Course Syllabus for DS 701: Exploratory Data Analysis

NOTE: This syllabus document contains the basic information of this course. The most current syllabus is available in the full course.

Course Description

This course introduces data science and highlights its importance in decision-making. Students will learn how to analyze data using the R programming language. During the course, students will learn how to import data into R, tidy it, conduct exploratory data analysis, develop visualizations, and draw statistical inferences. The course aims to teach data wrangling, visualization, and exploration with R.

Course Objectives

By the end of this course, you will be able to:

- Comprehend fundamental concepts in data science and analytics.
- Use R Studio for programming, using packages, debugging, and knitting .Rmd files.
- Effectively utilize packages from Tidyverse for R.
- Read/write data of various formats using R.
- Create effective visualizations using ggformula.
- Perform linear and logistic regression, and interpret and graph the model.
- Perform exploratory data analysis with the dplyr package in R.
- Perform data cleaning, management, and manipulation with R.
- Use conditional statements, loops, and control flow logic in R.
- Interpret p-values and confidence intervals.

Course Components

Homework 1, 2, 4-10: First, solve the problems in the provided .Rmd file. When you are done, submit your answers to the selected questions in the homework quiz for auto-grading. You are allowed three attempts (no time limit), and your highest score is counted. After the assignment is closed, everyone who submitted the assignment will gain access to a partial key, which you can use to check your work on the open-ended questions. The questions that do not appear on the homework quiz will not be graded, but you are expected to understand them to help prepare you for the final project and other courses in the UW Data Science program.

Quiz on Control Flow: This quiz will assess your understanding of for loops, while loops, and if/else loops. You will have 60 minutes to complete four questions. You are allowed two attempts. (Your second attempt will involve a new random selection of questions.) The higher of your two scores will be counted. The quizzes are graded automatically, so you will receive your score immediately. The problems in this quiz

follow a similar style to the last three problems on the Practice Quiz on Control Flow, so you are encouraged to complete that activity before taking this quiz.

Discussions: Discussions are your opportunity to contribute to and learn from your peers. We will engage in several discussions throughout the semester, as outlined in the course calendar. Note that every assessed discussion requires two posts: one for your original contribution to the discussion's central issues and at least two responses to your peers. Your initial post counts as 75% of each discussion grade, while your replies/peer responses count for the remaining 25%. To create a professional, open communication environment, you are expected to follow these online discussion guidelines.

Final Project: The final project for the course simulates a real-world scenario. You will utilize the R programming skills you honed during the course to perform exploratory data analysis and derive inferences from a given dataset.

Late Submissions (Assignments, Discussions): You are expected to submit all evaluated work (including graded discussions, assignments, and the final project) on time. In cases of extremely extenuating circumstances (i.e., documented circumstances clearly beyond your control), you may request late work supported with documentation. First, get approval from the instructor. Second, submit the late work. Third, notify the instructor of your submission. Late work not following the procedure will not be graded. No exceptions.

IMPORTANT: The course late submission policy does not apply to the final project. Late submissions are not allowed for the final project under any circumstances, so please plan accordingly.

Grading

Your mastery of course content is assessed using a variety of methods:

Activity	Percentage
Piazza Registration and VLD Test	2%
Self-introductions	3%
Homework 1, 2, 4-10 (5% each)	45%
HW3 Discussions 1 and 2 (5% each)	10%
Quiz on Control Flow	5%
Final project	35%
Total	**100%**

Final grades are assigned using the following scale:

90–100%	A
80–89%	B

60–79%	C
0–59%	F