# Course Syllabus: Visualizing & Analysing Data with R

## Hult University Summer 2023

Instructor: Ted Kwartler, MBA

Email: [edwardkwartler@fas.harvard.edu](mailto:edwardkwartler@fas.harvard.edu) [edward.kwartler@hult.edu](mailto:edward.kwartler@hult.edu)

Office Hrs: Available upon request

## Important URLs:

* Canvas

<https://mycourses.hult.edu/courses/3411791>

* The Github repository allows you to get all scripts, PowerPoints and data sets throughout course. For those not familiar with github, think of it like a shared drive similar to SharePoint or Dropbox but with added functionality for data and computer science. <https://github.com/kwartler/Hult_Intro2R>

### Prerequisites:

* Textbook[optional]: R for Everyone
* Software: R & R-Studio
* Access to git software to download data sets and class material or ability to download directly from the Internet

## Learning Objectives:

This course is an introductory examination into the principles and techniques of analytics using R, the IDE R-Studio and version control with git.

**Course Topics:**

1. Intro to R, R environment, installing libraries and loading packages
2. Basic objects in R, different data types, testing and changing types, importing data into the R environment
3. Basic data mining in R
4. Creating frequency histograms, analyzing different distributions, visualizing data
5. Basic machine learning algorithms
6. Creating automated report using R markdown, report streamlining
7. Creating an interactive dashboard in R

**Course Learning Outcomes:**

* CLO1: Learn how to program in R and how to use R for effective data analysis
* CLO2: Access online resources for R and import new function packages into the R workspace
* CLO3: Import, review, manipulate and summarize data-sets in R
* CLO4: Perform appropriate statistical tests using R

### Attendance:

Regular attendance is essential to the successful completion of this course. You are responsible for material covered in class even if you have not attended class. You should plan on spending at least three hours of independent study for each hour of class attendance.

### Code of conduct:

This course expects you to uphold and report violations of the Hult University code of conduct. Further, all assignments are the responsibility of **each individual pupil**. Utilizing forums, online Q/A resources, chatGPT, teaching staff, and/or the class professor to ask questions is (of course) acceptable but copying another peer’s work is considered a violation of the University code of conduct.

You are responsible for understanding Hult University policies on academic integrity and how to use sources responsibly. Not knowing the rules, misunderstanding the rules, running out of time, submitting "the wrong draft", or being overwhelmed with multiple demands are not acceptable excuses. There are no excuses for failure to uphold academic integrity.

### Accessibility

Your professor and Hult University are committed to providing an accessible, safe, diverse aca-demic community. If necessary, contact school administration for academic, classroom or other appropriate accommodations.

### Grading:

A course grade will be assigned on the basis of student performance on two equally weighted data-driven business case studies. Each assignment is graded out of 100 points and weighted according to the below information.

Assignment due dates are set within Canvas. Assignments are accepted up to 24 hours late with a one letter grade deduction. Any work submitted afterwards will automatically be assigned an F. Pupils are expected to manage their own time and submit their work accordingly. Failure to submit submissions through the University approved portal by the assignment deadline will be considered late and not accepted. Submissions to any other location will not be accepted.

### Graduate Student Assessments

1. EDA Cereal- Data set to be provided along with detailed instructions 50% of total grade
2. EDA & Modeling Case Study- Click Through Rate Data set to be provided along with detailed instructions 50% of total grade

### Assignment Presentations

Assignment information is contained in the course repository.

Although this is an analytical course, both assignments require presentations to a fictitious non-technical business leader. It is often the case that analytical professionals must demonstrate fluency, earn trust and articulate technical material effectively. Thus, both assignments require the script(s) to produce the technical outputs and a presentation.

Assessments involve using data to apply various methods and draw out insights and conclusions, and at least one case requires machine learning in addition to insights.

Each case will have the following work artifacts:

* Voice Narration – 20pts
  + Maximum 10min voice recorded slide presentation uploaded to youtube, or a voice over in the slide file, screenshare i.e. loom.com or shared in a similarly appropriate manner.
* Presentation Slides – 20pts
  + The presentation will be organized and describe, explore data, the problem statement, prior expectations and any insights identified
* R script – 20pts
  + The execution script used for the creation of any visuals, models or insights made during the presentation. Your code will be run and evaluated. If it causes errors, is inefficient, poorly commented, or organized points will be duducted.
* Written document - 20pts
  + Describe the insights uncovered with specific data validation, externally sourced information and statistics which may bolster the data driven insights uncovered
* Data Mining Process [Not a submitted artifact] – 20pts
  + Not a submitted artifact, but an assessment of the overall logical execution, robustness of the analysis, and construction of any models (if applicable) with evaluation. Points are assigned for utilizing analytical techniques from within and outside of class topics demonstrating conceptual understanding of the business implications and having a structured analytical effort as opposed to a meandering set of methods with difficult to understand and audit outcomes.

Essentially all supporting material including scripts, documents, visuals and/or presentation slides will need to be turned in for review.

## Classes

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Date | Start Time | Hr1 | Hr2 | Hr3 | Notes |
| May 22 | 5pm | Intro & Administrative | Intro to R | R, r-studio, git |  |
| May 23 | 5pm | How does GPT work? | Prompting & Productivity | Back to R: Object Classes |  |
| May 24 | 5pm | EDA: Data Sources,  Data Manipulation | EDA: Data Visualization | More EDA |  |
| May 25 | 5pm | Loops & Logical Operations | Custom Functions | Writing Packages |  |
| May 29 | NA |  | | | Case 1 DUE |
| May 30 | 5pm | Machine Learning Data Prep | Decision Tree | Random Forest |  |
| May 31 | 5pm | R Markdown | Flexdashboard | Library(officer) |  |
| June 1 | 5pm | Responsible & Trusted AI | Equity/Inclusion Modeling | |  |
| June 6 | NA |  | | | Case 2 DUE |

## Grading Scale

You earn the grade based on assignments according to the scale below. A student’s degree, certificate candidacy, or funding status will not have any impact on a course grade. “Needing an A” for any reason is not sufficient to earn an A grade. Assignments are graded according to a 0-100 point scale with rubrics included on the course site.

|  |  |  |
| --- | --- | --- |
| Low | High | Grade |
| 90 | 100 | A |
| 80 | 89 | B |
| 70 | 79 | C |
| 60 | 69 | D |
| 0 | 59 | F |