Go through

(1) homework problems,
(2) R code on Canvas,
(3) lecture slides,
(4) book.

Comprehensive exam. Similar in style to midterm.

Topic by topic (go backwards).

1. logistic regression and CART basics
2. database queries/SQL basics. How do you do these? Look at imdb example
3. What are JSON and XML? Why do you need to know about them?
    a. JavaScript Object Notation, is a minimal, readable format for structuring data. It is used primarily to transmit data between a server and web application, as an alternative to XML.
    b. XML: a metalanguage which allows users to define their own customized markup languages, especially in order to display documents on the Internet.
4. Text mining; Naive Bayes. What are methods for doing this? Explain the ideas clearly. E.g. spam versus ham
5. Regular expressions: how to write them to match patterns, solve problems
6. Representation of numbers, characters, html
7. Monte Carlo: What is it? how to use it to solve problems. Writing code to implement Monte Carlo
8. Functions, scope/environments
9. Lists
10. Basic data structures
11. Graphics, creating tables, exploratory data analysis