

Proposed Outline

Data Hack 2022 Outline

Learning Objectives: Students will learn data wrangling using the R programming language in addition to creating a solid foundation for good research practices (e.g., organization/reproducible research). Completion of the course will result in an understanding of how to begin the researching process of finding data, cleaning it, merging with other sources, analyzing, and creating effective visualizations.

Final Presentation Topic: Policy - choose any data set necessary to convey why this topic should be on the agenda for policy change.

Students will work in groups of 3-4 people on a project that conveys an important issue that deserves policy attention. Students will create visualizations/statistics using their own data that has been found and cleaned themselves.

Proposed Schedule

- Week 1: Introduction to R and Piping
 - Topics (R): Data types (characters/strings/logicals), importing data (excel/csv emphasis), common statistics (mean/sd/var), introduction to piping with mutating variables/summarizing/filtering.
 - Topics (Organization): Rprojects for file path management/collaboration, naming variables (e.g. snake_case and informative), the importance of READMEs, coding etiquette (leaving spaces, commenting etc.).
- Week 2: Continuation of Cleaning
 - Topics (R): Grouping, further filtering with conditions, cleaning dates,
 - Topics (Organization): Using RMarkdown for reproducible documents.
- Week 3: Introduction to Visualization with ggplot2
 - Topics (R): ggplot2 grammar of graphics, facet wraps, aesthetic mappings, reasons to use ggplot2.
 - Coding Check-in: Have us evaluate your code! Give us an example of a homework assignment you did (or cleaning exercise) and let us give feedback.
- Week 4: Regular Expressions and Scraping Documents
 - Topics (R): Using regular expressions along with extract, str_detect, str_replace, separate. In class activity of scraping a PDF documents using the tools learned.
 - Quick presentation on the DO's and DON'Ts of visualizations.
- Week 5: Spatial data with SF
 - Topics (R): sf package, raster vs. vector data, heatmaps, mapview.
 - Check-in: Does every group have their topic/path to data? Should have data source identified by this point.

- Week 6: Functions, Packages, and Programming
 - Topics (R): using functions to manage coding tasks and improve readability, if-statements, for-loops, creating packages to manage functions.
 - In class activity: create a package yourself!
- Week 7: Webscraping
 - Topics (R): rvest, css selectors, brief mention of Rselenium. For-loops and if-statements.
 - In-class activity of scraping data from a webpage and creating an effective visualization. Prize is a coupon for 1 redeemable handshake.
- Week 8: Merging Data + Regressions + Presentation of Models
 - Topics (R): merging using left join, right join, inner join, regressions with the fixest package, modelsummary and kableextra for presentation purposes.
- Week 9: Git/Github? - seems late in the quarter, but also don't really think they should be expected to use this.
- Week 10: Presentations + Reception