

DATA 606

Quantitative

Timeline for the Rest of the Semester

10/29 (Today): I am still missing final lit reviews!! Please send them in ASAP.

Draft of “Data” section in Data & Methods due; peer-review in class.

11/5: Draft of “Methods” section for Data & Methods due; peer-review using form (no class)

11/12 (No class): Turn in revised Data & Methods to Dan; feedback from him returned on 11/19.

11/19: Dan will go over Results, Discussion and Conclusion sections

11/26: Thanksgiving; work remotely

12/3: In-class peer-review of Results/Discussion/Conclusion of Quant

12/10: Turn in Results/Discussion/Conclusions of Quant to Runfolo (will provide feedback, but no further deliverables this semester).

How to assess the quantitative sections

We're going to do in-person for Data today, Methods next week, then you'll do a full round of peer-review with a worksheet on Data & Methods together like last time.

Focusing on the DATA section today.

-> Clarity and Precision

- Does the data match the research question, and is it clear why?
- Are the sources of each dataset properly cited?
- Is the way the dataset was constructed clear?
- Is the unit of observation and N explicitly defined?
- Are any features included in the source data well-defined (i.e., units, meaning, etc)?
- Are their figures that aid in your understanding of the data? Would it be helpful if not?

-> Data Quality

What is the total N? Is it sufficient to conduct the planned activities? Think about this in terms of both total N and information inter-dependence.

Are there any obvious biases or missing data?

Is the support (or scope) of the data clear? I.e., are limitations of the population well understood?

-> Reproducibility

Is the data available in a format and/or location that will promote reproducibility?

If not, is there a rationale for why?

If data is not available, are there synthetic analogues that might be provided?

Are any pre-processing scripts or codes (and, optimally, image environments) available? If not, is there a rationale for why?

Are data versions or timestamps recorded?