

Projects in Data Science: Teaching Skills

Fall 2025, Shilad Sen

Data Science technologies move fast. Learning each one independently, using "official" product documentation, would be challenging!

Fortunately, this isn't the case. A supportive learning culture has developed across many programming communities, particularly data science. Data Scientists regularly create tutorials, demos, and learning guides for new technologies. They do so to give back, strengthen their connection to a community or team, bolster their credibility and visibility, and learn new skills themselves.

In this assignment, your group will gain practice developing a learning guide for an "advanced" data science technology that is interesting to you all. You will test out your guide with a volunteer and refine it. Shilad may use some of your guides as activities later in the semester.

Step 1: Select a Topic

Send an email (or DM) to Shilad with a topic your team would like to teach the class on. I've included several example topics at the end of this document. Shilad will respond quickly with approval or ask you to suggest an alternative. You are welcome to synthesize other guides as long as you credit them.

Step 2: Draft the Learning Guide

Draft the tutorial as an RMD (probably!) or Google Doc (if not in R). Here are some guidelines:

- Include a **clear title and summary** describing what the reader will **learn**. If possible, make the writing conversational.
- The guide should include "**open-ended**" **tasks** that allow users to implement something that is compelling or interesting to them.
- The basic guide should take about 45 minutes for your classmates to walk through including open-ended tasks, with a clear sense of completion for the last basic task.
- The guide should include at least one optional **advanced task** beyond the basic task for learners who are curious and/or quick.
- A tutorial should give the reader an idea of technology's possibilities and the overall structure of possibilities it enables. Once readers know this, they can track down details important to their problems.

You will upload your first draft to Moodle.

Step 3: Test Drive the Learning Guide

Find a peer to follow your guide. It can be an MSCS major or anyone with the requisite skills to complete your guide. You will walk them through your guide. Let your classmate type, and try not to help them unless they need help or your guide is unclear. Take careful notes. I will ask you to share your notes.

After your guide tester finishes, ask them: How well did they learn what you said they would learn? Is there anything confusing? Is there anything they didn't learn that they wish they would have?

Meet with your group and compare notes on what went well and what needs improvement. Based on your experiences with classmates, refine your guide.

Upload your final Knit HTML to Moodle, and make sure it includes:

- Your final draft
- At the bottom:
 - Share each of your individual notes from your sessions with classmates
 - Describe the ways you improved your guide based on partners' feedback

After you are done, consider sharing your guide with the world! It's both valuable to the data science community and good branding for you!

Expectations:

| |
|---|
| Topic Selection |
| Initial topic is submitted within 2 days of assignment release |
| Responses to Shilad's emails happen quickly |
| Rough Draft |
| Draft is submitted on time |
| Draft has been completed in good faith and is suitable for feedback. |
| Final Draft |
| Learning goals are clear, complete, and accurate |
| Writing is clear, conversational, and engaging |
| Guide is easy to understand |
| Guide contains appropriate places for individuals to personalize their work |
| Careful, accurate notes were taken by team members |
| Team has appropriately responded to guide tester feedback |

Topic Ideas

- Text mining & analysis / Sentiment analysis (bag of words, stop words, etc.) (tidytext package)
- Basics of neural networks (nnet/tidymodels packages, keras/tidy models packages)
- Any LLM-related skills we haven't covered
- Python Pandas for tidyverse experts
- Interpretable Machine Learning (iml package)
- Creating an R Package (Host your data in R package)
- GIS / Mapping in R with sf package
- Gganimate
- Plotly