Day 2: Openings

Reading *Writing Science* by Joshua Schimel

Before workshop

1. Reach Chapter 5 of ***Style*** (online version)

Additional reading:

- Chapter 10 — *Writing Science in Plain English*

2. Identify thematic strings in a paragraph from an example paper

> "Even if you use Twitter only for professional purposes, consider opening up a little bit to show your followers you are a real person. People outside your field are not likely to follow you if your tweets are only about sharing events, articles, and positions in your own field. You need to add an extra ingredient—your opinions, or something personal—to what you share. One way to do this is through sharing failures: a rejected paper or job application, or even a spilt coffee. This is a great way to give and receive moral support from other academics."

I find this paragraph decent to read; it could be improved.

Characters I remember:

- real person
- followers
- your failures
- opinions

Conceptual center:

- persona
- likable
- relatable

These thematic strings that I recalled are indeed similar.

3. Identify thematic strings in a paragraph from your manuscript

3a. A paragraph that "hangs together"

> Those who have used R packages may know that, although leveraging existing tools can be advantageous, the initial challenge of finding a suitable package for a given task can obstruct potential benefits. Relatedly, new R users who are unfamiliar with the structure and syntax of the language may be hindered by the process of finding packages because they do not know where to search, what to look for, or how to sift through options. R packages are mentioned in a variety of places online, in print, and elsewhere. You can discover new packages any time you learn R-related topics, collaborate with other R users, or browse the internet.

Characters I remember:

- experienced R users
- new R users
- R packages

Conceptual center:

- searching

3b. A paragraph that seems vague

Old

A developer of an R package may intend for it to be private (exclusively for personal or professional use) or public (free and available for use by anyone) @rickert2018. If your task is specific to a line of research, consult colleagues to see if they have relevant (private) code they would be willing to share. Alternatively, literature in your field may either introduce R packages developed to solve a unique data science problem or mention packages used during the research process. The former may be published in the *Journal of Statistical Software*, *The R Journal*, or *BMC Bioinformatics*, for example, and search queries that include `"R package"` along with domain keywords will narrow results. The latter requires identifying authors whom have used R in their analyses, hence useful packages may be mentioned in the Methods and/or References sections of the article. Accordingly, formatted citations for R packages can be obtained in R with `citation(package = "...")`. You can search for packages directly by name in Google Scholar: the `Cited by` link displays the number of times a package has been cited and connects to a page with those publications. Lastly, conferences are another collaborative environment wherein you can learn about R packages. There are two major annual R conferences: rstudio::conf for industry and useR! for academia. Conferences in your field may foster connections with fellow scientists whom use R for similar tasks and help you collect information about packages related to your expertise. Talks and presentations at conferences are often recorded and made available online for playback at a later date.

New

A developer of an R package may intend for it to be private (exclusively for personal or professional use) or public (at no cost and available for use by anyone) @rickert2018. If your task is specific to a line of research, consult colleagues to see if they have relevant (private) code they would be willing to share. Alternatively, literature in your field may either introduce R packages developed to solve a unique data science problem or mention packages used during the research process. The former may be published in the *Journal of Statistical Software*, *The R Journal*, or *BMC Bioinformatics*, for example, and search queries that include `"R package"` along with domain keywords will narrow results. The latter requires identifying authors who have used R in their analyses; useful packages may be mentioned in the Methods and/or References sections of the article. You can search for packages directly by name in Google Scholar: the `Cited by` link displays the number of times a package has been cited and connects to a page with those publications. Lastly, you can attend conferences to learn about recent R package developments and applications. There are two major annual R conferences: rstudio::conf for industry and useR! for academia. Conferences in your field may foster connections with fellow scientists who use R for similar tasks and help you collect information about packages related to your expertise. Talks and presentations at conferences are often recorded and made available online for playback at a later date.

4. Identify the issue in paragraphs from example papers

> "Even if you use Twitter only for professional purposes, consider opening up a little bit to show your followers you are a real person."

The authors emphasize that sucessful Twitter users tend to have memorable online personas, but could be more concise in conveying their main points. Social media is a casual platform where people can be entertained and relate to others. Popular people create character-like personas that are unique, memorable, and believable; likability is linked to relatability.

> "A more subtle form of introducing hazard bias arises from issues of the definition of the hazard and assigning loss estimates (by the original data source) to predefined hazard categories within a database. This is most apparent in the management of complex events involving multiple hazards versus a singular hazard event. A tornado spawned by a hurricane is counted as a unique tornado event, but it could also be lumped together within the entire hurricane event, or both. Each loss database classifies events differently, especially when they involve multiple hazard types (Guha-Sapir and Below 2002). Inconsistent naming conventions and classification methodologies aggravate this problem and can result in different (and/ or artificial) hazard categories for similar, if not identical events. For example, Downton et al. (2005) reveal a $520 million"flood" loss in FEMA's database that was not in the NWS data. The discrepancy is a result of differences in how each agency defines what constitutes a flood event. In this case, the event (storm surge) was outside NWS's definition of a flood."

The theme at the end of the issue sentence is that hazard losses are difficult to estimate because of the ways in which disasters are defined and categorized in different databases.

> "More subtly, hazard bias arises because databases differ in how they define disasters and sort hazards into predefined categories."

The alternative is more concise and clearly links the issue to theme. It is no longer passive. The old information is at the beginning and new information is at the end.

5. Identify the issue in paragraphs from your manuscript

See manuscript

6. Identify the issue in every paragraph of your manuscript

See manuscript

After workshop (see manuscript for revisions)

7. Diagnose and revise a paragraph from an example paper

8. Revise the issue sentence in each paragraph of your manuscript

9. Read Chapter 6 of Style (online version)