STAT 351: Statistical Computing 2: Data Management and Visualization

Instructor: **email:**

Office Hours:

Credit Hours: 3

Other References:

- Rmarkdown: The Definitive Guide; available free online at https://bookdown.org/yihui/rmarkdown/

- Blogdown: Creating websites with R Markdown; available free online at https://bookdown.org/yihui/blogdow

- Tidy Text Mining; available free online at https://www.tidytextmining.com/

Prerequisites: Stat 251

Course Description: Intermediate statistical programming course. Computational skills for management, visualization and analysis of large and complex data which are necessary for modern statistics. Includes a wide range of topics necessary for data analytics, including harvesting data from websites and common data structures, setting up and working with databases, and designing interative data displays. Students will create an online portfolio to display projects which demonstrate their statistical skills.

Course Goals:

- Work with data throughout its lifecycle - cleaning, integrating with other data, and interactive visualization.

- Design and set up databases

- Utilize application programming interfaces for data access

- Present data and analysis products on the web in a reproducible format

- Design and construct interactive visualizations and documents to communicate statistical information and the consequences of decisions effectively

- Work with specialized types of data, visualizing and analyzing them appropriately

Grading:

Assignment(s)	Contribution to Final Grade
Blog posts and Homework assignments	40%
Project proposal	10%
Project rough draft	20%
Final project	30%

Grade	Final Percentage Range
A	94.0-100
A-	90.0-93.99
B+	88.0-89.99
B	84.0-87.99
B-	80.0-83.99
C+	78.0-79.99
C	74.0-77.99
C-	70.0-73.99
D+	68.0-69.99
D	64.0-67.99
D-	60.0-63.99
F	<60.0

Grading Scale: (table above)

Course Expectations: In this course, you are expected to have professional behavior. You are expected to attend all class meetings, be curious, ask questions, seek opportunities to learn, and be open and responsive to constructive feedback. In addition:

- Be an active participant–statistics is not a spectator sport!

- Be committed, take your work seriously

- Engage with the in-class activities and homework sets

- Help others–if you understand the material being discussed, practice your mentoring skills. This does not means sharing answers, but instead helping others understand the concepts.

- Complete assigned readings.

You are also expected to exhibit a professional demeanor (language, attitude) toward others. Disagreement during discussions is welcome and often productive in developing a deeper understanding of the concepts being discussed. However, disagreement does not warrant yelling or disrespectful language or behavior. Unprofessional behavior will not be tolerated, and appropriate actions will be taken to prevent future occurrences.

Blog posts and Homework assignments: You will build and maintain a portfolio using blogdown in this class; each week, you will complete a small analysis, written response, or other demonstration of skills learned in class and applied to new data. These assignments are intended both to reinforce lectures and to help you build a platform that you can use to demonstrate your skills to employers.

Project proposal: A proposal for your semester project will be submitted in week 6 of the course. Your proposal should include a research question, at least two data sets, an assessment of data quality, and a proposed analysis strategy. The proposals will be evaluated by your peers as well as the instructor.

Project rough draft: A rough draft of your semester project will be submitted in week 10 of

the course. The rough draft should include a detailed description of your data (sources, contents), documentation of the data cleaning and integration process, and exploratory analysis of your data, with an outline of the statistical analysis method selected and anticipated trouble spots.

Final project: Written report and presentation of project. Projects will be peer-evaluated as well as instructor-evaluated.

Emergency Response:

- Fire Alarm (or other evacuation): In the event of a fire alarm: Gather belongings (Purse, keys, cellphone, N-Card, etc.) and use the nearest exit to leave the building. Do not use the elevators. After exiting notify emergency personnel of the location of persons unable to exit the building. Do not return to building unless told to do so by emergency personnel.

- Tornado Warning: When sirens sound, move to the lowest interior area of building and seek shelter. Stay away from windows and stay near an inside wall when possible.

- Active Shooter

  - Evacuate: if there is a safe escape path, leave belongings behind, keep hands visible and follow police officer instructions.

  - Hide out: If evacuation is impossible secure yourself in your space by turning out lights, closing blinds and barricading doors if possible.

  - Take action: As a last resort, and only when your life is in imminent danger, attempt to disrupt and/or incapacitate the active shooter.

- UNL Alert: Notifications about serious incidents on campus are sent via text message, email, unl.edu website, and social media. For more information go to: unlalert.unl.edu.

- Additional Emergency Procedures can be found here: emergency.unl.edu.

Tentative Course Outline:

Week	Topics
1	Blogdown, continuous deployment
2	Data documentation: codebooks, reproducibility, sources
3	Data storage: databases and database operations (queries, joins), setting up a SQLite database, documenting database relationships and schema
4	A brief introduction to HTML, Javascript, and CSS: data on the web
5	Web scraping, application programming interfaces, REST/SOAP protocols
6	Advanced uses of literate programming - parameterized reports, interactive markdown documents, bookdown; project proposal due
7	Integrating data into web pages: web graphics, interactive tables, and advantages of web-based data displays
8	Interactive graphics (with Shiny, plotly, Tableau, d3)
9	Software engineering principles: modular design, tests, code documentation, versioning, release cycles
10	Special data types: text mining, project rough draft due
11	Special data types: images
12	Special data types: time
13	Special topics (TBD based on class interest)
14	Project presentations
15	Project presentations