Week 4

Week 3 Recap

- Finished overview and examples of Opening conversation
- Had mock ending conversation.
- Bring 2 copies of writing project proposal to class.

For next week: - update & finalize writing project proposal (think about the story) - Read article about statistical pitfalls

Writing Science: Excerpts from Chapter 3 - Making a Story Sticky

- S: Simple
- U: Unexpected
- C: Concrete
- C: Credible
- E: Emotional
- S: Stories

Simple

- Ideas that stick tend to be simple. A simple idea contains the core essence of an important idea in a clear compact way. Simple ideas have power.
- It is important, however, to distinguish simple messages that capture the essence of an issue from those that are just "simplistic."
- A simple idea, therefore, is one that finds the core of the problem. It takes no special talent to see the complex in the complex. Cutting through the clutter to see the simple in complex is what distinguishes great scientists from the merely competent.
- Part of being simple is expressing your thoughts in language that builds off ideas that your readers already know.

Unexpected

- Science doesn't advance by great leaps but by many small steps, each of which makes its own contribution. We make a good story by identifying the knowledge gap we fill.
- You frame a knowledge gap by using what is known to identify the boundaries of that knowledge. It's like framing a window - build the structure to support the area you will fill in. Identifying a knowledge gap creates curiousity. Filling in that gap creates novelty.

Concrete

- Simple has power, but concrete adds mass to that power.

- The importance of bing concrete might seem an obvious and inherent characteristic of writing science. After all, science is about data, and data are concrete. But science is also about ideas, and ideas are abstractions - the antithesis of concrete.

Credible

- We establish the credibility of our ideas by grounding them in previous work and citing those sources.

- We establish the credibility of our data by describing our methods, presenting the data clearly, and using appropriate statistical methods.

- We establish the credibility of our conclusions by showing that they grow from those credible data.

- We build a chain that extends from past work into future directions. A break anywhere in that chain makes the whole endeavor lose credibility.

Emotional

- This is an awkward one for scientists. To do good science you must be dispassionate and objective about your work. There is, however, one emotion that is not only acceptable in science but fundamental to it: curiousity.

- If you don't ask an engaging question, and instead just offer new information, you appeal to another, weaker emotion.

- The E element of the SUCCES formula is thus closely aligned with U. Unexpected things create curiosity, so use that link to your benefit.

Stories

- This whole book is about telling stories - about seeing your work as a story and presenting it that way. But stories are modular; a single large story is crafted from a collection of smaller story units, threaded together.

Writing Science: Excerpts from Chapter 4 - Story Structure

All stories have a beginning, a middle, and an end.

There are four cour elements that underlie the structure of all stories:

- **Opening:** Whom is the story about? Who are the characters? What is the larger problem you are addressing?

- **Challenge:** What do your characters need to accomplish? What specific questions do you propose to answer?

- **Action:** What happens to address the challenge? In a paper, this describes the work you did.

- **Resolution:** How have the characters and their world changed as a result of the action? This is your conclusion - what did you learn form your work?

- OCAR functions are as central to a scientific paper as they are to a work of fiction. A good paper describes the larger problem and the central "characters": (O); it frames an interesting question (C); it presents your research plan and results, developing the action (A); and it leaves the reader with an important conclusion about how our understanding of the world has changes as a result of the work (R).

OCAR can be mapped into a scientific paper:

- Introduction has three subsections
  - Opening: This is typically the first paragraph that introduces the larger problem the paper is targeting. What is the context, and what are the characters we are studying?
  - Background: What information does the reader need to understand the specific work the authors did, why is it important, and what will it contribute to the larger issue? I consider this an extension of the O, as it fleshes out introducing the characters.
  - Challenge: What are the specific hypotheses/questions/goals of the current work?

- Data / Methods / Results: describe the action (what did you do) and then describe the findings.

- Discussion / Conclusion: Develops the resolution. What did it mean, and what did you learn?

ASCCR

- Recall we spent quite a bit of time focused on POWER (lots of acronyms in this class...), which was one subcomponent of ASCCR, specifically the S(tructure)

Attitude

- The Attitude of collaboration starts with two concepts of emotional intelligence: self-awareness and self-management (Goleman, 2012).

- Collaborative statisticians should be aware of their attitudes and emotions and should choose to act on attitudes that enhance their effectiveness as collaborators and be mindful of—but refrain from acting on—attitudes and thoughts that detract from collaboration.

Structure

- Use Prepare - Opening - Work - Ending - Reflection for collaborative meetings.

Content

- Every statistics project can be summarized in three parts: Qualitative (Q1) - Quantitative (Q2) - Qualitative (Q3). Note these also map to story telling structure.

Q1

- Q1 is the project's initial Qualitative component, which sets the foundation for the Quantitative component of the project (Q2) and the implementation of the solution (Q3). Specifically, there are seven aspects of Q1, inspired by the questions of George Heilmeier's Catechism (Hahn, 2014), relevant for every applied statistics project.

1. *What is the domain problem?*

2. *Why is the domain problem important or interesting?* Understanding the reasons the domain expert finds the problem important and/or interesting can clarify the problem and help the statistician identify the best quantitative (Q2) methods for solving it.

3. *How will the eventual solution be used?* Eliciting this information can clarify a complicated project and identify the implementation of the solution as an ultimate goal toward which to work. This information can also inform future decisions about which statistical methods (Q2) to use.

4. *What potential data could solve the domain problem?* I.e., what data, if it were available and accessible, would help answer the underlying research questions or guide the business or policy decisions? This is an important theoretical exercise that is too often skipped in statistics and data science collaborations. It may be that the solution to the domain problem becomes simplified by collecting new data and that the domain expert and data scientist overlook this by focusing only on data already collected. On the other hand, the impossibility of collecting certain data can illuminate the intricacies of the domain problem and help guide the collection or analysis of alternative data.

5. *What data have been collected?* Why were the data collected originally? For what purpose? When and where were the data collected? Who or what collected the data? How were the data collected, i.e., with what instrumentation/methods? Understanding the data and their collection will help the statistician appropriately explore, visualize, model, and analyze the data (Chatfield, 1995).

6. *What may be the qualitative relationships between variables, for those observed and unobserved?*

7. *Which types of statistical analyses or machine learning techniques would be most useful to the domain expert?* Which would not be useful? Could new statistics or data science methods be developed to more usefully answer the domain question? If the domain expert does not understand or cannot defend the choice of a specific technique, that technique may not be useful, and a different (simpler or more familiar) method may need to be used (Stallings, 2014). If the simpler method is not appropriate for the problem and the data, the more advanced method should be used and sufficient time spent explaining the new method to the domain expert. Statisticians and data scientists have the potential to drive collaborative projects to develop new technical methods tailored to better answer the particular domain questions and generalizable to answer other domain questions.

Q2 Technical skills in statistics and data science and how to effectively apply them are the sine qua non of statisti- cal collaboration. Applying these skills to quantitatively explore, visualize, model, and analyze data to solve a research/business/policy problem in a reproducible manner is the Content of collaborations that occurs after the problem has been well defined and understood (Q1) and before the solution to that problem is determined, explained, and implemented (Q3). Keeping a reproducible record of decisions made on how the raw data have been tidied and analyzed is essential when there is the potential for more than one person to work on the data.

Q3 Translating the (Q2) data analysis into useful qualitative answers to the original research question or the implementation of a data-driven decision completes the Q1Q2Q3 process. A statistician who sufficiently understands the domain problem (Q1) and appropriately analyzes the data (Q2) is not finished until she explains to the domain expert— qualitatively—what the analysis results mean (Q3). A collaboration is sufficiently completed (Q3) only after the statistician and domain expert jointly interpret the quantitative results (Q2) to answer the original research question (Q1)

Communication

- Communication is key to effective interdisciplinary collaborations.

- *Asking great questions:* a great question does 2 things: 1) The question creates shared understanding between the statistician and domain expert by eliciting information useful for answering the research/business/policy questions, and 2) The question is asked in a way that strengthens the relationship with the domain expert.

- *Listening, Paraphrasing, and Summarizing:* When a domain expert answers questions, the statistician must, of course, listen to the domain expert to understand the content of the response, paraphrase the response to clarify content, and then summarize to solidify the content.
  - Focus as much as possible on the domain expert while listening. Manage distractions. Be patient.
  - Listen to the domain expert's words without jumping ahead to their implications
  - Paraphrase whenever the domain expert communicates a key concept or whenever one is confused or uncertain
  - State the intent behind attempts to paraphrase and summarize (e.g., "Let's see if I understood correctly; you are attempting to. . .")
  - Use visual aids (e.g., a whiteboard) when paraphrasing and summarizing so the domain expert can both hear and see the restatement or summary
  - Summarize conversations completely before moving on to the next topic.

- *Explaining Statistics to Nonstatisticians*

Relationship We believe that collaborative statisticians should attempt to become more aware of the domain experts' attitudes and emotions and should choose their own words and actions to manage/strengthen the relationship. Some practical advice on how to build stronger relationships with domain experts includes:

-Express authentic interest in developing and building a strong relationship with the domain expert

- Learn and use the language of the domain expert's discipline

- As mentioned in the section on Attitude, respect the skills the domain expert brings to the collaboration because strong relationships are built on respect

- Be aware that some relationships might not be worth continuing. If a domain expert does not exhibit respect for the statistician, it may be worthwhile to end that relationship and focus on strengthening others.

- Building strong relationships requires time, patience, and trust. Act trustworthy to gain trust.