DATA 606

Quantitative

Timeline for the Rest of the Semester

10/22 (Today): Email me your revised intro/lit reviews (“final draft”). I’m going to share these with your advisors as well. We’re going to go over the quantitative section of your literature reviews.

10/29 (Next week): Draft of “Data” section in Data & Methods due; peer-review in class.

11/5: Draft of “Methods” section for Data & Methods due; peer-review

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 Runfola (will provide feedback, but no further deliverables this semester).

Purpose of Quantitative Analysis

- Required Component of Prospectus
  - Goal is to illustrate you (a) have mastered the ability to understand and implement quantitative techniques common in your fields, and (b) can communicate these results.
  - Data collection, processing, modeling and rigour are all discussed by the committee.
  - Can take the form of a new analysis you produce, or a replication study of past research, depending on your advisor.

Structure

- Your quantitative section should mirror sections commonly found in academic papers - Data, Methods, Results, Discussion and Conclusion.
- Each section should explicitly cover the stated topic: no making critical inferences in Results, no presenting new results in your discussion, etc.

Data

- It should be exhaustive - i.e., if you are using a pre-trained neural network, the input data into that network should be briefly cited here.
- It should be structured along major data types if you're integrating information - i.e., in my student's case, that's frequently a sub-section for satellite imagery, then another sub-section for survey information.
- Make sure your number of observations is explicitly stated and clear.
- Make sure your unit of observation is explicitly stated and clear.

Methods

- Should carefully document every stage of your process - in 90% of cases, this should involve a figure that is a flow chart that shows how data inputs and model outputs link together.
- Should not be a literature review; cites should be rare at this point, and techniques you're using should either be novel or already described in your literature review.
- Your training procedure should be very well described, if relevant.
- In the introduction to this section, you should provide an overview to the reader.
- Any techniques you use to handle imbalance data or splitting should be described here (*not* your data section, which should be focused on the raw data).
- All pre-processing, ETL pipelines, sampling, normalization, and other techniques you apply to the data need to be carefully documented here.
- You should have an explicit section describing model evaluation (F1, precision, recall, accuracy), with a particular focus on describing how you will interpret these metrics in your use case.

Results

- Should - in excruciating detail - spell out your results with tables and figures.
- There should be nothing normative in this section - the text should largely just be describing your tables, figures, and any other key results.
    - Words like “good” or “bad” or “highly accurate”, etc., don’t have any place here. Save the commentary for your discussion.
- How I normally approach: generate your tables and figures, then write a narrative describing the most important elements of them to the reader. Again, don’t draw any conclusions here.

Discussion

- Generally broken into multiple sections
  - Overall summary of findings, including any “blockbuster” ideas or concepts that emerge.
  - At least one section drilling in on key/interesting findings. I.e., “This seems to work really well in certain conditions” or “accuracy varies across different dimensions”.
  - At least one section on Limitations and/or Biases
    - Anything you can think of - small datasets, training procedure risks, validation risks, etc.

Conclusion

- 1-2 paragraphs.
- Overall summary of what you attempted to do, the results, and your core finding(s).

Supplementary

- Optional, but can be helpful for many analyses.
- Should describe any additional tests you did to explore the robustness of your findings, or additional figures you created that didn’t make it into the main text.
- Doesn’t have to have a lot of writing; can mostly be figures/tables and captions. Generally one-two sentences per figure/table is sufficient, and may be a bit redundant with captions.