# DSCI 523: Programming for Data Manipulation



## Contents


- License
- Course Learning Outcomes
- Teaching Team
- Lecture Schedule
- Deliverables
- Class Schedule & office hours
- Policies


Program design and data manipulation using industry-standard software tools designed for statistical work. Organizing, filtering, sorting, grouping, reformatting, converting, and cleaning data to prepare it for further analysis. This course is not eligible for Credit/D/Fail grading.

**Course Webpage** https://pages.github.ubc.ca/MDS-2024-25/DSCI_523_r-prog_students/README.html

**Course GitHub repo** https://github.ubc.ca/MDS-2024-25/DSCI_523_r-prog_students

Skip to main content

# License



# Course Learning Outcomes

By the end of the course, students are expected to be able to:

1. Read data into R from vanilla (e.g., `.csv`) and non-standard plain text files, as well as common spreadsheet file types (e.g., `.xls`).
2. Manipulate a single data table in R, to do things such as:
    - filtering rows based on a criterion or combination of criteria;
    - selecting variables;
    - creating new variables and modifying pre-existing ones;
    - rearranging the observations or variables in a deliberate way (e.g., sorting).
3. Define tidy data and explain why it is an optimal format for data analysis involving rectangular data in R.
4. Transform data into the tidy data format in R using `tidyr`.
5. Understand the key data structures in R.
6. Compare and contrast these relationships to the relationship between `vectors` and `data.frame` objects in R. Move data fluidly between these object types.
7. Manage and manipulate data with dates and times, missing values and categorical variables in R. Rename data frame columns.

Skip to main content

9. Translate fundamental programming concepts such as loops and conditionals into R code.
10. Use the split-apply-combine approach in R to iterate over and summarize data by groups.
11. Understand how to write functions in R, document them correctly and assess if they are correct via unit testing.
12. Know when and how to abstract code (e.g., into functions) to make it more modular and robust.
13. Use a functional programming style as another means of code abstraction in R.
14. Use metaprogramming (in particular, tidy evaluation) to make use of tidyverse functions inside custom written functions in R.
15. Produce human-readable code that incorporates best practices of programming and coding style.

# Teaching Team

## Section 001

| Position | Name | Slack Handle |
|---|---|---|
| Lecture & Lab Instructor | Tiffany Timbers | @tiff |
| Teaching Assistant | Samir Damji | @Samir Damji |
| Teaching Assistant | Maria Stephenson | @Maria |
| Teaching Assistant | Ramin Rezaeianzadeh | @Ramin Rezaeianzadeh (TA) |
| Teaching Assistant | Riya Eliza Shaju | @Riya Eliza Shaju |

Skip to main content

## Section 002

| Position | Name | Slack Handle |
|---|---|---|
| Lecture & Lab Instructor | Gittu George | @gittu |
| Teaching Assistant | Meltem Omur | @Meltem Omur (TA) |
| Teaching Assistant | Tony Liang | @Tony Liang |
| Teaching Assistant | Atabak Eghbal | @Atabak |
| Teaching Assistant | Ngoc Bui | @ngoc |

# Lecture Schedule

We will be employing a lot of active learning in this course, as you learn programming best by doing! Typically there will be assigned readings & videos that should be reviewed before each lecture. During synchronous lecture meeting times, I will start with a live demonstration related to the videos you watched beforehand, and then we will work in small breakout groups on a lecture worksheet (a Jupyter notebook) that allow us to practice what was covered in the assigned readings & videos. This synchronous class will be recorded and shared afterwards.

Skip to main content

| Lecture | Topic | Required readings | Required videos | Supplementary resources |
|---|---|---|---|---|
| 1 | Reading data, single data frame manipulations & tidying data in R | Introduction to Data Science<br>• chapter 1<br>• chapter 2, sections 2.0-2.5 inclusive<br>• chapter 3, sections 3.0-3.5 inclusive | Lecture 1 videos | • Data Import Cheatsheet<br>• Data transformation cheat sheet<br>• STAT 545 (chapter 5)<br>• Relevant chapters of R for Data Science |
| 2 | Key datatypes & operators in R | Not applicable | Lecture 2 videos | • Base R cheat sheet<br>• Advanced R (chapters 2-5) |
| 3 | Working with dates, strings & factors in R | STAT 545 (Data Analysis 2 section) | Lecture 3 videos | • Dates and Times Cheatsheet<br>• Work with Strings Cheatsheet<br>• Factors with forcats Cheatsheet |
| 4 | Two table joins & base R control flow | STAT 545 (Chapter 15) | Lecture 4 videos | R for Data Science (chapter 13) |

Skip to main content

| Lecture | Topic | Required readings | Required videos | Supplementary resources |
|---|---|---|---|---|
| 5 | Tidy control flow in R | STAT 545 (section 7.8) | Lecture 5 videos | • R for Data Science (section 5.6) |
| 6 | Functions & testing in R | R for Data Science (chapter 19) | Lecture 6 videos | • Chapters 6 - 8 of Advanced R<br>• Testing chapter of R packages |
| 7 | Mapping & nested data frames in R | | Lecture 7 videos | • RStudio Apply/map functions Cheat Sheet<br>• R for Data Science (section 21.5)<br>• R for Data Science (section 25.3 - 25.5)<br>• Advanced R (chapter 9) |
| 8 | Tidy evaluation in R | Programming with dplyr | Lecture 8 videos | • RStudio Tidy Evaluation Cheat Sheet<br>• Advanced R (Metaprogramming |

Skip to main content

See the lecture learning objectives for a detailed breakdown of lecture-by-lecture learning objectives.

# Deliverables

You are responsible for the following deliverables, which will determine your course grade:

Skip to main content

| Assessment | Weight | Due Date |
|---|---|---|
| Lab 1 | 10% | 2024-09-07 18:00 PT |
| Lab 2 | 10% | 2024-09-14 18:00 PT |
| Lab 3 | 10% | 2024-09-22 11:59 PT |
| Lab 4 | 10% | 2024-09-28 18:00 PT |
| Worksheet 1 | 1% | 2024-09-07 18:00 PT |
| Worksheet 2 | 1% | 2024-09-07 18:00 PT |
| Worksheet 3 | 1% | 2024-09-14 18:00 PT |
| Worksheet 4 | 1% | 2024-09-14 18:00 PT |
| Worksheet 5 | 1% | 2024-09-22 11:59 PT |
| Worksheet 6 | 1% | 2024-09-22 11:59 PT |
| Worksheet 7 | 1% | 2024-09-28 18:00 PT |
| Worksheet 8 | 1% | 2024-09-28 18:00 PT |
| Pre-lecture quizzes | 1% | Before each lecture |
| iClicker | 1% | During each lecture |
| Quiz 1 | 25% | 2024-09-17 - 2024-09-20 |

Skip to main content

# Class Schedule & office hours

See calendar.

# Policies

Please see the general MDS policies.

Next
Lecture 1 - Introduction to R via the tidyverse ›