fail [Gelman 2004].

A canonical example of this is a set of four datasets known as Anscombe's Quartet [Anscombe 1973].

Each dataset consists of 11 observations of two variables –  $x$  and  $y$ . The  $x$ - and  $y$ -variables have the same mean and variance between each dataset. Within each dataset, the  $x$ - and  $y$ -variables are also identically correlated and fit by the same regression line. However, visual inspection of the datasets reveal strikingly different structure in each dataset (Figure 1). In particular, the second and third dataset (going clockwise around Figure 1) show a suspiciously quadratic pattern and a pattern driven by a single datapoint, respectively.

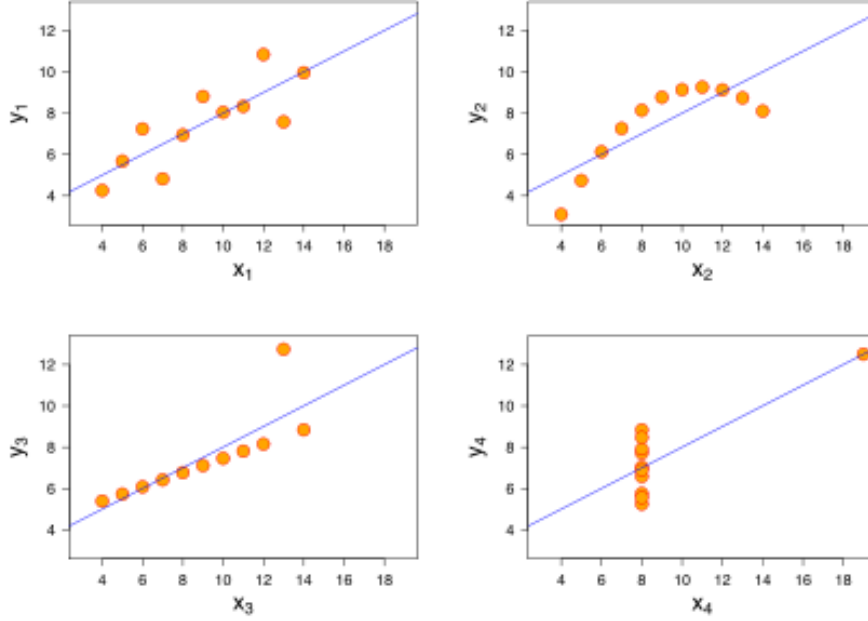


Figure 1: Anscombe Quartet

### Limitations of static visualizations in neuroscience

Static visualizations – visualizations where the state of the visualization cannot change by interaction or animation – have been the *de facto* standard in neuroscience. Static visualizations are used at the early stages of analysis in examining the quality of raw signals (e.g. voltage changes on an electrode, BOLD signals in fMRI), in formulating preliminary hypotheses and in communicating refined analyses in publication. In general, they play a central role in the iterative, sense-making process of data analysis and communication of results.

Advances in technology and computing have made generating static visualizations easier, but those same advances have led to more data, more complex analyses and more complex hypotheses [Freeman 2015]. In electrophysiology, implantation of multielectrode arrays with upwards of 100 electrodes are becoming common [Miller and Wilson 2008, Einevoll et al. 2012?, Boyden ????], Siegel et al. 2015] and the number of simultaneously recorded neurons is projected to double every seven years [Stevenson and Kording 2011]. Laminar recordings have also become standard and add a spatial dimension of data per electrode. Whole brain two-photon imaging experiments in zebrafish can yield up to 1.2 TB of data per hour [Freeman 2015]. In fMRI, the amount of data per brain is already high and there have been efforts to scan more than 1000 subjects [Van Dijk et al 2012, etc].

Consequently, analyses are complex, because with more data, there is greater statistical power to resolve finer differences in the data; we can partition the data into smaller subsets (more dimensions) and still not be overcome by noise [Gelman ???]. With greater ability to resolve differences in the data, we can ask more complex questions and form more complex hypotheses.

An example of this is electrocorticography studies in which grids of intracranial electrodes are placed across wide swaths of cortex. These grids span multiple brain areas and can measure both local field potentials and action potentials (Figure 2a). Given enough data, this allows us to ask questions about the properties at different spatial scales (multiunits, local field potentials, brain region summaries) and how they relate (e.g. correlation and coherence between spatial scales and between brain regions, see Figure 2b). Moreover, we can ask questions about how these change over time and/or relate to experimental conditions. This results in a high-dimensional partitioning of the dataset with many complex interrelations.

So why is this a problem for static visualizations? Complex analyses and hypotheses necessitate either an increase in the number of static visualizations or further summarization of the data (dimensionality reduction) to deal with the number of dimensions – often both. For example, a common visualization of spiking data from a neuron would be a raster plot (Figure 2a) or as a summary – a histogram (Figure 2c). These are typically visualized with respect to a particular experimental stimulus or event (e.g. a saccade); each event requiring a new visualization. To investigate the firing rate of 1000 neurons in multiple conditions implies visualizations for each neuron (Figure 2d) or aggregation in some form (e.g. summarization by brain area, see Figure 2e) once the visualization becomes too ineffective to support perceptual comparison. Because analysis is an iterative process, a typical analysis might require hundreds of visualizations as different sets of experimental conditions are examined or as more data is added.

A large number of static visualizations results in more time spent switching between visualizations, which can not only extend the time to analyze the data, but also has a meaningful impact on our ability to explore and understand the data. For example, Liu and Heer (2014) found that even a 500 millisecond delay between visualizations could reduce the amount of the dataset explored and affect the number of hypotheses and observations formed. Similarly, Brutlag (2009) found that users performed fewer web searches if there was as little as a 200 millisecond delay in the return of search results.

Summarization, while often necessary, can obscure complexity and variability in the data – as in the case of Anscombe’s dataset. Furthermore, summarization does not obviate the need to understand and check statistical assumptions. This can be a problem particularly with large datasets, where going back and forth between raw data and summaries is particularly difficult because of the amount of data.

**Interactive visualizations can help us quickly make comparisons and deal with complexity**

**Dynamic visualizations can help us understand large quantities of data**

**Web enabled visualizations are easily shareable**

**Prior work**

- [Jeremy Freeman](#)
- [pycortex webGL fMRI viewer](#)
- New York Times

**Goals of our visualizations**

**Case Study: RasterVis**

RasterVis is a D3-based visualization tool for quickly viewing, grouping and summarizing spike rasters for many neurons.

This tool allows you to: + Generate and change between rasters for many neurons. + Quickly view rasters aligned to experimental trial events. + Add Gaussian-smoothed peristimulus time kernel density estimates with arbitrary smoothing. + Group spikes based on experimental factors.

**Case Study: SpectraVis**

**Case Study: glmVis**

**Future Challenges**

**References**