Final Project

Overview

The best way to learn to code and work with data is through applying the skills to your own research question. As a result, the final project is a major component of your performance assessment in this course. So you can get feedback throughout the semester, your final project grade consists of six components:

1) Topic proposal
2) Data source identification
3) Literature review and workflow plan
4) Peer review
5) Final report
6) Final presentation

See the syllabus for the point breakdown and due dates for each of these final project components.

The objective of the final project will be to complete a fully reproducible workflow that uses data to address your chosen environmental question. The project must illustrate all of the following tasks:

- Some form of data access / reading data into R
- Integration of multiple datasets to address the question
- Use of dplyr to manipulate and summarize the data in relevant ways
- Initial data visualization with ggplot2
- Final, publication-worthy visualization with ggplot2
- RMarkdown writeup, with final submission as both a clean html or pdf file, and a "show your under-the-hood work" version of the file
- Overall clean and clear presentation of the workflow, code, and explanation

1. Topic selection

We will be building toward the final project throughout the class. At this stage I need enough information from you to tailor the data sources we cover and the examples we use in class. A full project proposal, with questions and identified data sources, will be due later in the semester. For now, please submit your responses to the following prompts.

If you do not have your own data:

1) General topic of interest (e.g., climate, species conservation, evolution, genetic diversity, hydrology, human demography, etc)

2) Potential questions within that topic area (e.g., how do climate forecasts differ regionally? How does environmental change influence species composition?, how does phyogenetic diversity vary by system?, etc)

3) Potential datasets within that topic area (e.g.,specify relevant government databases, academic data repositories, etc)

If you do have your own data:

1) The questions you would like to answer with your data

2) The structure of the dataset (give as much information about the data as the str() function would return)

2. Data source identification

Please state your research question(s) and identify the specific datasets you will be using to address your question(s). I expect at least two datasets will be combined to address your questions. Provide the data sources (full data source reference, including web url) and sufficient metadata to convey who collected the data, how it was collected, and what each column contains. In addition, please provide an overview of the structure of the datasets (give as much information about the data as the str() function would return).

3. Literature review and workflow plan

To progress toward the final project, please prepare a literature review and develop a workflow plan. A description of each is below. We will peer-review these prior to the final project to help each other with the workflows.

Literature review

I expect the literature review will be around 5 well cited paragraphs that do the following:

1. Introduce the problem and explain why

- Set the stage for the problem
- Put the concept and question into context
- Lots of big-picture citations (such as reviews) in the first paragraph

2. Past work and data available on the project

- Who has addressed this problem, and what did they do it?
- What are the data available to address this problem?
- How has the data available and/or methods changed recently?

3. Purpose of the study

- Further refine your approach (e.g., what data will you combine, how will you address the question)
- Justify why this is needed now (e.g., visualization to test a new dimension of the question or better convey an old one)

4. Hypotheses/questions

- List these clearly and in a logical order
- Make hypotheses directionally using predictions (e.g. "I predict N will reduce plant diversity" rather than "I predict N will change plant diversity")

5. Literatured cited

- At least 10 relevant, peer-reviewed citations
- Citations are scientifically formatted (any format is acceptable, as long as you are consistent)

This is a nice reference on scientific writing: Turbek, Sheela P., Taylor M. Chock, Kyle Donahue, Caroline A. Havrilla, Angela M. Oliverio, Stephanie K. Polutchko, Lauren G. Shoemaker, and Lara Vimercati. "Scientific Writing Made Easy: A Step-by-Step Guide to Undergraduate Writing in the Biological Sciences." The Bulletin of the Ecological Society of America 97, no. 4 (October 2016): 417–26. https://doi.org/10.1002/bes2.1258.

Workflow plan

Describe the workflow you will use tidy your raw data, manipulate and summarize it in relevant ways, and visualize it. This should be a descriptive step-by-step description of what you plan to do to answer your research question(s). Please include any cleaning steps (e.g., "remove non-species such as 'miscellaneous litter' from the species column") as well as aggregation steps (e.g., "count the number of entries by plot and year to calculate species richness"). The goal here is to develop a logic to your workflow before you code.

4. Peer review

You will conduct a peer review of another student's final report, which will be due one week after the draft. You will complete the "Final Project Peer Review" form and submit the form to the author and on Blackboard.

5. Final report

Submit both a "cleaned-up, report-worthy"" html or pdf AND the "all work shown, echo = TRUE" files

I. Introduction (Literature review)

The literature review will be around 5 well cited paragraphs that do the following:

1. Introduce the problem and explain why

- Set the stage for the problem
- Put the concept and question into context
- Lots of big-picture citations (such as reviews) in the first paragraph

2. Past work and data available on the project

- Who has addressed this problem, and what did they do it?

- What are the data available to address this problem?
- How has the data available and/or methods changed recently?

3. Purpose of the study

- Further refine your approach (e.g., what data will you combine, how will you address the question)
- Justify why this is needed now (e.g., visualization to test a new dimension of the question or better convey an old one)

4. Hypotheses/questions

- List these clearly and in a logical order
- Make hypotheses directionally using predictions (e.g. "I predict N will reduce plant diversity" rather than "I predict N will change plant diversity")

II. Approach/Methods

- Provide a description of the data sources, including who collected the data and how it was collected
- Describe how you will use the data to address your questions
- Include R chunks with data import, data tidying, and any important preliminary visualizations (for the "cleaned up" version these will be in non-exported R code chunks; they should be shown in the "all work shown" version)

III. Results

- Report the key findings (but save the interpretation and contextualization for the discussion)
- Provide evidence to answer all questions or hypotheses posed in the introduction
- Include R chunks that contain data summaries and visualizations
- Have these chunks export captioned, publication-worthy figures

IV. Discussion

The discussion will be around 3-5 well-cited paragraphs that do the following:

1. Provide a big overview or summary of consequences of strongest results (1 paragraph)
2. Expand thoughts on results and hypothesis (2-4 paragraphs)
3. Discuss the limitations of the approach a/o potential future directions (1 paragraph)
4. Summarize the conclusion or take home message (1-2 paragraphs)

Goals for this section include:

- Address hypotheses (are they resolved?), questions, aims, and/or limitations
- Discuss the consequences or implications of results
- Make comparisons with previous findings (support or contradict)
- Announce your study's contribution to the current field
- Draw broader conclusions
- The comprehensive literature review should be in the introduction, here it is necessary to link to past work but no need to introduce too many new citations

V. Literature cited

Please follow MLA citation guidelines. This will include in-text, parenthetical citations and a literature cited section at the end. Aim for at least 10 peer-reviewed cited works.

This is a nice reference on scientific writing: Turbek, Sheela P., Taylor M. Chock, Kyle Donahue, Caroline A. Havrilla, Angela M. Oliverio, Stephanie K. Polutchko, Lauren G. Shoemaker, and Lara Vimercati. "Scientific Writing Made Easy: A Step-by-Step Guide to Undergraduate Writing in the Biological Sciences." The Bulletin of the Ecological Society of America 97, no. 4 (October 2016): 417–26. https://doi.org/10.1002/bes2.1258.

6. Final presentation

During the last class period, each student will deliver a 10 minute presentation on their final project. The presentation should provide the audience with a background on the topic that motivates the research question. The research question and a brief overview of the data and approach should be clearly stated. The presentation must include compelling visual representations and a summary of the main findings.