



Reading to write

Deborah Nolan and **Sara Stoudt** present a framework for learning the art of statistical storytelling

We've cleaned and wrangled the data, we've made picture-perfect plots, and we've performed our statistical analysis. Now it's time to share our findings with the world. This is a step that novices and experts alike balk at. That blank screen of our future manuscript is intimidating.

Instead of staring at a blank screen waiting for inspiration to hit, we can take action. In fact, even now, as you read this magazine, you can get ready to write. Learning how to write can start with reading data-related articles and examining how the author organises and writes about their findings. When we “read to write”, we identify the main components of a data analysis and notice how the author brings these components together to form

a logical and compelling story. During this deliberate process, we discover examples and templates that we can use to organise our own work and write about our findings.

It can be daunting to develop an “ear” for writing. However, reading to write is a skill that can be honed through practice. To help you get started, we have developed a three-step framework. We walk through each of these steps as we read “Nightingale’s overlooked Scutari statistics”, an article appearing on pages 29–33 of this issue of *Significance*.

Reading once, twice or more

Whether we read to learn how to write or to learn about a new study, we typically don’t read an article straight through from beginning

to end. Instead, we make several passes over the article, each time understanding it in greater detail. We like to begin by paging through to see how long it is and reading the section headings. Knowing this helps pace our reading; it gives a sense of where we will find the various parts of the story. Next, we often book-end the article and read the abstract, introduction, and the conclusions to get an understanding of the core of the argument. After that we examine figures and tables and their captions to uncover the basic statistical argument. Finally, we read the paper in order from beginning to end. In this final round, we already have a sense of the organisation and main gist of the argument; now our focus is on how well the authors convince us of the correctness and value of their findings.

Map the organisation

Paging through “Nightingale’s overlooked Scutari statistics”, we find that the submitted version (before design and layout) covers eight pages of A4 paper. There is an introduction and then a series of section

headings: ““The operative causes of our great mortality”” (which seems to give a description of Nightingale’s data); “Policies on evacuation to Scutari” (possibly a background section, which interestingly is placed after the data description); “The sick at Scutari” (appears to contain the heart of the article); and “The wider context” (probably a discussion section). These headings provide a roadmap through the article and give us more context than traditional names, such as introduction, background, and findings. They tell us that the author investigates Nightingale’s data, brings in historical records, and addresses a particular question about the condition of

sick soldiers transferred to Scutari. Sketching this roadmap, as in our Figure 1, can help the novice writer organise their own work.

Next, we skim the introductory and concluding sections to learn more about the topic: the author’s aim to investigate some of Nightingale’s lesser known data and argue for its importance. Specifically, the author explores Nightingale’s perhaps controversial conclusion that the higher death rates at Scutari hospitals were due to overcrowding and poor hygiene.

The abstract and introduction are often the hardest to write, and the most important because a reader uses them to decide whether or not to keep reading. Notably this

introduction grabs our attention by associating the author’s work with the signature plot of Nightingale’s (“her polar area chart”), despite the plot not being directly relevant to the analysis in this article. The introduction starts with a bigger picture overview of Nightingale’s investigation and then further narrows to the task at hand, a particular annex of her report concerning regiment mortality. In the conclusion, the author takes the opposite approach, going from the article’s specific takeaway to the broader implications of Nightingale’s findings. Her campaign improved civilian life expectancy because it led to improvements in overcrowding, drainage, water supply, and ventilation.

Now we take another pass over the article to “read” the tables and plots and assess how well they support the author’s analysis.

Identify statistical elements

After we’ve obtained an overview of the article’s structure, we look for the main building blocks of the data analysis. Some of these elements are explicitly included in the article; others are not. This investigation helps us understand the choices that an author makes in writing about their analysis, and as a reader, we can reflect on whether particular omissions impact the credibility of the conclusions or whether any included details are superfluous to the main story.

The author has included some of their data in a table in the “operative causes” section of the article. He could have simply provided the scatter plot in Figure 1, but we appreciate the ability to peruse the raw numbers. We do wish the author included more information, such as the date of transfer and the reason for the order of the rows. Nonetheless, his Figure 1 confirms the impression we get from glancing through the numbers in the printed table (and the full online table): nearly all regiments have a lower death rate in Crimea than in Scutari.

The “Policies” section contains a bar chart and a line plot, Figures 2 and 3, respectively. The bar chart splits the soldiers sent to Scutari according to whether they were wounded or sick. This chart shows us that the vast majority of transfers were sick, and sick soldiers had a much higher death rate compared to the wounded. The author has identified a confounding factor.

At first glance, Figure 3 looks a bit overwhelming. It contains two line plots on

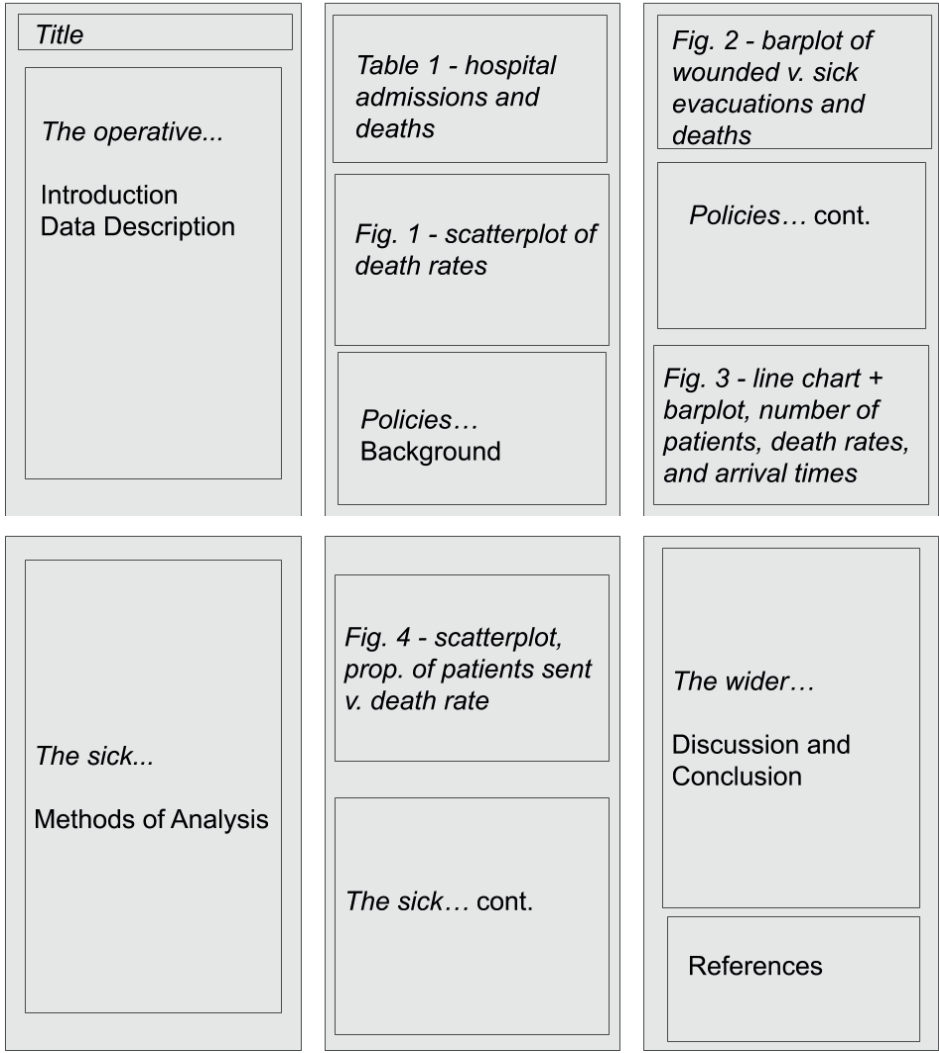


Figure 1: Map of “Nightingale’s overlooked Scutari statistics”. By drawing a schematic of an article, we can see at a high level where different components of the argument are placed in relation to one another. We can also see proportionally how much space is devoted to different components.

► different scales (the left axis corresponds to the number of patients and the right to the death rate). Furthermore, the x-axis marks dates, but their distances are not proportional to the gap in time, and above these dates are bars. Fortunately, the legend and title help us parse the graph: time is on a nonlinear scale and the height of each bar corresponds to the number of wounded transferred on that date. Sometimes, informative plots take time to decipher. After further reading of this plot it becomes clear that the wounded arrived when hospitals were much emptier, and the death rate tended to be lower at those times. We also get a sense of the seasonal variation; in late January and February, the death rate was at its peak.

The final figure (Figure 4) seems a bit curious because the scatter of points does not reveal a pattern. We notice that after this plot, several correlations are provided (between percentage of patients transferred to Scutari and death rates in Crimea, and between death rates in Crimea and at Scutari). These correlations are accompanied by *p*-values, but aside from Figure 4, plots of pairs of these variables are omitted. We think this figure could also be left out without taking key information away from the reader.

Figure 3 appears to contain the core message. Now that we have looked over the plots, we have enough context to read the details of the argument.

Examine the argument

Many of us may think that since science is objective, we don't need to make an argument, we simply need to present our findings. However, the Nightingale article definitely argues a case.

An article is an extended argument, and as with any argument, there are (at least) two sides: ours and others'. The Scutari article presents evidence in the context of a debate over Nightingale's examination of hospital quality and particularly whether a direct comparison of mortality in two different sets of hospitals, those in Crimea and those in Scutari, is justified on the basis that one set of hospitals (Scutari) may have "received the most serious cases".

An author's choice of words can impact the strength of their argument. Sloppy writing can easily lead the author into making stronger claims than are supported by the analysis.

When reading, we pay close attention to the precise wording used. We expect the author to accurately describe their findings and demonstrate a balanced understanding of the implications. That doesn't mean that the writing must be dry and formulaic. The Scutari article uses vibrant imagery: Wyatt's medical reports contain "harrowing details", Nightingale's work is likened to a modern-day "whistleblower", the sick soldiers had a "nightmare journey" from Crimea to Scutari, and were cared for in "pest-houses" for hospitals. This language makes for interesting reading, but we recognise that it also sets us

in a frame of mind to be sympathetic towards Nightingale's claims.

Since a writer does not directly converse with the reader, they must foresee a reader's potential objections, questions, and comments and address them ahead of time. As well as anticipating possible pitfalls, we anchor our claims to that of others. If readers don't know what to compare our argument to, they may not remain convinced when presented with other work that contradicts ours. In the Scutari example, the author makes his argument by piecing together historical records. He points out that no source is cited for the

The "Reading to write" template

Use this template to guide you in reading articles relevant to your own field. Knowing what to look for will help you keep track of lessons learned from other writers. Your notes can then be referenced when you are ready to write.

Map the organisation

Make a map of the article and mark the locations of the objective and motivation, data description, methods, statistical analysis, graphs and tables, related findings made by others, discussion of limitations, generalisations, and impact, and conclusions.

- **Where is context provided (e.g., do related findings motivate the topic or are they provided later in a discussion)?**
- **How much space is dedicated to each type of analysis (graphs, tables, statistical findings)?**
- **Is there one key graph? Where is it located in the paper? What message does it convey?**
- **How often and where is the impact of the study discussed?**

Identify the statistical elements

- **What data are used, including, as appropriate, study design, protocol for data collection, and non-response?**
- **What statistical graphs are provided? Do they address exploration, formal analysis, or comparison and assessment?**
- **What are the main outcomes of the analysis?**
- **Were any details of the analysis excluded?**

Examine the argument

- **How do the results presented relate to the rest of the scientific literature? Do they support existing understanding, fill in gaps, offer evidence in a debate, or provide a new approach or new evidence?**
- **How thoroughly are findings described? Does the level of detail match the audience?**
- **Does the analysis support the conclusions? Do the author's claims seem adequately justified, overblown, or unduly cautious?**
- **How does the level of detail in describing the findings and impact vary across the article?**
- **How do the specific word choices strengthen or weaken, clarify or confuse, and support the argument?**



Deborah Nolan holds the Zaffaroni Family Chair in Undergraduate Education and is professor of statistics at the University of California, Berkeley.



Sara Stoudt (@sastoudt) is a statistician at Smith College with research interests in ecology and the communication of statistics.

assertion that Scutari hospitals received the most serious cases, while presenting documentation that at least one hospital, at one point in time, seems to have kept its sickest patients in Crimea. However, he also discusses other documentary evidence that shows that at least one army regiment did send its sickest patients to Scutari.

The author brings additional data (albeit historical records) to make his argument; his analysis is placed in a broader historical context by combining both qualitative and quantitative information from letters and other historical records. A discussion of the data and methods may seem fairly standard and objective, but the careful description of the provenance of the data, detailed records for two regiments, and descriptions from the medical records are essential to the argument. This historical context lends credibility to the analysis and is woven into his examination of statistical graphs.

The author intersperses his defence of Nightingale's original claim with Nightingale's own commentary, allowing her to weigh in and defend herself. The author also systematically states possible ways for Nightingale's claim to be invalidated and goes through each in turn, using a historical reference or a plot and pausing to clear up other questions that arise. In some cases, he concedes that an "explanation cannot be verified". He is forthright about the shortcomings of the missing data.

Let's step through a particular paragraph as an example:

Why was the mortality among the (seriously) wounded at Scutari so low compared to that of the sick? Figure 3 suggests that this discrepancy was at least partly due to the relative timing of the arrival of wounded and sick. The great majority of the wounded arrived in November, when the Scutari hospitals were not full. The death rate was low. After the fighting stopped, and winter set in, the hospital became overcrowded with sick, not wounded, men, and the death rate rose dramatically. This is consistent with Nightingale's statement to the Minister of War, quoted earlier, that the main cause of "our great mortality" was overcrowding which aggravated the sanitary problems caused by inadequate drainage and ventilation. The correlation between

hospital population and death rate in Figure 3 is 0.78 ($p=0.014$).

The first sentence anticipates a question that a reader may have after looking at Figure 2. The next four sentences answer this question, pointing the reader to the important takeaways from Figure 3. The penultimate sentence connects the findings back to Nightingale's own work and includes a quote from Nightingale herself. The final sentence of the paragraph adds a quantitative finding to support the qualitative argument made in the rest of the paragraph.

The author investigated multiple ways that Nightingale's argument might be weakened and found no evidence to cast doubt on her claim. However, he is careful not to overstate his findings, reminding the reader that "the data and evidence we have considered here do not allow us to rule out the possibility that the most serious cases were transferred to Scutari".

Discussion

When we read an article, we expect to be convinced of the importance and validity of the findings. We consider the appropriateness of the analysis, the generalisability of the conclusions drawn, and whether others' work has been adequately and convincingly presented. Do the article's insights support, counter, or extend current views in the field? To help make these assessments, we pay close attention to the writer's word choice and tone. We look for convincing language that doesn't reach too far beyond what the data can support and sections that work together to tell a credible and cohesive story.

We have created a template that you can use to organise your thoughts as you read (see box). Our approach and template aim to help you tease out how a writer makes their argument. Graff and Birkenstein also provide a framework for examining the structure of written arguments,¹ and in Chapter 13 of their book, Gillen addresses the topic of making arguments using data.²

Others have also advocated learning to write by reading the works of others. Heard suggests that you: offer to read a paper for a peer/colleague; participate in (or create) a group to read a paper together and discuss the writing as well as content; and extend the development of your writing "ear" by reading beyond science topics (e.g., news stories and

popular magazines).³ Like Heard, as we read, we often save snippets of writing we like, and we take care to include the source of this material so we don't accidentally forget those words are not our own (and thereby avoid inadvertent plagiarism).

As you read through this article, you might have noticed that *Significance* is not a traditional research journal. Contributions are expected to be written in an accessible and engaging style. When you apply this read-to-write framework yourself, we suggest reading articles in publications where you plan to publish. That way you can acquaint yourself with the interests and knowledge of the readers you want to reach so that you can help find a good match for your intended audience.

Conclusion

Researchers often think that since science is objective, they do not need to craft an argument for their findings. However, an effective communicator must convince others that their approach is well founded and that their final message matters. By recognising technical arguments in others' work and deconstructing an author's approach to making a compelling case for the validity and importance of their work, we gather a toolbox of strategies to wield ourselves. We can then use these tools to face the blank page with confidence. Happy reading and writing! ■

Note

Deborah Nolan and Sara Stoudt are the authors of an upcoming new book, *Communicating with Data: The Art of Writing for Data Science*, to be published by Oxford University Press.

Disclosure statement

The authors declare no conflicts of interest.

References

1. Graff, G. and Birkenstein, C. (2010) *"They Say / I Say": The Moves that Matter in Academic Writing* (2nd edition). New York: W. W. Norton.
2. Gillen, C. (2010) The data suggest: Writing in the sciences. In G. Graff and C. Birkenstein (eds), *"They Say / I Say": The Moves that Matter in Academic Writing* (2nd edition). New York: W. W. Norton.
3. Heard, S. (2016) *The Scientist's Guide to Writing: How to Write More Easily and Effectively throughout Your Scientific Career*. Princeton, NJ: Princeton University Press.