# STAT 251: Statistical Computing 1: Data Wrangling

**Instructor:** **email:**

**Office Hours:**

**Credit Hours:** 3

**Other References:**

- R for Data Science (Wickham & Grolemund)
- Python Data Science Handbook (Vanderplas)

**Prerequisites:** Stat 151

**Course Description:** Techniques for processing, cleaning, and visualizing messy data. Topics include data reduction strategies, data transformations, combining multiple data sources, and special types of data (text, spatial, dates and times, hierarchical).

**Course Goals:**

- Use appropriate visualizations to explore and assess data and its applicability to a problem
- Write code to reshape and reformat moderately complex and/or messy data in a reproducible manner.
- Create graphical displays to explore data, assess statistical models, and present model results
- Adapt pre-existing code for sophisticated visualizations to new data.
- Implement an algorithm or procedure for data modification given in pseudocode
- Write pseudocode to describe and document modifications made to the dataset
- Access documentation and source code to determine how software works (or why it doesn't)
- Identify problems in a dataset that limit the analyses which are appropriate for the data.

**Grading:**

| Assignment(s) | Contribution to Final Grade |
|---:|:---|
| Midterm | 20% |
| Participation | 10% |
| Homework | 50% |
| Project | 20% |

**Grading Scale:**

| 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 |

**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.

**Midterm:** One in-class, open note exam will be given during the course of the semester, on dates noted on the Tentative Course Outline on the last page. This exam will require that you write code to solve problems utilizing the course material. You are expected to take exams at the scheduled times. If this is impossible due to extreme circumstances (illness, death in the family, previously scheduled activities vital to academic program), please notify me and provide appropriate documentation. No make-up exams will be given if I am not notified prior to the examination. You will be required to obtain a note from your physician or advisor explaining the nature of the conflict.

**Participation:** All students are expected to attend and fully participate in class activities. Participation will be determined based on a combination of class attendance and activities.

**Homework:** Approximately 8-10 homework assignments will be made over the course of the semester. You will typically have one week to work on each of the assignments. The only way to learn how to program is to practice working problems, and homework is therefore an essential part of the course. Homework must be submitted in the file format specified, and should run or compile as submitted.

**Project:** The course project will allow students to apply the course material to data which interests them. More details about the project will be given after Exam 1; there will be at least one draft submission in the second half of the semester to provide you with feedback about your project before the due date.

**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 | Tools for code and data management (Rmarkdown, version control, reading in data) |
| 2 | Code Wrangling (commenting, basic functions, getting help, debugging) |
| 3 | Grammar of graphics (layers, variable mappings, aesthetics) |
| 4 | Types of charts |
| 5 | Working with subsets of data (select, filter, logical operators) |
| 6 | Grouping, transformations, and summaries |
| 7 | Exam 1, Missing data (detecting missingness, types of missingness, imputation and other solutions) |
| 8 | Reshaping data (structural transformations, presenting data summaries in tables) |
| 9 | Dates and times, String operations |
| 10 | SQL joins (full, left, right, anti, semi) |
| 11 | Getting data from the web |
| 12 | Complex data structures (hierarchical data, json, xml) |
| 13 | Functional programming (Split-apply-combine, map-reduce) |
| 14 | Spatial data (maps, heatmaps, contour plots, common spatial formats), Project due |
| 15 | Multivariate data and dependence diagnostics |