Final Projects

Winter Institute in Data Science

Ryan T. Moore

2020-01-06

1 Final Projects

The final project is an opportunity to answer a social science question of interest to your group. Your group will collect and analyze data that enable your analysis to answer or inform your question. Data can come from any source, such as a work project, an academic replication repository like https://dataverse.harvard.edu, public data such as https://opendata.dc.gov or https://nces.ed.gov/datatools/.

2 Work Products

Your final project should include

- Social science data
- Application of data science methods of modeling or discovery (statistical inference, linear, generalized linear, generalized additive, Bayesian regression, machine learning, clustering, text methods, network data methods)
- Data analysis that is consistent with data science practices from class.
- Code that embodies modern data science workflow procedures.
- A report that clearly describes your social science question, the intellectual or political context of the question, the data collection procedure, interesting or important features of the data wrangling or processing steps, your statistical methods, and your results. It should have an abstract that summarises these features in a handful of sentences.
- A GitHub repository with your code, data (if applicable), and a report on your question, methods, results, and interpretations. These should be in folders `code`, `data`, and `docs`, respectively. You may include additional elements as needed.
- Optionally, use of cloud or distributed computing

3 Presentations

Presentations will take place Saturday morning. Each group will have 20-25 minutes for its presentation. Roughly half to two-thirds of this time can be formal presentation, with ample time for questions and discussion from instructors and colleagues at the end.

4 Other Guidelines

We expect there to be some variety in the length of the report, and groups should avoid adding extraneous information unrelated to their work. The report should provide readers enough context to a) understand the analysis, b) be motivated about the importance of the question, and c) comprehend the intellectual debate in which the paper participates. It is difficult to imagine doing this well, and describe your methods, results, and interpretations in under 5 pages. Aim for quality.

Any citation style is acceptable, as long as it's reasonable, consistent, and used in an academic field. We

recommend adopting the norms of the substantive field to which your paper should contribute.
Visualizations are not required, but are encouraged. They can add a great deal to the reader's ability to understand your problem or findings.
Talk to the instructors if you need help with any of these aspects and/or have other questions.

5 Groups

1	2	3	4	5	6	7
LucasA	Zeinabou	Mark	AndrewZ	JessicaK	Lauren	JessieG
Hannah	Bryce	Marc	Erin	Ethan	Olan	Kathleen
Eddie	Carine	Jocelyn	Tanesia	Cameron	Hubbert	Kelly
LucasG	Milika	Sophie				Kate