



Teaching Advanced Data Science in R: Successes and Failures

Lisa Lendway, Assistant Professor, NTT
Macalester College, St. Paul, MN USA

Course logistics

- Advanced Data Science in R is a new course for Junior and Senior Applied Math and Statistics majors.
- This year, the course took place Mondays, Wednesdays, and Fridays from 11AM-1PM on Zoom for 7.5 weeks.
- Course material was posted to the course website, and students were expected to read through that on their own.
- Class time was dedicated to working on exercises or the final project, student presentations on R packages, and listening to guest speakers from various industries.

Course website: <https://advanced-ds-in-r.netlify.app/>

Successes and Failures

- Next, I will talk about successes and failures within each course learning objective.
- I'll end by discussing successes and failures that were unrelated to the course learning objectives.

Course website: <https://advanced-ds-in-r.netlify.app/>

Git/Github

Make using git and GitHub a habit. Get more comfortable with using it when collaborating with others - branching, pull requests, etc.

Successes

- Students used it for each of their assignments, providing a link to the repo each time.
- They were all collaborators on the group project.

Failures

- In general, they didn't use branches or make pull requests. This is likely because I didn't require that they do it.

Create a website

Create a website and add work from the course that you can share with future employers, advisers, collaborators, etc.

Successes

- They all created websites!

Failures

- Many were VERY basic, including only a landing page with their name. Next time, I should make more specific requirements.
- Most of them didn't choose to post anything on them. Again, I didn't require this.

Machine Learning methods

Become familiar with more machine learning models and methods and the associated R code.

Successes

- Students learned the {tidymodels} framework, stacking, interpretable machine learning methods, boosting, and using H2O in R.

Failures

- Reviewing machine learning via {tidymodels} was more difficult than I realized it would be since they were new to the code.
- We didn't have time for deep learning.

Learning on their own and teaching others

Gain confidence in learning new R functions on your own and experience teaching others to use them.

Successes

- Students gave group presentations on a package or set of functions. The list of topics I gave them to choose from was broad enough to satisfy all student interest.
- Students did a lot of problem solving on their own, without my help, to prepare their materials.

Failures

- I should check in with the group before their presentation to assure the topic covers the right amount of material and to make sure they aren't using out-of-date code.

Put a model into production

Understand what it means to "put a model into production".

Successes

- Students saved a model they created using the `saveRDS()` function and used the saved model in a shiny app.

Failures

- In the future, use Docker and the {plumber} R package to more closely resemble what it's like to put a model in production.

Importance of data ethics/justice

Appreciate the importance of data ethics/justice and integrate it into all aspects of data science.

Successes

- There was a reading or viewing exercise on each homework assignment about data ethics/justice.
- Student reflected on what they watched/read.

Failures

- Reserve some time in class for discussions or maybe even for watching/reading.
- The reflections weren't directly related to the problems on their assignment. Find ways to link them together better.

Other things that went well

- Students enjoyed having the speakers from industry every week. It gave them many different perspectives and connections.
- Class format worked pretty well. It was nice that they had a lot of time in class to work through problems, which is where they ran into a lot of questions.

What didn't go well?

Controllable:

- Needed to start the projects earlier so they had time to make some mistakes at the beginning and start over if they desired.
- Homework assignments were too long. Make those shorter and make project work part of the weekly assignments.
- SO MUCH more I wanted to cover and in SO MUCH more detail. It might be better to skip some topics or try to incorporate some into their machine learning course. For example, if they learned {tidymodels} in that course, I could skip it in this course.

Uncontrollable:

- The 7.5 week intense module, rather than a 15 week semester felt too condensed.

THANK YOU!

Thanks for listening to this brief talk. If you teach similar material, I'd love to chat about your experiences. Reach out by email, llendway@macalester.edu, or on Twitter @lisalendway.

Course website: <https://advanced-ds-in-r.netlify.app/>