A $\sum$ of Introduction to data science

This course has run a couple of times since I got here in 2021.

It's fairly stable. Elements come from a variety of pre-2016 books, when R was still untainted by the "tidyverse". Since, the language has been tainted by an unfortunate artificial split in two subcultures.

This is not Python 2 vs. Python 3 — that was a central design decision — while the splitting of R is due to the machinations of RStudio (now "Posit"), the company who got big with its IDE for R and R notebooks.

Where to go from here?

1. Experiment with using DataCamp as the "textbook" instead of splitting my attention.
2. Python, already added at the end of DSC105, becomes "the other language".
3. A number of special topics will be covered:
   - Natural language processing
   - Importing and exporting data
   - Data science on the command line
   - Integration of SQL and C++
4. The project will focus on exploring and presenting (interactively) an R and an equivalent Python package.
5. I'll bring Emacs + Org-mode back as the main coding platform.