

DATA 552

Communication and Argumentation

Lecture 1: Critical Consumption

Dr. Vikas Menghwani, Assistant Professor of Teaching

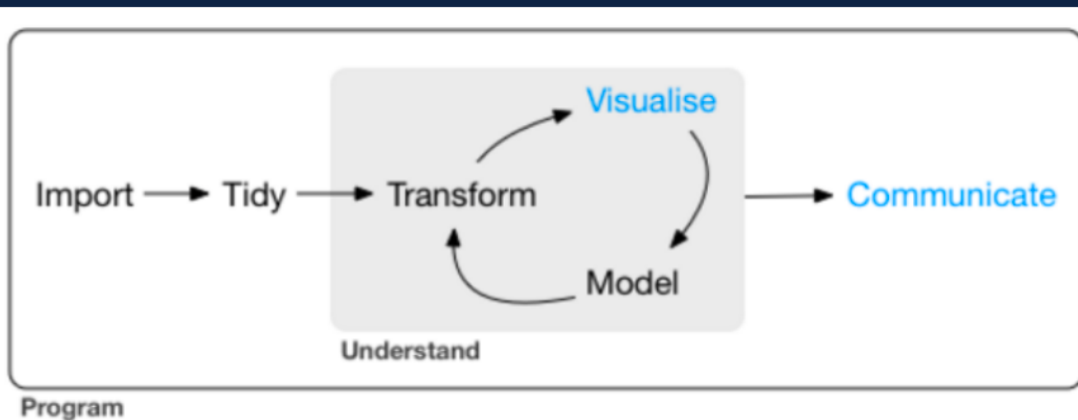


Which of the options below demonstrates the best way for a data scientist to communicate?

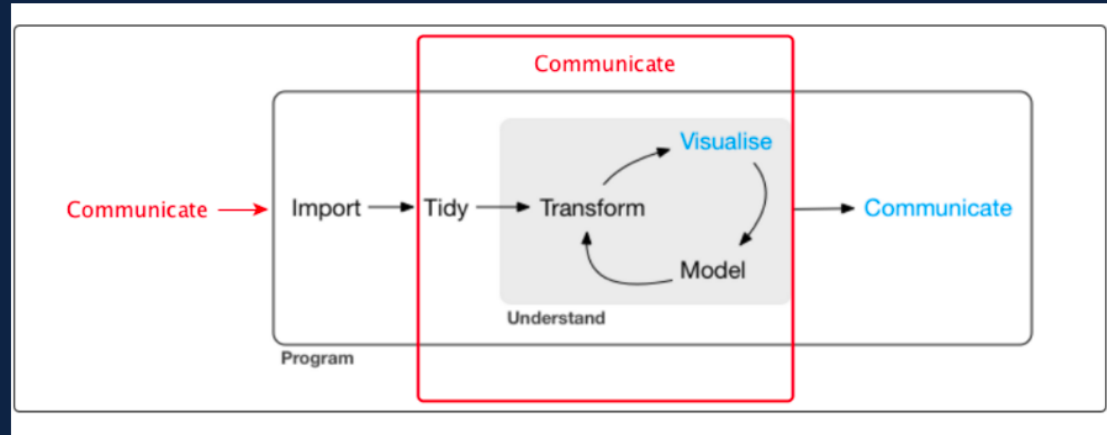


- | | |
|----|---|
| 1. | If you know more about programming, you can devise more efficient solutions to computational problems. |
| 2. | $r = \frac{1}{v + \epsilon},$ <p>where r is the average runtime of a Python programmer's solutions to a sample of ten computational problems, v is their score on a Python vocabulary quiz, and ϵ is a Gaussian error term.</p> |
| 3. | <pre>get_estimated_runtime <- function(vocabulary_score) {
 1/(vocabulary_score + rnorm(n = 1, mean = 0, sd = 1))
}</pre> |

Communication



Communication



Learning outcomes

After completing this course, students will be able to:

1. Critically **consume** communication in data science contexts.
2. **Write** effectively about data science topics for a variety of audiences.
3. Make **calibrated claims** about data science results.
4. Reason about and apply principles of **narrative** when reporting on data science results.
5. **Design and deliver** clear and informative **oral presentations**.



Assessment

Date	Title	Grade Weight
2024-02-26	Lab 1: Critical consumption and rhetoric	20%
2024-02-29	Quiz 1 (Canvas)	5%
2024-03-04	Lab 2: Initiation of a Data Science Project	20%
2024-03-11	Lab 3: Report on a Data Science Project	20%
2024-03-14	Quiz 2 (Canvas)	5%
Final week	Project Presentations	20%
End of Module	Portfolio	10%



** Quizzes will need lockdown browser*

- Late submissions of assignments carry a penalty of 10% per day (max 3 days)
- Deadlines are known well in advance, most deliverables are completed in-lab/class.

Portfolio document

In-class activities need to be recorded in the portfolio document, to be submitted end of term



Big Picture

- What is required for strong communication in statistics/data science?
- The same as communication in any environment — you need to figure out what the ‘story’ is and who you’re telling it to.
- Both tasks require careful thought and reflection.
- And both need to be considered **throughout** the data science lifecycle.



Big Picture

- What is required for strong communication in statistics/data science?
- The same as communication in any environment — you need to figure out what the ‘story’ is and who you’re telling it to.
- Both tasks require careful thought and reflection.
- And both need to be considered **throughout** the data science lifecycle.



Course outline

- Critical consumption
- Rhetorical situation
- Writing – tone, language, and style
- Communication in data science projects
- Speaking and presenting



Communication is all about sending and receiving signals



Critical consumption

- Statistical Literacy: Possess a deep understanding of statistical methods to critically evaluate the appropriateness and robustness of analytical techniques used.
 - Analytical Thinking: Dissect data and research methodologies to understand the mechanics and assumptions behind them.
 - Logical Reasoning: Use logic to distinguish between correlation and causation, and to identify logical fallacies or errors in argumentation.
- Skeptical Mindset: Approach all data, analyses, and research findings with a *questioning attitude*, seeking evidence and validation.
 - Contextual Analysis: Consider the broader context in which data is collected and presented, including socio-economic, political, and cultural factors.
 - Source Evaluation: Scrutinize the origins, credibility, and quality of data sources, including the potential for bias or conflict of interest.



Contextual analysis

- From: Bergstrom and West's callingbullshit.org
- Consider this claim (circa 2016, USA):

“Food stamp fraud is at an all-time high, with cases this year including a state lawmaker and even a millionaire. According to the USDA, \$70 million of taxpayer money was wasted in 2016 due to food stamp fraud.”
- What's your reaction to this?
- Do you think this is the reaction the claimant is looking for (source knowledge might help)?
- Assuming none of this is an outright lie, can I contextualize it better?



Contextual analysis

- First, “all-time high” is a powerful part of headlines such as this, but is it adjusted for inflation? Adjusted for size of program?
- \$70 million is way more money than I’ll ever see, so this certainly seems like a big problem...
- **Fermi estimation**, AKA back-of-the-napkin estimation, is a powerful tool to contextualize statistical results that are often rhetorically(!) placed.



Contextual analysis

Bergstrom & West propose the following approximations to important variables...

- 300 million Americans (it's 330)
- 10% on food stamps (it's 15%)
- \$1k spent yearly per person in program



This would suggest total program cost is roughly \$30 billion.

So, the rampant fraud amounts to approximately $70\text{mill}/30\text{bill} = 0.0023$, or roughly 0.2% of all benefits paid.

Turns out, real number is more like 0.1%

Contextual analysis

- Does the additional context change your reaction?
- That may still depend on biases/stances outside of the statistical contextualization regarding fraud...
 - ‘I support universal food stamps. Food for free, free for all! I’m pro-food.’
 - ‘I’m against any food stamp program. Stamp out the stamps, work for food! I’m pro-jobs.’



Summary: Food stamps

Considering our skills again...



- Reaction to original claim: Concern (lots of fraud). Perhaps skepticism (biased source, or biased recipient, aka me).
- Variables: Percent of population on food stamps. Population size. Yearly cost per person. Yearly cost for whole program. Percent of program lost to fraud.
- Note that variables are rarely given directly in a claim. Coming up with them yourself is an exercise in understanding an argument.

Skeptical mindset

Sci Tech Daily, [July 2022](#)



“AI Algorithm Predicts Future Crimes One Week in Advance With 90% Accuracy”

- What is your reaction?
- What all do you want to know about this claim?

Bias

Confirmation Bias: a bias which confirms a pre-existing belief

“As cities around the nation grapple with the effects of climate change, the latest survey results offer a beacon of hope: a 60% approval rating for the newly implemented climate policy. This overwhelming support is a clear indication of the public’s alignment with progressive environmental initiatives.”

Selection Bias: specifically selecting an area "in the below example the tech sector was selected whereas a state is likely to have many sectors"

"The state’s economy is on an upswing, with the tech sector reporting an impressive 15% increase in employment this past year. This surge in high-tech jobs is a testament to our state’s thriving innovation ecosystem and forward-thinking policies.”



Portfolio activity

- Consider the following link: [NBA Home Court Advantage](#)
- Reactions: In your portfolio, note your overall impressions of the article. What's surprising? Confusing? Predictable?
- Variables: In your portfolio, note some variables from the article. What are some key relationships between variables? Are there other potential variables of interest that could help contextualize?





THE UNIVERSITY OF BRITISH COLUMBIA

