

DATA 552

Communication and Argumentation

Lecture 6: End of a project

Dr. Vikas Menghwani, Assistant Professor of Teaching

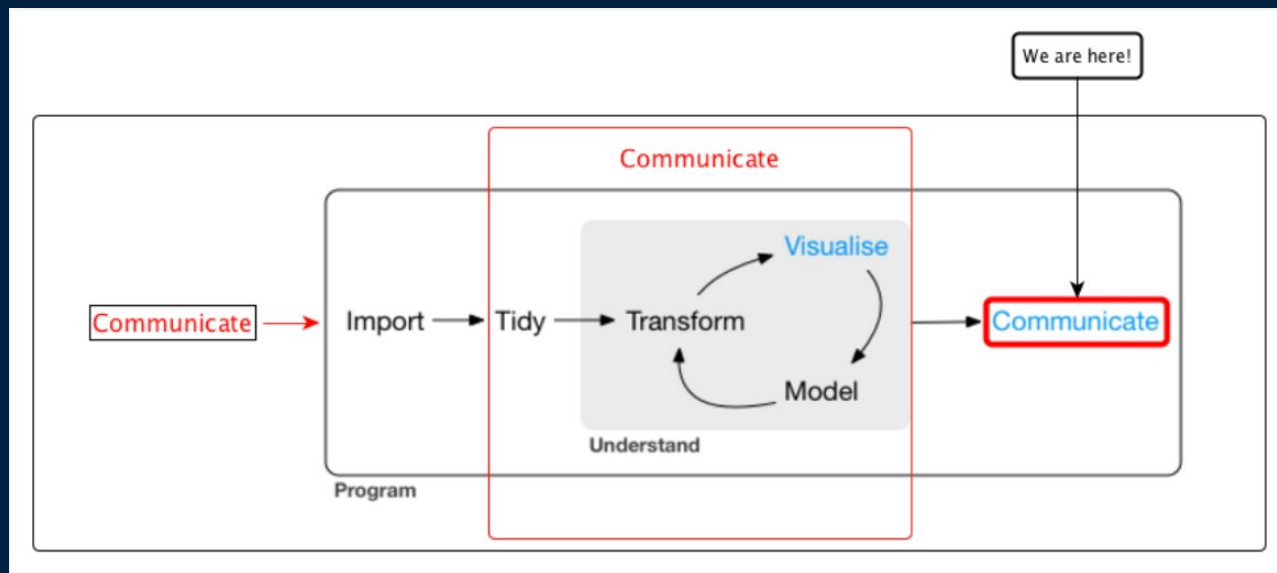


From last class..

➤ When you assign a probability to a claim (e.g. “I am 70% certain that it will rain tomorrow”), what are the two main ways to interpret that probability?

UNCERTAINTY MATRIX		Level of uncertainty			Nature of uncertainty	
Location of uncertainty		<i>statistical uncertainty</i>	<i>scenario uncertainty</i>	<i>systemic uncertainty</i>	<i>knowledge-related</i>	<i>variability-related</i>
<i>Context</i>	Assumptions about ecological, technological, economic, political, or social context					
<i>Expert judgment</i>	Narrative uncertainty or experience uncertainty					
<i>Model</i>	Model structure: relations					
	Model parameters: choice and representation					
	Model input: data, drivers					
<i>Data</i>	Availability, gaps, quality					

End of a data science project



End stages of a project..

- Separate relevant information from irrelevant information.
- Turn meaningless numbers into meaningful numbers.
- Structure the relevant information so that the decision-maker understands and trusts it.

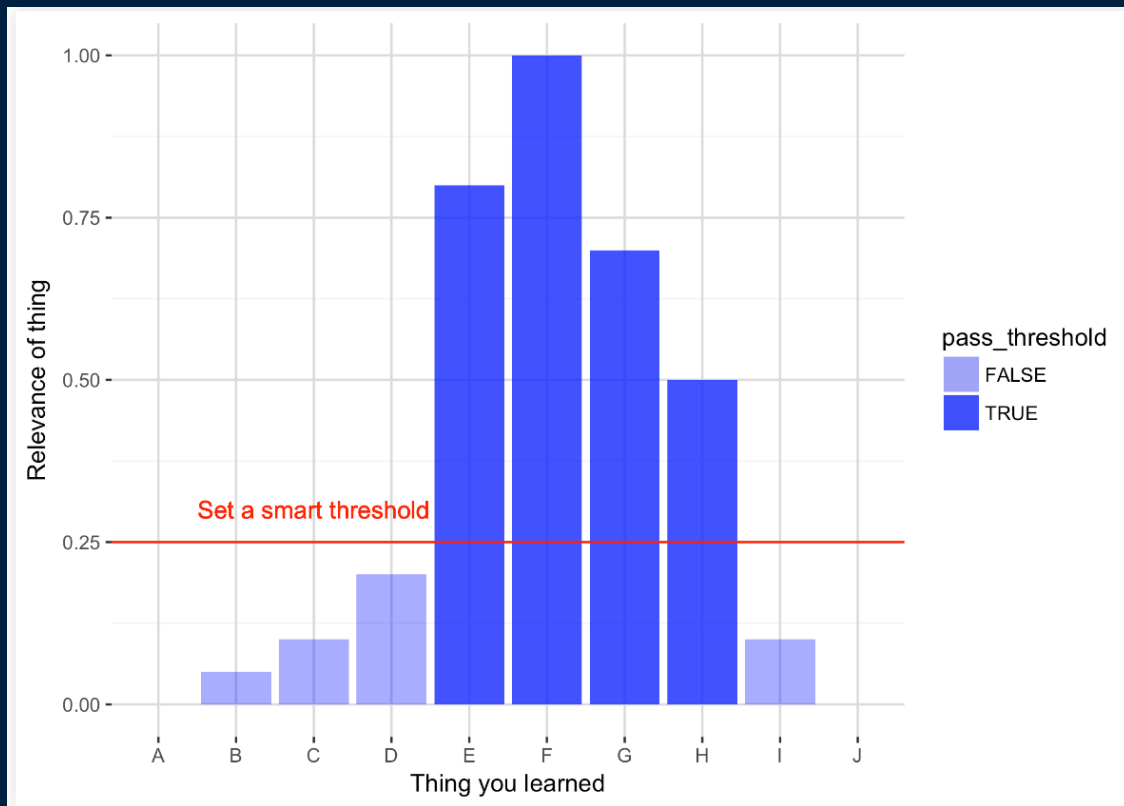
Relevance: a rough and extremely brief guide

- Relevant to you (as a data scientist) is not necessarily relevant to the decision maker(s)

Relevant	Rarely relevant
What you learned about the variables that are part of the decision.	Everything you learned while doing research.
The results of the project.	The history of your involvement with the project.
Accurate interpretations of statistical tests.	Technical terminology used by academic statisticians.

Relevant != relevant enough.

- Don't overwhelm with details of **minor** relevance at the risk of losing focus on issues of **major** relevance
- Items of minor relevance can be elicited through discussion (presentation) or appendix (reports) for those that are interested.
- You should be prepared to explain them, but they shouldn't be focused on in the main deliverable.



Relevant != relevant **enough**.

Examples:

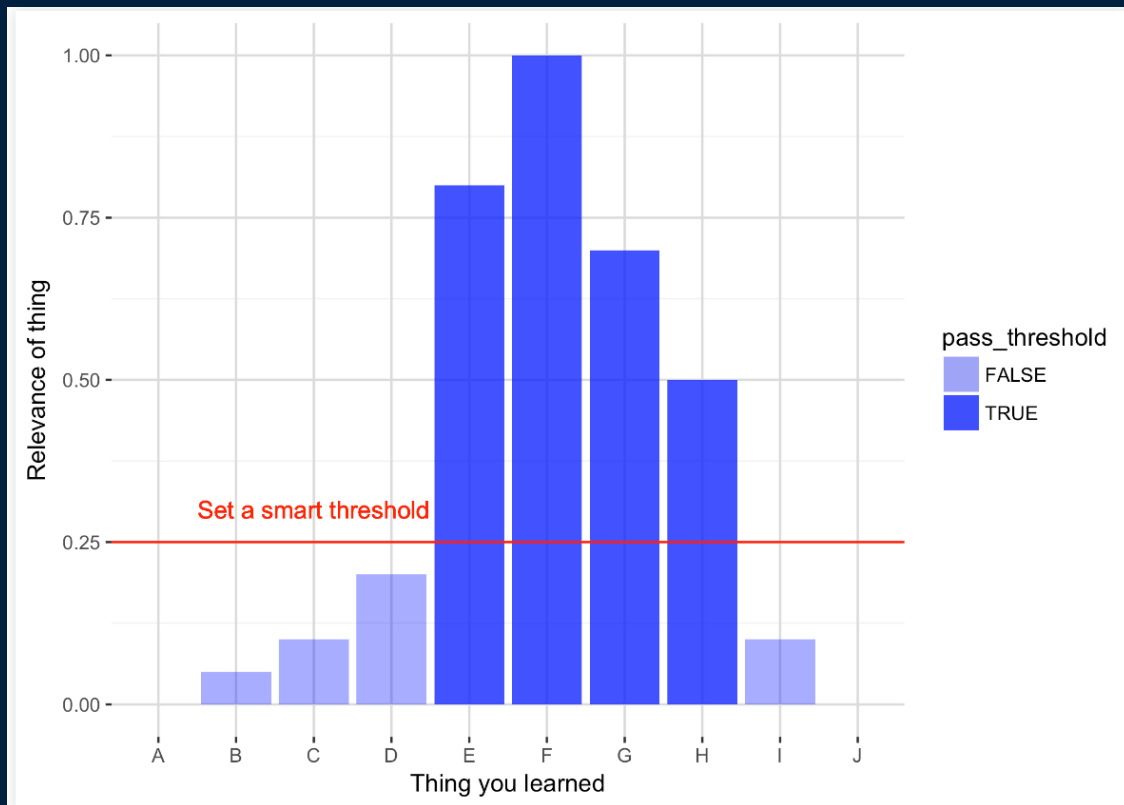
Project Management

Marketing Pitch

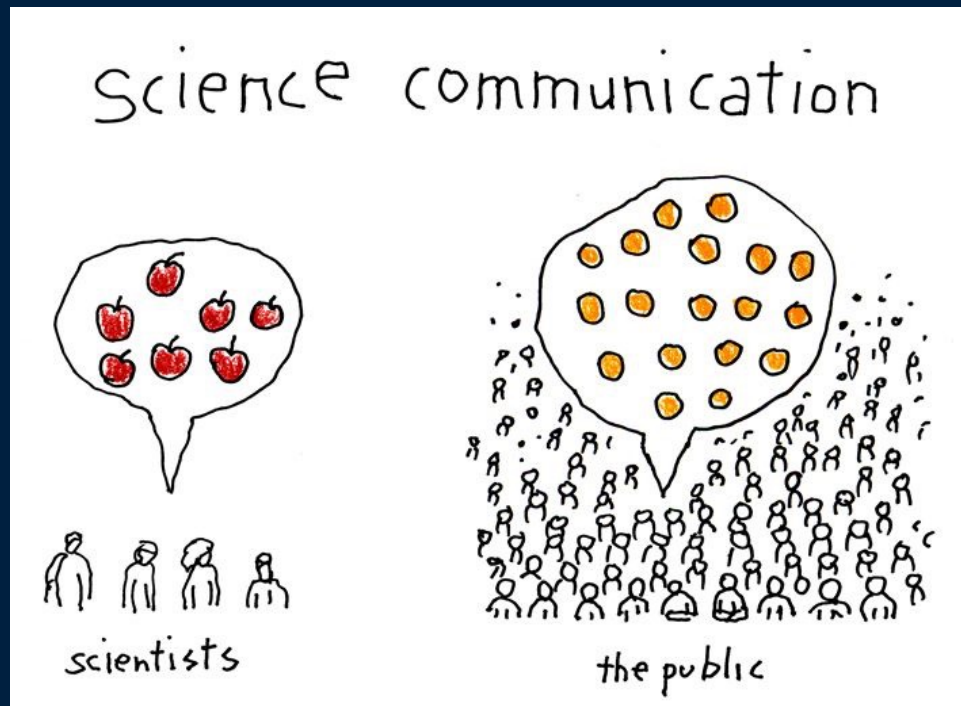
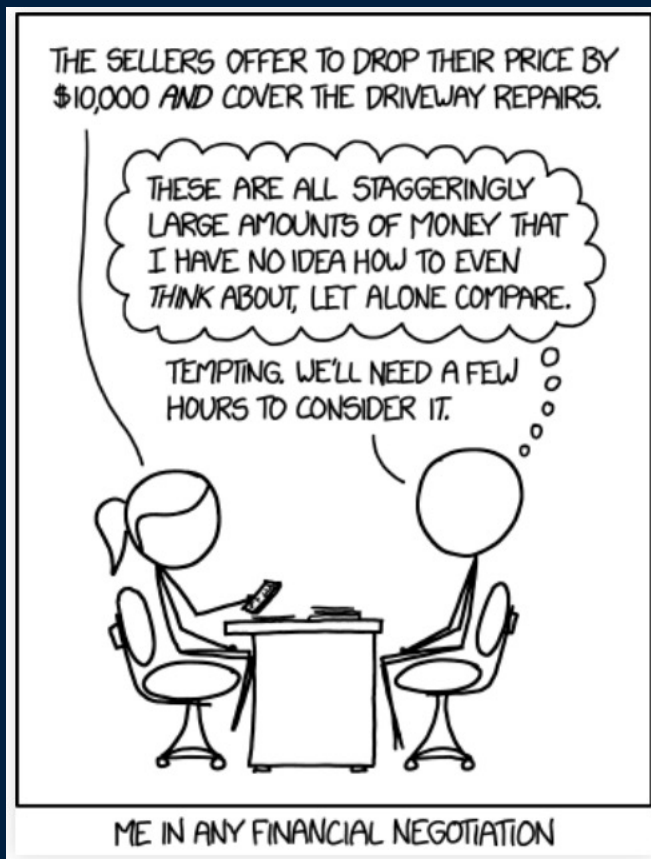
Research Presentation

Business Proposal

Technical Support Documentation

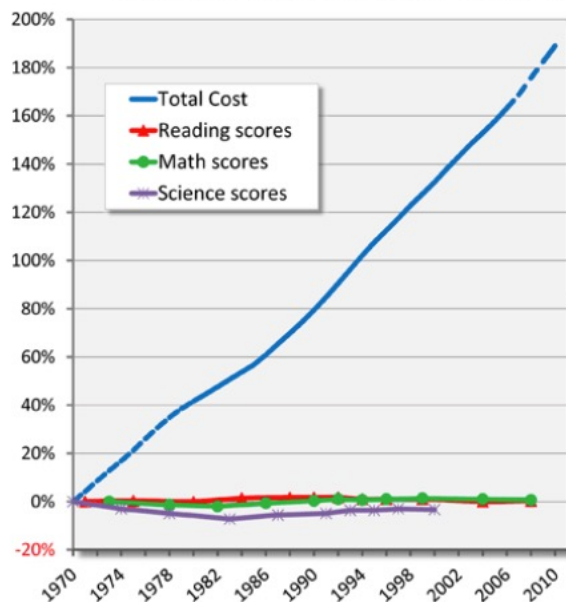


Meaningless numbers



Bringing meaning into numbers

Trends in inflation-adjusted total cost of a K-through-12 public education and achievement of 17-year-olds (U.S.)



Cato Institute

"Total cost" is the whole amount spent on the K-12 education of a student graduating in the given year. We graph the percent change in that amount--and in test scores--over time.

Data sources

(test scores): NAEP, *Long Term Trends* reports, U.S. Department of Education (cost): *Digest of Education Statistics 2011*, Table 191, U.S. DOE, CPI adjusted to constant 2012 \$. Missing values linearly interpolated or extrapolated.

Prepared Sept. 2012 by:

Andrew J. Coulson,
Director, Cato Institute Center
for Educational Freedom

Which would you prefer?

Sending your child to a 2016 school? Or sending your child to a 1975 school, and getting a check for \$5,000 every year?

Importance of narrative

First, let's quickly reflect on how often our mind wanders during..

- my lecture or any lecture
- your favorite binge-able TV show, or while reading a novel

➤ Why? Narrative is engaging. I cannot compete with narrative built purely for entertainment.

➤ What is narrative? Is it the same thing as storytelling?

Importance of narrative

Narrative includes both the core substance of the “story”, as well as its structure

Consider the following claims:

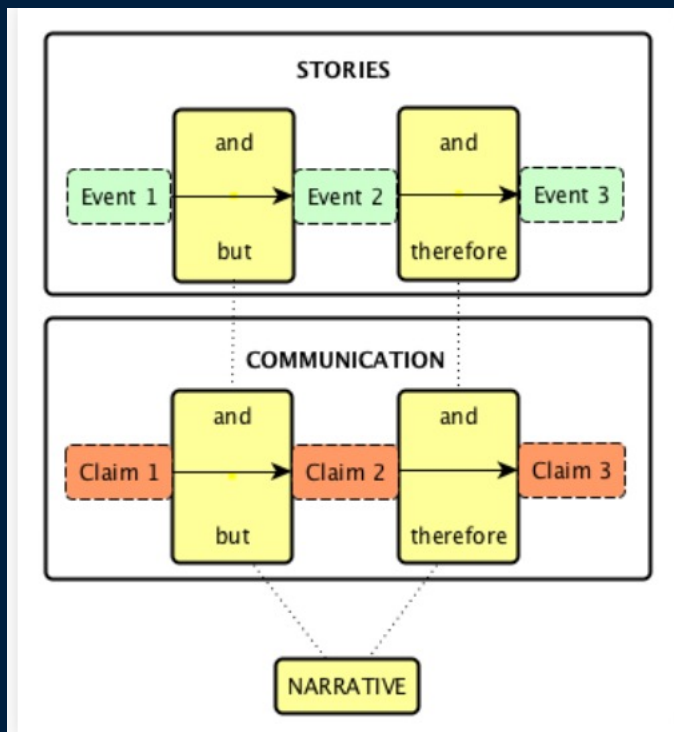
Structure #1	Structure #2
You don't want to carry around needless weight. And the weather forecast says there is a medium chance of rain today. And you really don't want to get wet. And you should should bring a raincoat.	You don't want to carry around needless weight. But the weather forecast says there is a medium chance of rain today. And you really don't want to get wet. Therefore , you should should bring a raincoat.

Trey Parker (creator of South Park) on good narrative:

“I call it the rule of replacing *ands* with either *buts* or *therefores*”.

Naïve narrative	Better narrative
“this happened, and then this happened, and then this happened, and then this happened...”	“this happened, therefore this happens. But this happens...”

- "I lost my job, and then I started a business, and then I became successful.
- " Better: "I lost my job, therefore I had the chance to start a business. But it was through overcoming numerous obstacles that I finally became successful."
- "but" is used to introduce a conflict or pivot, and "therefore" to show causality, creating a more compelling and structured narrative.



Structure

Narrative stories have an ebb and flow to them

➤ Beginning → Middle → End

➤ Momentum → Conflict/Tension → Resolution

➤ Introduction → Methods/Results/Analysis → Conclusion

While narrative is a good communication tool, it is not necessarily a good thinking tool



Consider the two sentences below

Sentence 1	Sentence 2
Eric is a friendly guy, and he gets frustrated easily.	Eric is a friendly guy, but he gets frustrated easily.

Is there any difference between the factual information conveyed by the two sentences?

Sentence 3	Sentence 4
Eric gets frustrated easily, and he's a friendly guy.	Eric gets frustrated easily, but he's a friendly guy.

While narrative is a good communication tool, it is not necessarily a good thinking tool



There's a reason narrative should enter primarily at the end of a project.

- Getting caught up in it while you're still in the middle stages of a project can create unnecessary biases/confusion.
- Consider framing your work in your mind: "I thought about fitting a linear model, and I also thought about using random forests, and I'm unsure what supervised model to move forward with."
- Versus: "I thought about fitting a linear model, **but** I also thought about using random forests, **therefore** I'm unsure what supervised model to move forward with."
- The narrative structure here is both suffocating and confusing. Why?

While narrative is a good communication tool, it is not necessarily a good thinking tool



- Suffocating: Using 'buts' and 'therefores' here vaguely hints that there are no other options. Using 'ands' leaves your brain a little more wiggle room to consider **other models** not appearing in that brainstorming session.
- Confusing: The existence of random forests aren't a good reason to be unsure if a linear model is appropriate for the task at hand. The narrative structure provides that confusing causation.
- The point: When you're assessing evidence (deciding what's true for **yourself**), it is good practice to avoid the narrative structure.
- Once you have decided what is true, then it is time to build the narrative structure for others.

While narrative is a good communication tool, it is not necessarily a good thinking tool



➤ Another caution: interesting details often make for better narratives, but worse hypotheses.

➤ Daniel Kahneman and Amos Tversky presented subjects with the following scenario:

“Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.”

➤ Which is more likely?

- Linda is a bank teller.
- Linda is a bank teller and is active in the feminist movement.

At the end of the project

It is good practice to consider the following steps at the end of your project

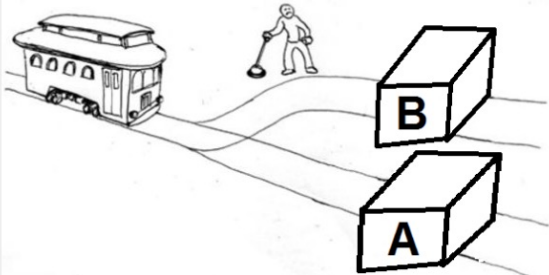
- Initially: **de**-narrativize your findings. This aids your ability to determine what's true
- Finally: **re**-narrativize your findings. This aids others in understanding what you believe to be true.

Portfolio

➤ Consider the thought experiment:

You see a trolley headed towards box A, but you can pull a lever to switch it to box B. You know in both boxes there are people tied up on the track. You want to know if you should switch the trolley to box B so you pay someone to test if there are more people in box A than box B. The test comes back that box A **MIGHT** have more people but the test was non-significant at $p = .061$

Do you still switch over the track or does marginal significance not matter?



Step 1: **De**-narrativize. Remove all the connecting words to identify the core propositions. You should end up with seven bullet points.

Step 2: **Re**-narrativize. Place your own connecting words to advise the decision maker on whether or not to pull the lever.



THE UNIVERSITY OF BRITISH COLUMBIA



Beginning

- We have identified a decision that will be augmented or automated as a result of our project.
- We have identified the alternatives, the objectives, and the context of that decision.

Middle

- We have made our assumptions explicit, through testable predictions.
- We have tested our predictions and noted our results.
- We have documented our research, analysis, and development.