CMU ONLINE
Carnegie Mellon University

Syllabus Overview Sample: Intro to Data Science Computing Workflows

Program: Applied Data Science

Faculty: Dr. Shannon Gallagher, Lecturer and Special Faculty, Department of Statistics & Data Science

Course Number: 36643 **Units:** 6

Zoom Class: Weekday Evening, ET

Course Description

Learn how to apply computational thinking to data processing and analysis problems through common programming languages (R). Topics include defining and manipulating vectors, lists, and data frames; processing strings; input and output data; writing functions; and working with data-and-time-based data.

Course Learning Objectives

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

- Write basic R code to load, manipulate, and summarize data.
- Use tidyverse functions to group, aggregate, and summarize data.
- Use basic R data structures such as vectors, lists, and data frames.
- Create simple reproducible reports by combining R and text in Quarto.
- Build reproducible workflows that load, transform, and output data

Faculty Introduction



Dr. Shannon Gallagher is a Lecturer and Special Faculty at the Department of Statistics & Data Science at CMU. Her research interests include statistical methods for AI, biostatistics, and reproducible workflows. Prior to this appointment, she was a Sr. Machine Learning Research Scientist at the Software Engineering Institute at CMU. She received her PhD in Statistics in

2019 and completed her post doc at the National Institute of Allergy and Infectious Diseases.

Course Topics

Topic	Learning Objectives
Introduction and R Basics	● Create R scripts and run the code they contain. ● Define variables containing simple values like numbers, strings, and vectors. ● Call simple built-in R functions and use the values they return. ● Use RStudio to write and run simple R code.
Presenting & Quarto	● Create and run code chunks in Quarto documents. ● Render Quarto documents to HTML and PDF formats. ● Insert formatted text in Quarto documents, such as headings and emphasis. ● Use R functions to format tables into human-readable form.
Data Transformation	● Filter a provided data frame to identify observations of interest to a research question. ● Use a data frame and R functions to calculate aggregates and groupings of interest. ● Join multiple data tables together to facilitate analysis. ● Arrange data in “tidy” format to facilitate easy visualization and analysis.
Workflow	● Write your data transformation/processing as an R script. ● Create RStudio projects to organize your scripts and data files. ● Use relative paths to refer to files within your projects. ● Deliver reproducible scripts that can be reused by others.
Reading & Writing	● Read data provided in various formats, correctly reading different types of columns. ● Identify problems in data files, such as missing values or incorrect column types. ● Write processed output back to files in various formats. ● Enter simple data frames by hand.
Data Types	● Comfortably manipulate vectors of numbers, strings, and factors ● Convert between data types as needed to facilitate a data analysis ● Perform arithmetic operations on numeric vectors ● Format strings and vectors of strings to produce textual output ● Rearrange factors to facilitate analysis and plotting
Programming	● Write functions to do simple transformations on vectors and data frames ● Call functions and use their outputs to perform tasks ● Use iteration to repeat tasks on many datasets, columns, or items

	● Build complex data pipelines that use functions and iteration to compose multiple tasks across many data files

---

NOTE: This is a sample designed to provide an overview of topics covered in the course based on previous course syllabi. All students will be provided an updated and final syllabus at the start of the course. Faculty, content, and schedule are all subject to change.