LTI/DSP 110:

Introduction to Data Science

# INSTRUCTOR & CLASS INFORMATION

Instructor: Harrison Dekker

Email: hdekker@uri.edu

Time: Tuesday and Thursday, 8:00 AM

Classroom: Library 130

Office hours: Monday, Wednesday, Friday, 8:00 AM

Office: Library, AI Lab

# COURSE DESCRIPTION

To study problems in modern society we need to understand how to use the vast amounts of data available in the world today. Modern literate citizens need to understand how we interpret data by considering how data are generated and analyzed. In this course students gain experience using data to answer questions about the world. Students will learn to obtain, explore, visualize, and analyze data across an array of fields, including physical, natural and social sciences. Real data will be used to address real, pressing, social questions. This course will focus on the R computing language. No prior programming experience is necessary.

# LEARNING OUTCOMES

At the end of this course students enrolled in LTI/DSP 110 should be able to…

* Interpret basic graphs and tables.
* Formulate a data-oriented research question.
* Find trustworthy, publicly accessible data.
* Use the R programming language to manipulate, summarize, and visualize numeric and text data.
* Conduct a data analysis project and communicate the results using an organized, documented, and reproducible workflow.
* Identify potential ethical issues associated with a data project or activity.

In addition, the course meets the requirements for General Education Learning Outcome B3 – Apply the appropriate mathematical, statistical, or computational strategies to problem solving.

# TEXTBOOKS

OpenIntro Statistics - Fourth Edition*.* Diez, Barr, and Cetinkaya-Rundel. 2019.Available from <https://www.openintro.org/book/os/>

R for Data Science. Wickham and Grolemund. 2017. Available from <https://r4ds.had.co.nz/>

Optional: Modern Dive: Statistical Inference via Data Science. Ismay and Kim. 2019. Available from <https://moderndive.com>

# CLASS SESSIONS

The syllabus lists the topics that will be covered in each class session and the activities (readings, videos, quizzes) that should be completed before class time. Classes will include lectures as well as individual and group activities.

# PROBLEM SETS

Problems sets will be made available at least one week before the stated due date. Assignments will generally involve taking an assigned data set and doing an analysis similar to in-class work but may also include short research and writing projects, as well as exercises drawn from the textbooks. Seven problem set will be assigned over the semester and the lowest grade will be dropped. Late work will not be accepted.

# QUIZZES

Quizzes will be conducted on Brightspace and must be completed before class. Questions will be based entirely on material from assigned readings and videos.

# GROUP PROJECT

The group project will involve working in groups of three or four students, choosing a research topic, selecting a relevant data set, and conducting an exploratory analysis, and presenting the results in class. Deliverables will include a short topic proposal presentation and written abstract due in week 9, a six-minute final presentation given in class in week 13 or 14, and a complete set of code and data files to be submitted before the last day of class. Detailed instructions for the project will be available on Brightspace at the beginning of the term.

Students will have an opportunity to express a preference for potential group partners, but the instructor will have final say in group composition.

# MIDTERMS & FINAL EXAM

The two midterms will be take-home and administered on Brightspace. The final exam will be conducted in person.

# ASSIGNMENTS & GRADING

|  |  |
| --- | --- |
| Quizzes (12) | 10% |
| Problem Sets (6 highest scores) | 30% |
| Group Project | 20% |
| Midterms (2) | 20% |
| Final Exam | 20% |

A: 100-94% A-: 93-90% B+: 89-87% B: 86-83% B-: 82-80%

C+: 79-77% C: 76-73% C-: 72-70% D+: 69-67% D: 66-60% F: <60%

# OUTLINE

[subject to change]

|  |  |
| --- | --- |
| **Unit 1: Data Basics** | |
| *Week 1* | |
| Thursday,  September 9 | Course and subject overview: Computing requirements; Web and library resources for this course |
| *Week 2* | |
| Tuesday, September 14 | Numerical and categorical variables (continuous, discrete, ordinal); Topics: Associated variables; Independent variables   * Chapter 1.1 - 1.3 (*OpenIntro Statistics)* * Video: ["Case Study: Using Stents to Prevent Strokes"](https://www.youtube.com/watch?list=PLkIselvEzpM6pZ76FD3NoCvvgkj_p-dE8&v=nEHFF1ADpWE) * Video: ["Data Basics: Observations, Variable, and Data Matrices"](https://www.youtube.com/watch?v=Mjif8PTgzUs&list=PLkIselvEzpM6pZ76FD3NoCvvgkj_p-dE8) * Video: ["Data Collection Principles"](https://www.youtube.com/watch?v=2N_bkiyTiXU&list=PLkIselvEzpM6pZ76FD3NoCvvgkj_p-dE8) * Video: ["Designing Experiments"](https://www.youtube.com/watch?v=g7JGe_ykB3I&list=PLkIselvEzpM6pZ76FD3NoCvvgkj_p-dE8) * **Quiz 1** |
| Thursday,  September 16 | Topics: Explanatory variables; Observational and experimental studies; Generalizable experimental results; Correlation and causation.   * Chapter 1 (*R for Data Science*)   **REMINDER: Problem Set 1 due 9/19 midnight** |
| *Week 3* | |
| Tuesday,  September 21 | Topics: Scatterplots; shape/center/spread/outliers; distribution shapes; histogram and box plots   * Chapters 2.1-2.2 (*OpenIntro Statistics*) * **Quiz 2** |
| Thursday,  September 23 | Topic: Data science ethics   * *Data Ethics Canvas* *User Guide* (Brightspace) * Video: ["Introduction to Data Ethics Canvas"](https://www.youtube.com/watch?v=wwEvtjMiW0Q)   **REMINDER: Problem Set 2 due 9/26 midnight** |
| *Week 4* | |
| Tuesday,  September 28 | Topics: How to find data sources; How to find published research based on a specific data source   * “[Eleven quick Tips for Finding Research Data](https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006038)” * [“Evaluating Statistics & Data - Data, Datasets, and Statistical Resources - Research Guides at Carleton College"](https://gouldguides.carleton.edu/c.php?g=146834&p=964943) * “Data Reference Worksheet” (BrightSpace) * **Quiz 3** |
| Thursday,  September 30 | Topics: Manage a research project; files and directories; TIER Protocol   * [“Making Data Wrangling Useful Again”](https://www.sciencedirect.com/science/article/pii/S2405896315001986) * [“TIER Protocol: Specifications”](https://www.projecttier.org/tier-protocol/specifications-3-0/) * Chapter 8 (*R for Data Science)*   **REMINDER: Problem Set 3 due 10/3 midnight** |
| **Unit 2: Exploring Climate Data** | |
| *Week 5* | |
| Tuesday,  October 5 | Topics: Case studies: climate data; Use data files with R; Identify data and variables   * ["Here’s Where Australia’s Destructive Wildfires Are Burning"](https://www.nytimes.com/interactive/2020/01/02/climate/australia-fires-map.html) * ["Australia fires: A visual guide to the bushfire crisis](https://www.bbc.com/news/world-australia-50951043)” * **Quiz 4** |
| Thursday,  October 7 | Topics: Data visualization with R and ggplot2: syntax, aesthetic mappings, facets, geometric objects, transformations   * Chapters 2, 3.1-3.7 (*R for Data Science)*   **REMINDER: Complete take-home Midterm #1 by 10/10 midnight** |
| *Week 6* | |
| Tuesday,  October 12 | Topics: Data visualization with R and ggplot2: labels, annotations, scales, zooming, themes, saving plots)   * Chapter 28 (*R for Data Science)* * **Quiz 5** |
| Thursday,  10/14 | Topics: Tidy data introduction, data transformation (filter, arrange, select)   * Chapters 4, 5.1-5.4 (*R for Data Science)*   **REMINDER: Problem Set 4 due 10/17 midnight**  **REMINDER: Group project assigned: See Brightspace for details** |
| *Week 7* | |
| Tuesday,  October 19 | Topics: Data transformations: mutate, summarize, grouped summaries, grouped mutate; introduction to Markdown   * Chapters 5.5-5.7 (*R for Data Science)* * **Quiz 6** |
| Thursday,  October 21 | Topic: Communication: RMarkdown, knitr   * Chapter 27 (*R for Data Science)*   **REMINDER: Problem Set 5 due 10/24 midnight** |
| **Unit 3: Exploring Public Health Data** | |
| *Week 8* | |
| Tuesday,  October 26 | Topic: Case studies: public health data   * [“Why is life expectancy in the U.S. lower than in other rich countries”](https://ourworldindata.org/us-life-expectancy-low) * **Quiz 7** |
| Thursday,  October 28 | Topic: Exploratory data analysis -- part 1   * Chapters 7.1-7.4 (*R for Data Science)*   **REMINDER: Group Project topic proposals due in class on 11/2** |
| *Week 9* | |
| Tuesday,  November 2 | Topics: Exploratory data analysis -- part 2; Group project topic proposal presentations   * Chapters 7.5-7.7 (*R for Data Science)* * **Quiz 8** |
| Thursday,  November 4 | Topic: Temporal data   * Chapters 16.1-16.4 (*R for Data Science)*   **REMINDER: Take-home Midterm #2 due 11/7 midnight** |
| *Week 10* | |
| Tuesday,  November 9 | Topics: Advanced tidy data; table pivoting; reshaping   * Chapters 12.1-12.6 (*R for Data Science)* * **Quiz 9** |
| Thursday,  November 11 | Topics: Table formats; Cross tabulation   * Video: "[How to create a crosstab with row and column percentages](https://www.youtube.com/watch?v=tGVTIrNeWio)" * [“tabyls: a tidy, fully-featured approach to counting things”](https://cran.r-project.org/web/packages/janitor/vignettes/tabyls.html)   **REMINDER: Problem Set 6 due 11/14 midnight**  **REMINDER: Group project update due by 11/12 midnight** |
| **Unit 4: Exploring Social Survey Data** | |
| *Week 11* | |
| Tuesday,  November 16 | Topics: Case studies: social science data; Introduction to survey data   * “The simple truth about the gender pay gap”   + [Summary](https://www.aauw.org/resources/research/simple-truth/)   + [Full Report](https://www.aauw.org/app/uploads/2020/02/AAUW-2018-SimpleTruth-nsa.pdf) * **Quiz 10** |
| Thursday,  November 18 | Topics: Categorical variables; General Social Survey   * Chapters 15.1-15.3 (*R for Data Science)* * ["GSS Data Explorer: Key Trends"](https://gssdataexplorer.norc.org/trends) * Video: ["How to use the General Social Survey…"](https://www.youtube.com/watch?v=hyirTanipAc)   **REMINDER: Problem Set 7 due 12/5 midnight** |
| *Week 12* | |
| Tuesday,  November 23 | Topics: Modifying factor order; factor levels   * Chapters 15.4-15.5 (*R for Data Science)* * **Quiz 11** |
| Thursday,  November 25 | *Thanksgiving Break – No Class* |
| **Unit 5: Exploring Non-numeric data** | |
| *Week 13* | |
| Tuesday,  November 30 | Topic: Strings   * Chapters 14.1-14.2 (*R for Data Science)* * **Quiz 12** |
| Thursday,  December 2 | Topic: Text mining   * Chapters 1.1-1.6 (*Text Mining with R*)   **Group presentations #1** |
| *Week 14* | |
| Tuesday,  December 7 | Topics: Course Wrap-up  **Group presentations #2** |
| TBD | **Final Exam** |

# ADDITIONAL INFORMATION

# POLICIES

Any student with a documented disability should contact me as soon as possible so that we may arrange reasonable accommodations. As part of this process, please be in touch with [Disability Services for Students Office](http://www.uri.edu/disability/dss/) at 330 Memorial Union, 401-874-2098.

Students are expected to treat faculty and fellow classmates with dignity and respect. Students are responsible for being familiar with and adhering to the published “Student Code of Conduct” which can be accessed in the [University Student Handbook](http://web.uri.edu/studentconduct/student-handbook/).

Students are expected to be honest in all academic work. A student’s name on any written work, quiz or exam shall be regarded as assurance that the work is the result of the student’s own independent thought and study. Work should be stated in the student’s own words, properly attributed to its source. Students have an obligation to know how to quote, paraphrase, summarize, cite and reference the work of others with integrity. The following are examples of academic dishonesty.

* Using material, directly or paraphrasing, from published sources (print or electronic) without appropriate citation
* Claiming disproportionate credit for work not done independently
* Unauthorized possession or access to exams
* Unauthorized communication during exams
* Unauthorized use of another’s work or preparing work for another student
* Taking an exam for another student
* Altering or attempting to alter grades
* The use of notes or electronic devices to gain an unauthorized advantage during exams
* Fabricating or falsifying facts, data or references
* Facilitating or aiding another’s academic dishonesty
* Submitting the same paper for more than one course without prior approval from the instructors.

If you need to miss class for any sanctioned reason (e.g. religious holiday, University-sanctioned event, serious illness, accident, personal tragedy, disability) please contact me in advance if possible to make alternate arrangements. If you miss class for an unsanctioned reason please discuss this with me to determine whether arrangements can be made. If the reason is too personal to discuss with me please discuss options with your academic advisor.

**Classroom Protocol**

* Be prepared to discuss assigned readings.
* Be kind to others. Remember that people have different backgrounds and experiences.
* If you must come in late, please do not disrupt the class. Please ensure all cell phones, pagers, and other electronic devices are set to silent.
* Students are expected to come to class. If you are unable to attend due to illness, severe weather, religious holiday, personal or family emergency, or sanctioned University event please let me know as soon as possible.