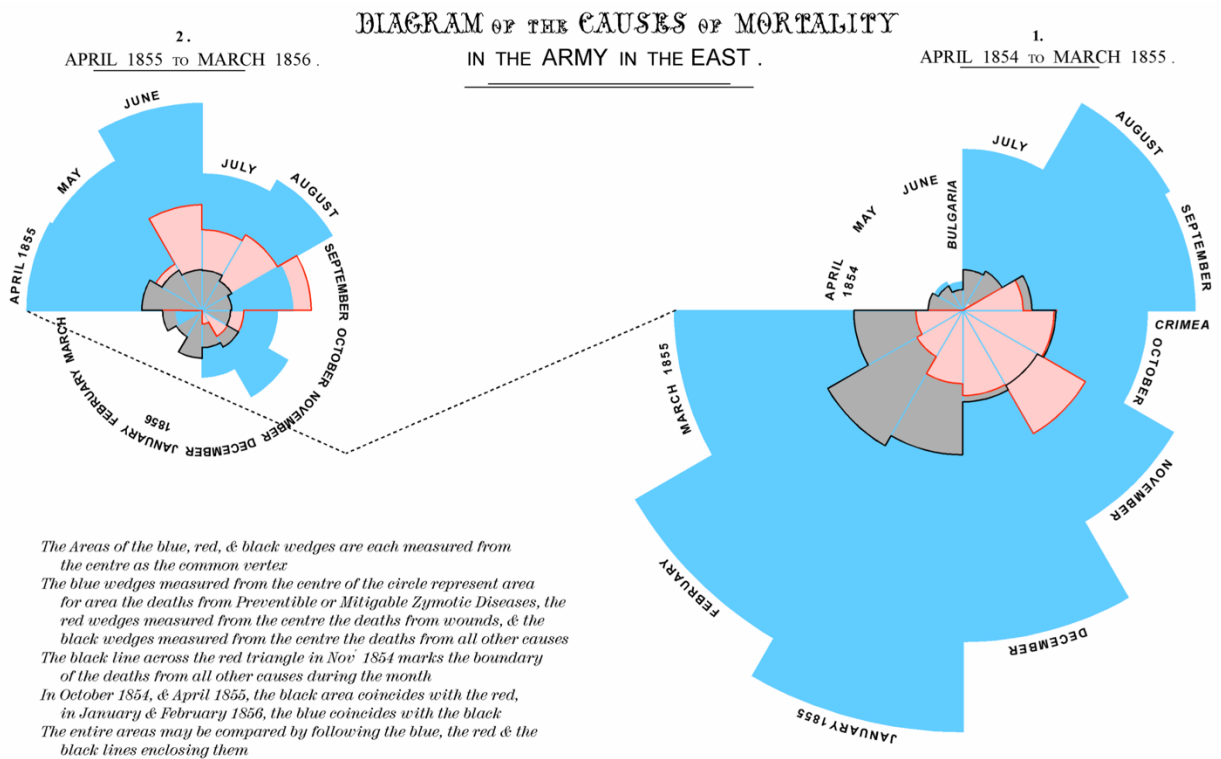


Title: Data Visualization as Narrative

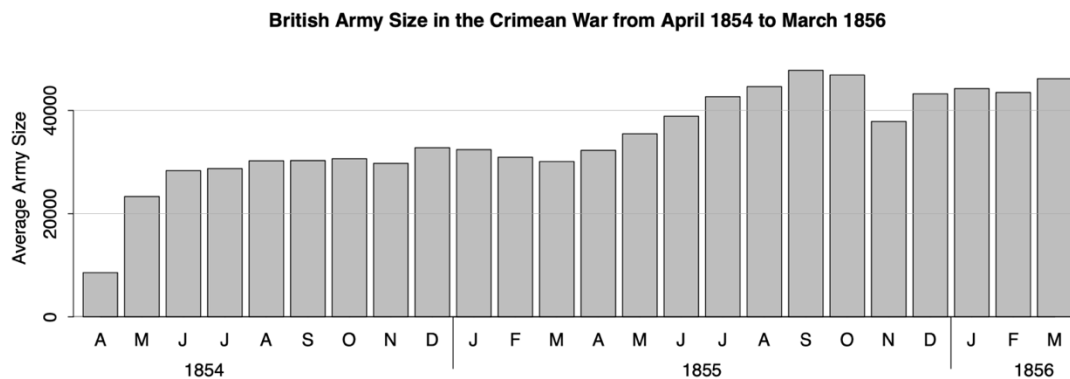
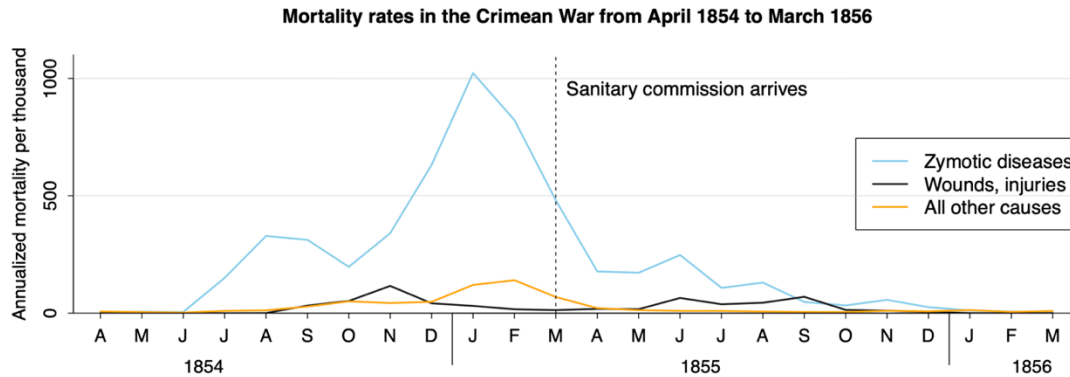
By: Andrew Gelman and Helen DeWitt

Data visualization, as Alberto Cairo has written, is a functional art. Scientists and policy analysts spend a lot of time visualizing data – firstly, to understand the data they are collecting and analysing; secondly, to explain their results to others.

Here is a classic display of British casualties in the Crimean War (1853–56), constructed at the time by the statistician, and founder of modern nursing, Florence Nightingale:



Here is a modern redrawing:



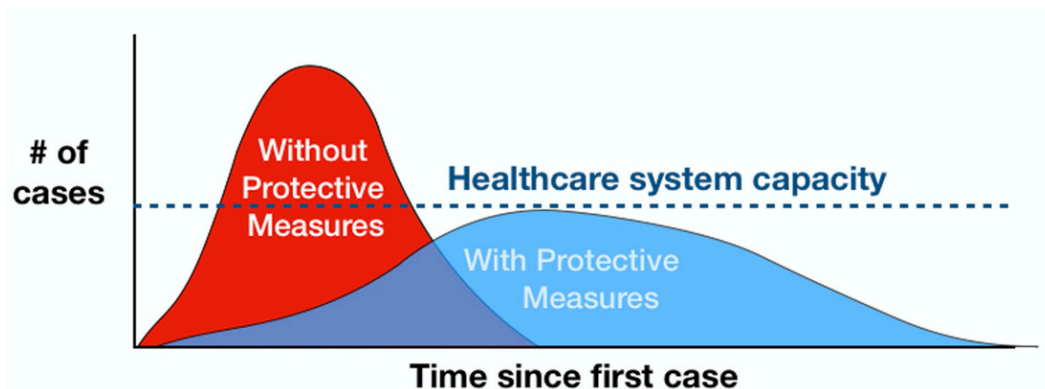
The modern version more clearly shows the patterns in the data: for example, that the disease rate did most of its decline *before* the sanitary commission arrived; this key bit of timing is less apparent in Nightingale’s graph. But the beauty of that century-and-a-half-old visualization serves its own function, by highlighting that most deaths in the war were the consequence of disease and, most importantly, by drawing the reader into the story. The earlier visualization succeeds, paradoxically, by presenting something of a puzzle that invites reader input. Neither display is ‘better’ in an absolute sense.

One implication of this division of labour between clarity and drama is that the search for exemplary or beautiful visualizations can be counterproductive. You can’t ask scientists, statisticians or graphic designers to devise their own versions of Charles Minard’s classic 1869 graph of Napoleon Bonaparte’s failed Russian campaign (1812–13) – a version popularized in the late 20th century by Edward Tufte – any more than you would want writers of scientific papers regularly to produce works of art: not all stories lend themselves to this sort of depiction.

It is often the more challenging visualizations that are the most appealing. This suggests that active reader participation is a key part of the visualization experience. Nightingale's circular graphs have puzzles (decoding each line), plots and subplots. They even feature allusion and metaphor: the 12-month annual cycles are reminiscent of the 12 hours on a clock face and suggest the circularity of time – even though, from a statistical standpoint, they are a distraction in this particular plot, given the lack of annually-recurring cycles in the actual data.

If we consider data visualization as narrative, then what, exactly, is the story behind a static graph? Arguably, it is a narrative of discovery that engages the viewer in a theory-meets-data, theory-loses-data, theory-gets-data storyline mirroring the scientific process. It is said that most of the action in a comic strip occurs between the panels; similarly, the active part of a graph takes place not on the page or screen but in the viewer's mind. Thus, even a static visualization has a temporal dimension, with the narrative unfolding during the reader's process of understanding.

In science fiction we talk about a 'sense of wonder', and it can often seem this is what scientists and graphic designers are aiming for in their data visualizations. But let's forget a moment about wonder. Instead, let's talk about urgency.



One of the central images of the COVID-19 pandemic has been the 'flatten the curve' graph, which has been humanly compelling even though it is not an image taken from real life. Versions of this graph have been widely circulated, showing the relation between healthcare resources and two curves: one the likely rate at which the virus would spread without social distancing or other control measures; the other, the flattened curve. There have been criticisms of this visual as

oversimplified, but it really did make vividly clear why the rapid move to lockdown was seen as so urgent, why slowing the rate of infection was a matter of life and death, and why, from one week to the next, we went from self-regulated social distancing at the supermarket to compulsory masks, one-way aisles, plastic barriers at the tills and floor markers for correctly spaced queueing. In the meantime, the news was full of stories about the shocking shortage of PPE in hospitals and the shortfall of childcare for hospital workers: in other words, the horizontal line that supposedly represented what the healthcare system could accommodate was itself contingent on behaviour and policy.

When it is a matter of life and death, getting data viz right – getting the data right and communicating it intelligibly – is an urgent matter. No doubt this is what impelled Nightingale to communicate by inventing a graphic representation: to our eyes it may look delightfully bonkers, but it was prompted by the horrific levels of non-combat deaths in the Crimea, about which something could be done. Given the magnitude of the issue in question, Nightingale's graph is too complex to be readily and effectively deciphered – a modern data visualist should reject her approach out of hand for the goal of conveying trends in the data. Yet, posterity still deems it a classic, its potency rooted in its ability to convey a sense of urgency in an unusual way.

So we have a vague sense of unease. In the current moment, people who have never previously given much thought to data visualization will see information presented in narrative form without fully appreciating how the effectiveness of its communication can, in some cases, be a matter of life and death. At the same time, we would like people who've never thought about it to be converted from easygoing accepters of data viz as a kind of scientific-looking ornament, to critical and exacting readers capable of being moved to urgency by the visual narrative of expectation and surprise.