From Topic to Question
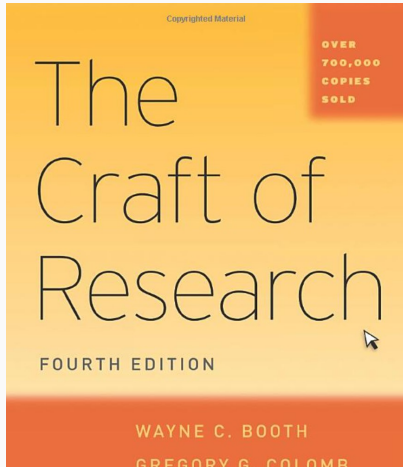
Jonathan Tannen

Agenda

- From Topic to Question (1.25 hours)
- Break (0.25 hours)
- Git & Github (1 hour)

How do you take a broad topic, and narrow it down to an answerable question?

From Topic to Question

Ask a **question** that solves a **problem** that you can convince readers to care about.

Possible question types:

- **Disagreement**: Some people think X, but other people think Y.
- **Structure and Composition**: What patterns exist? What typologies?
- **Claim**: Unmeasured claim about how things work.
- **Optimization**: Can we do this better?

From Topic to Question

Avoid questions that are. . .

- Settled fact (you could look them up)
- Unanswerable with data
- Answers are dead ends

Is your question answerable with data?

Think about your question and your final product. Will it answer the question? What will skeptics say?

- Is your question to broad to have an "answer"?
- What would the ideal dataset look like?
- What will a skeptic say?
- State a possible finding: "I find that an increase of X is correlated with an increase of Y. . . "

From Hofman, J.M., Watts, D.J., Athey, S. et al. Integrating explanation and prediction in computational social science. Nature 595, 181–188 (2021).

Table 1 | A schematic for organizing empirical modelling along two dimensions, representing the different levels of emphasis placed on prediction and explanation ·

	No intervention or distributional changes	Under interventions or distributional changes
Focus on specific features or effects	Quadrant 1: Descriptive modelling Describe situations in the past or present (but neither causal nor predictive)	Quadrant 2: Explanatory modelling Estimate effects of changing a situation (but many effects are small)
Focus on predicting outcomes	Quadrant 3: Predictive modelling Forecast outcomes for similar situations in the future (but can break under changes)	Quadrant 4: Integrative modelling Predict outcomes and estimate effects in as yet unseen situations

From Topic to Question

Ask your question in the following form. . .

1. Name your topic: "I am studying _____."
2. Add your question: "I want to know _____ because _____."
   a. This is a (prediction, feature) question that assumes (no intervention, yes intervention).
3. Add an example finding: "A possible finding is _____. My data is capable of showing that because _____."
4. Implication: "A hypothetical way my result will be used is _____."

Think, Pair, Share

1. 10 min: Individually write your question in the form of the prior slide.
2. 30 min: In groups, share your question and proposal. Group members, critically ask:
   a. Is this question answerable with data?
   b. Would the proposed analysis convince a skeptic?
   c. Would the result be useable?
3. Come back, share out results.

Assignments

- Next week, 1/28: GitHub Repo with raw data.
- 2/4: Data Summary Analysis and 10-minute Presentations.
- 2/11: Project Proposal 1.