BACKGROUND NOTE

Creating effective data visualisations

By Jeff Knezovich | June 2017

Although think tanks use a wide variety of research techniques, recent technological advances have put a focus on working with big data. But working with large datasets – not just so-called 'big data' – can be challenging. For example, it's often difficult to make sense of that much raw data. It's a classic 'wood for the trees' problem – in other words, having so much detail that it's difficult to see the bigger picture. That's one reason why think tanks are investing heavily in data visualisation capacities and techniques. Not only do data visualisations potentially help support the research process itself, but they can also help to communicate large quantities of information to a wider, less technical audience. But what does this look like in practice for think tanks?

Data versus information

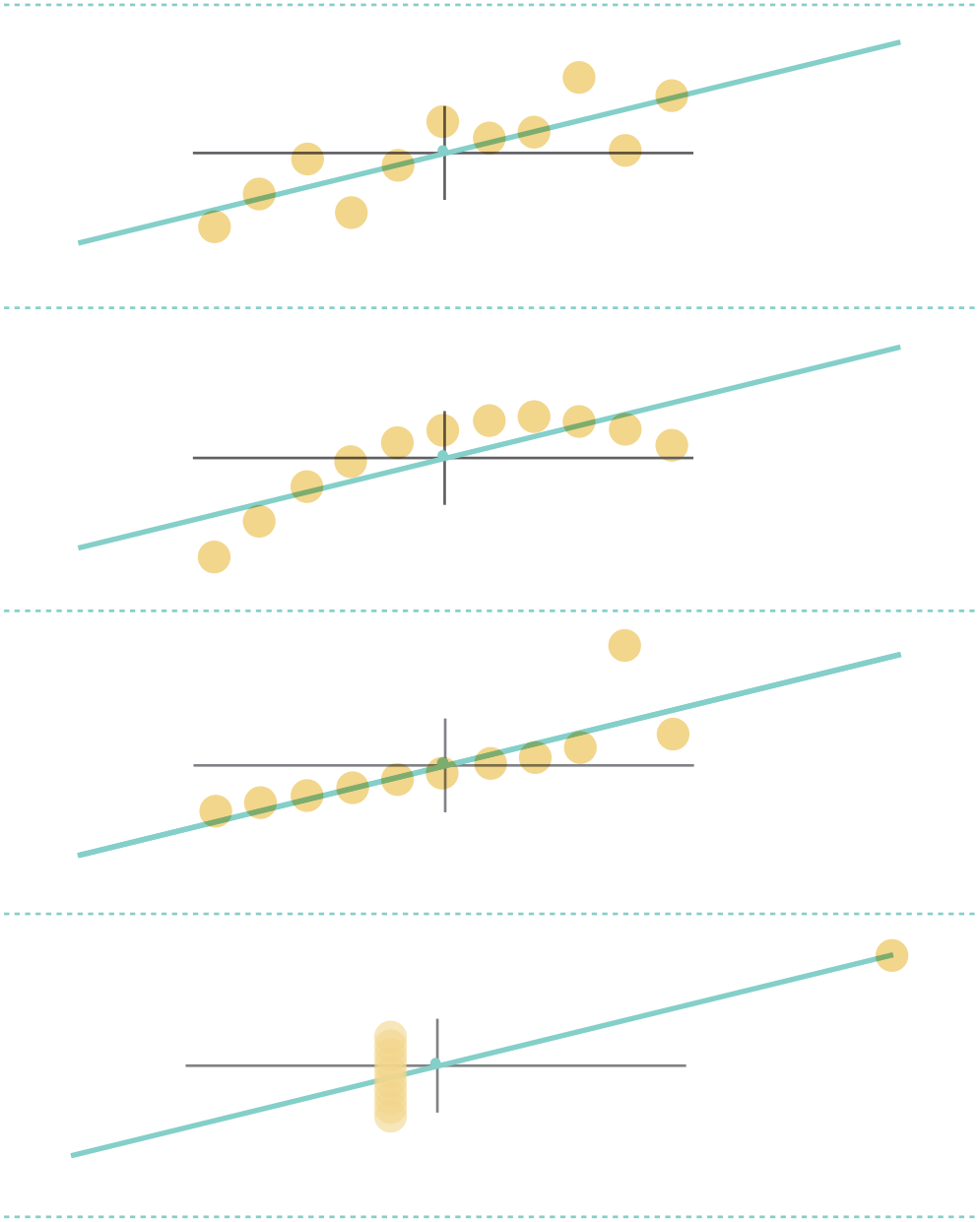
Data, by definition, are the lowest level of abstraction – the bits that make up the byte. That means that data are often incomprehensible on their own. In order to understand data ¬– to turn it into information – it is usually important to abstract it further, to read across multiple data points, possibly even multiple data sets.

One way we do this is through statistics. Looking at a data set, we might start describing it by looking at the average value. We might compare that to the median value to see how our data are distributed. We may even add standard deviations to gain a better understanding. But these descriptors still only give us part of the picture. Consider these four datasets:

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

In each case, the mean of x is 9 and the mean of y is 7.50. And if we want to get into fancier statistics, the correlation between x and y is 0.816. They also share similar variance. All these statistical descriptors are the same, so they must all be the same data set, right?

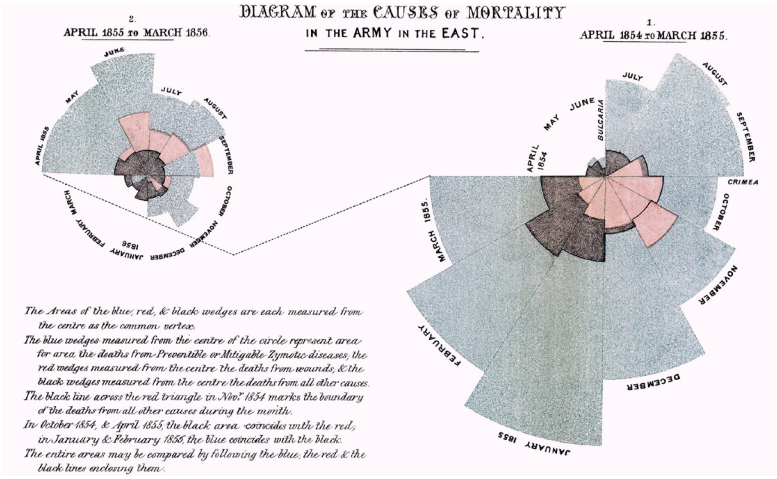
Absolutely not. Known as Anscombe's quartet, when visualised, the differences among the data sets become immediately apparent:



And that's the power of data visualisation – they are a very high level of abstraction that allow us to **immediately understand relationships between data points**. Visualisations help us to spot patterns, trends and outliers. They put the data into context by presenting them at the same scale. More advanced visuals can help us associate data through colour and shapes too.

Data visualisations and policy influence

The ability of data visualisations to effectively communicate data make them a powerful tool for think tanks. Indeed, some big changes have been attributed to the power of data visualisation. Take, for example, Florence Nightingale's 'coxcomb', or 'rose' diagram of the causes of British military deaths during the Crimean War:



The diagram shows that most of the deaths were caused by preventable disease, and not battle wounds. She used it as a lobbying tool with British parliamentarians to argue for better sanitary conditions – initially in the army, but eventually back on the home front, too. It contributed to the establishment of modern nursing and better sanitary conditions across the UK, as codified in the Public Health Acts of 1874 and 1875. One example of many of data-driven policy influence. We will share more of them during the course.

What kinds of visualisations are there?

The term 'data visualisation' refers to the process of creating graphical representations of data, both quantitative and qualitative in nature – though quantitative data lend themselves particularly well to visualisation. But the type of the data doesn't necessarily dictate what format these visualisations ultimately take.

One way to think of different types of data visualisations is by the format they take and by their level of user interactivity – either static, motion or interactive.

- **Static visualisations:** These visuals don't have any interactive elements, and therefore tend to be thought of as some of the simplest forms of visualisation. Static representations of quantitative information tend to be charts of some sort – for example pie charts, line charts or bar charts usually present in traditional papers. But a static visualisation may not be limited to a single chart. They may include a set of charts in a panel. Or they may be sets of icons, especially for more qualitative data. Or they may be a combination thereof, which are often referred to as infographics. For an example of a think tank infographic that combines different types of static visualisation, consider Don't limit her possibilities, from JumpStart Georgia.
- **Motion graphics:** Motion data visualisations are designed to tell a specific story with data. They generally walk a user from point a to point b, explaining or interpreting the data as they go along – this is often done with an audio track, but sometimes this may be done through text. These might take the format of videos or animated gifs, which are particularly useful for social media and in other web content. A great example of this type of visualisation is Visualizing the past, present and future of carbon emissions by the World Resources Institute.
- **Interactive visualisations:** Sometimes it is best to empower users to discover their own story through data – this is where interactive visuals are at their best. As the name implies, these visuals change and react to user input. Clicking on one part of a chart may change another part of a chart – either extending it, zooming in or out, or filtering the data shown. Many think tanks have used this approach to great effect. One good example is Mapping Czech crime, by Otevrena spolecnost.

The table below gives a high-level overview of the characteristics of these different types of visualisation.

Static	Motion	Interactive
Tells a clear story	Tells a story	Users 'create/discover' their own story
Linear	Linear	Multiple pathways
Doesn't move or change	Moves/changes by itself	Moves/changes based on user input
Useful for highlighting key facts	Useful for explaining more complex data	Useful for allowing users to explore large data sets
More icon based	Mix of icons/images and charts/maps	Often based on charts or maps
Simple charts are found throughout standard think tank publications. Infographics have been more typical for posters and social media content.	Motion can be employed in different ways to tell a story, but by definition are limited to digital outputs. These are great videos, but can also be a great way to introduce interactive visualisations.	Most often found as a standalone page or microsite for a think tank project or report. These can also be used as dashboards for real-time monitoring.

What goes into an effective data visualisation?

At the heart of any effective data visualisation should always be the end-user and the objective of the visualisation; in that sense, data visualisations are not that different from any other communication output. As we will cover in more detail in the course, data visualisations are targeted at the subconscious mind and are designed to not only be immediately comprehensible but also aesthetically pleasing. This can be particularly important when working to engage policy-makers, as they tend to have limited time and short attention spans.

Given this, the need to adhere to these basic communication principles in data visualisation is perhaps more important than in some other contexts, if only because they try to convey as much information as simply as possible using both text and graphics. It's a concept that Edward Tufte, one of the 'fathers' of modern visualisation, refers to as 'data-ink ratio'.

> "Data-ink is the non-erasable core of a graphic, the non-redundant ink arranged in response to variation in the numbers represented."
> Edward Tufte

Striking the right balance between adding the information and context needed for the intended audience to understand the visualisation and keeping it simple enough to let the data speak for themselves is the key to creating effective data visualisations. And if there are more complicated objectives that require more complex visuals, ensuring that the user interface ¬– the buttons, the sliders, the filters, the input boxes ¬– make sense to the user is critical.

If this all sounds a bit difficult, that is because it is. And to make matters worse, the skills required to create effective visualisations for think tanks are diverse and not often found in the same person.

As I've noted before, there are four main skills needed to produce visualisations:

- **Research:** This includes everything from strong data literacy skills to strong understanding of the context from which the data spring. In terms of data literacy, this means being able to merge and tidy datasets, as well as knowing what sorts of statistical analyses are appropriate to run. And in terms of context, it's about having knowledge about the area of study – whether it be social or cultural context, the political environment (as are more typical in development studies) or the physical, biological or chemical processes at play.
- **Technology:** Technology plays a role in all parts of data visualisation: from collecting and scraping data, to processing and storing data, and straight through to drawing charts and graphics. Some technologies require more skills than others. Some might require mark-up or coding. Some are simple drag-and-drop interfaces that make visualisation easy. We will cover a range of technologies during this course, mainly focused on the visualisation part of this process. No prior coding knowledge will be required, but it may be helpful!
- **Design:** In terms of visual design, it might be about knowing the appropriate types of visuals for the data or about understanding chart design fundamentals. But it's about balance and flow. And it's also about appropriate use of colour, typography and other visual cues. And as for user experience, there are a number of elements to get right, from navigation to information structuring.
- **Communication:** The number of visualisations I've seen that either don't have a clear message or a clear purpose is staggering. Communicating the right messages, or, in the context of interactive visualisations, understanding how to nudge users in the 'right' direction as they explore visuals is very important. At the same time, it's important not to over-simplify of misconstrue the data.

For think tanks, building the capacity to create effective data visualisations is often about finding and creating teams that collaborate across these four areas. It's difficult to find a single person who is strong at all these skills!

However, one of the benefits of the modern explosion in data visualisation are the tools and technologies that support people to create these visualisations. Don't have a design background? That's ok, some tools like Canva or Carto will make your data beautiful. Don't have lots of money to create fancy visuals? That's ok, there's a lot that can be done with Excel and GoogleSheets. Don't know how to code? That's ok, TableauPublic uses a graphical interface to allow you to explore your data. Already know the fundamentals of visualisation and looking to make bespoke, interactive visuals? There are coding libraries that allow for more refined user engagement.

This will be an entry-level course, so we're not expecting you to have all these skills already. That said, based on the description above, it would be good to start thinking about the following questions:

- What are the balance of these four skills in my think tank?
- Which should we focus on if we want to build our capacity for data visualisation?
- Are there any of these skills we are completely lacking?
- Are there ways we can get around not having some of these skills in house, and if so, which are the critical ones for us to actually have?

These will inform our discussions during the course and may provide you with new ideas on how to move data visualisation to the next level.

What comes next?

In this course we will look at how to produce effective data visualisations to support research and communication. We will examine more closely what types of visualisations there are, and what makes a good data story. And most importantly, we will look at different tools for developing visualisations, turning theory into practice. Everyone will have the opportunity to make their own, based on the tools we discuss and will get personalised feedback on their submission.

This course will start online in our discussion forum. You are invited to briefly introduce yourself to your fellow students. We ask you to include – either by link or attachment – an example of your favourite data visualisation and a bit of reasoning behind why that is.

This course includes three webinars that will last around one hour each. The first one starts with further discussion about what makes a good, or effective, data visualisation. It will be slightly more theoretical in approach and will also provide an overview of design best practices.

The second webinar will move straight into the practical side of things and will learn how to structure and manipulate data in Excel and Google Sheets. We will look at how to improve the charts produced using these common technologies to make sure they adhere to the design lessons covered in the first webinar.

For the third webinar, we will look at how to make interactive visualisations with Tableau Public, covering the basic concepts and elements of Tableau Public and an example of creating an interactive dashboard.

We will be providing samples and datasets for you to work along with as we go along. But it would be good to be thinking about what data you have available, as the final practical exercise for the course is the development of a data visualisation of your choosing based on the tools and techniques learned during the course.