

Final Project

AFRE 891 SS 25 Michigan State University

Project Description

Through the second half of the semester, students will work on an independent project. The project can take one of two forms:

1. Replication

- Students may choose to center their project around replication of a published research paper.
- Students are tasked with identifying a paper of interest with publicly-available replication materials whose methods align with our covered course content.
- Students must then fully replicate the paper, and in the process create an R Markdown file that walks others through the full replication process.

2. Data Acquisition

- Students may alternatively choose to conduct a data acquisition project.
- Students choosing this option are tasked with using data acquisition methods covered in this course (e.g. web scraping, APIs) to obtain a dataset of their interest, creating a well-documented R Markdown file that conducts the data acquisition.

At the end of the semester, students will give an 8 to 10-minute, conference-style presentation on your project, in which you will summarize what you did and communicate your experience.

- For replication projects, this will involve discussing the motivation for choosing that paper, what the replication steps looked like, how the replication package was laid out originally, what you had to translate/modify to get it working in R, and whether you were able to replicate things as-is or if it took additional steps.
- For data acquisition projects, this involves discussing the motivation, identifying the source of the data, and walking us through the steps required for acquiring the data (the scraping workflow or how you interacted with the API).

Each presenter will be assigned a discussant, who will review the final project submission and provide a brief 3-minute oral feedback following the project's presentation.

While the final presentation is structured to mimic the format of an economics conference, I will only be grading on the portion of content that is relevant to our course content (I enjoy separating hyperplanes as much as the next person, but the emphasis of your presentation should be on the empirical work).

In addition to the final research presentation, students will submit **replication packages** (code and utilized data) that will allow for the replication of all figures and tables presented in the research presentation. If there are privacy or data sharing concerns regarding your data, students should make arrangements with me in advance of the final presentation.

Proposal

Due end of day Friday, February 28 At the end of February prior to Spring Break, each student is expected to submit a one-page **proposal**, outlining the proposed project and planned direction. The proposal

should contain

Paragraph 1: Introduction

- Brief motivating background
- Introduce planned project
 - If the project is a **data acquisition** project, give a brief overview of the desired final dataset
 - If the project is a **replication** project, identify the paper that you chose and why you selected it.

Paragraph 2: The Involved Steps

- Discuss your project's involved steps/scope
 - If the project is a **replication** project, state the figures/tables and the analysis you'll need to do in order to replicate them (can be a list)
 - If the project is a **data acquisition** project, state the motivation for compiling the proposed dataset and the source of the data, and any expected manipulation/wrangling needed to get it into a workable dataset. or inputs to a machine learning method, feel free to include them
-

Replication Package Material Submission

Due end of day Sunday, April 20 Prior to the final presentations, students must submit a **replication** package that includes all datasets and R scripts used to complete the project.

If this is an **replication** project, this should include

- The underlying data used for the project
- R Scripts/RMarkdown files used to produce all output that will be presented during the final presentation
- Readme Markdown or RMarkdown file that provides an overview of the folder structure, and states the tasks completed by each included R script and the order in which included scripts should be run to replicate output (if only one single script, discuss where to find the portions of code that produce the figure/table)
- A copy of or link to the paper you're replicating

If this is an **original data** project, this should include

- All R Scripts/RMarkdown files used to produce the dataset
 - Readme Markdown or RMarkdown file that provides an overview of the folder structure, and states the tasks completed by each included R script and the order in which included scripts should be run to replicate output (if only one single script, discuss where to find the portions of code that produce the figure/table) Make sure your scripts/RMarkdown files are appropriately commented, all necessary packages are identified/loaded, and all filepaths are defined relative to the main project folder.
-

Discussant

Due end of day Friday, April 27 Prior to the final presentations, discussants should make sure to do the following:

1. Attempt full replication. - Try to run all included scripts and see if you are able to replicate the project. For data projects that involve web scraping, you do not need to perform complete replication - just make sure that you are able to obtain at least a row or two of data using the supplied code. For replication projects, attempt to replicate all output (tables, estimates, and figures)
2. Prepare a 3 minute discussion (slides or no slides, your choice) to present to the class following the main project presentation - Comment on the overall project and what you liked about it (the idea, the ease of replication, etc.) - Whether or not you were able to successfully replicate the project, and if not at what steps you failed to fully replicate - Any ideas for future extensions or considerations

Presentations

Final Timing TBD (during Finals Week) Finally, prepare an 8 to 10-minute, conference-style presentation on your project. This presentation should follow the standard progression of a conference talk, providing motivation, describing the project goals or research questions, give an overview of the approach to data collection or replication, and discussion of challenges/takeaways (replication) or planned next steps (data project).

8-10 minute presentations are a difficult length, as they require you to efficiently communicate what you're doing, why you're doing it, and why it's exciting. I recommend keeping your deck small (aim for no more than 1 slide per minute) and focusing on what's new and interesting about what you're doing. You can take as given knowledge on anything that we've covered in class, so to the extent that you're paralleling lecture content you can give a brief mention and expand into more detail on anything that goes beyond what we've specifically covered.