Module 3: R

Week plan

# Overview

R is a great way to get started with data science: a popular language, open source and free, R was designed for statistical work and is enhanced by its IDE, community, and specialised packages. The first part of this module teaches R with a focus on manipulating and visualising data. Students will get set up with a functional RStudio workflow, use different file types, transform data tables, import and manipulate data, use functions and loops, create data visualisations, make a Shiny app, and learn how to solve problems with their programming. Both base R and tidyverse methods are taught. To work reproducibly, students will create R Projects.

The second part of the module will cover the ethics of consent, Equity, Diversity & Inclusion (EDI) training, and professional skills including presentation, project management, and data security. Finally, the module will conclude with an industry case study.

# Length

5 days.

# Learning outcomes

1. Setting up and using R and RStudio.
2. Manipulating and visualising data.
3. Fixing errors.
4. Understanding consent in data-based studies.
5. Making presentations and managing projects.

# Key texts

* Alexander, 2022, *Telling Stories with Data*, CRC Press.<https://www.tellingstorieswithdata.com/>
* Alexander (eds), 2021, *DoSS Toolkit*,<https://rohanalexander.github.io/doss_toolkit_book/>.
* de Graaf, 2019, *Managing Your Data Science Projects: Learn Salesmanship, Presentation, and Maintenance of Completed Models*, Apress.
* Healy, 2018, *Data Visualization: A Practical Introduction*, Princeton University Press
* Timbers et al., 2021. *Data Science: A First Introduction.*<https://ubc-dsci.github.io/introduction-to-datascience/>
* Wickham and Grolemund, 2017, *R for Data Science*, O'Reilly.<https://r4ds.had.co.nz/>
* Wickham, 2021, *Mastering Shiny*, O'Reilly.<https://mastering-shiny.org/>
* Wiley, Matt, Wiley, Joshua F., 2020, *Advanced R 4 Data Programming and the Cloud*

# Schedule

· 01 Hello World! *(Beginner)*

*An introduction to R, working in R, R code basics, and R file types*

o Getting set-up (Alexander (eds), 2021, Chapters 2-4)

§ R

§ RStudio

o R basics (Wickham and Grolemund, 2017, Chapter 4)

o File types:

§ scripts (Wickham and Grolemund, 2017 Chapter 6)

§ RMarkdown (Wickham and Grolemund, 2017 Chapter 27)

§ Hello World! (Alexander (eds), 2021, Chapters 5, 6, or 7)

· 02 Work practices *(Beginner)*

· Errors (Alexander (eds), 2021 Chapters 13)

*Strategies for understanding and resolving errors*

o Getting help (Alexander (eds), 2021 Chapter 10)

o Using Stack Overflow (Alexander (eds), 2021 Chapters 11-12)

o Making reproducible examples (reprexes) (Alexander (eds), 2021 Chapters 14)

o Debugging (Alexander (eds), 2021 Chapter 15)

· Reproducibility

*Working in R using R projects to create work that others can reproduce*

o R projects (Wickham and Grolemund, 2017 Chapter 8)

o Coding practices (Alexander (eds), 2021 Chapter 19)

· 03 Data in R *(Beginner)*

*Understanding key data types*

o Vectors (Wickham and Grolemund, 2017 Chapter 20)

o Tibbles (Wickham and Grolemund, 2017 Chapter 10)

o Strings (Wickham and Grolemund, 2017, Chapter 14)

o Factors (Wickham and Grolemund, 2017, Chapter 15)

o Dates and times (Wickham and Grolemund, 2017, Chapter 16)

o Missing values (Wickham and Grolemund, 2017 Chapter 5)

· 04 Manipulation *(Beginner)*

*Using tidyverse to manipulate data tables to view row and column subsets, reorder, and create new columns; summarise data tables and subsets*

o Filtering (Wickham and Grolemund, 2017 Chapter 5, Timbers et al. 2021, Chapter 3.6)

o Arranging (Wickham and Grolemund, 2017 Chapter 5)

o Selecting (Wickham and Grolemund, 2017 Chapter 5, Timbers et al. 2021, Chapter 3.5)

o The pipe (Wickham and Grolemund, 2017 Chapter 5 & 18; Timbers et al. 2021, Chapter 3.8)

o Mutating (Wickham and Grolemund, 2017 Chapter 5, Timbers et al. 2021, Chapter 3.7, 3.10)

o Summarising (Wickham and Grolemund, 2017 Chapter 5, Timbers et al. 2021, Chapter 3.9)

§ Counting

§ Proportions

o Grouping (Wickham and Grolemund, 2017 Chapter 5)

o Cleaning (Alexander, 2022, Chapter 11)

· 05 Wrangling *(Intermediate)*

*Using data from saved files and databases, combining data sources, reformatting data tables, and handling larger data tables*

o Importing data (Wickham and Grolemund, 2017, Chapter 11; Timbers et al. 2021, Chapter 2) & Saving data

o Pivot (Wickham and Grolemund, 2017, Chapter 12; Timbers et al. 2021, Chapter 3.4)

§ Joining data (Wickham and Grolemund, 2017, Chapter 13)

o data.table (Wiley and Wiley, 2020, Chapter 7;<https://cran.r-project.org/web/packages/data.table/vignettes/datatable-intro.html>)

· 06 Programming *(Intermediate)*

*Writing custom functions and iterators; simulating data*

o Functions (Wickham and Grolemund, 2017, Chapter 19)

o Loops (Wickham and Grolemund, 2017, Chapter 21)

o If/else logic (Alexander (eds), 2021, Chapter 47)

o Simulation (Alexander (eds), 2021, Chapter 62 - 63)

· 07 Visualization *(Intermediate)*

*Using ggplot to build visually appealing and clear bar charts, histograms, and scatter plots*

(Wickham and Grolemund, 2017, Chapter 3; Healy, 2018, Chapter 3; Alexander, 2022, Chapter 6)

o ggplot essentials and grammar

o Bar charts and histograms

o Scatter plots

· 08 Shiny *(Advanced)*

*Creating a basic interactive web application*

(Wickham, 2021, Chapter 1)

· 09 Ethics

*Understanding the importance of consent in data collection and experiments*

o What is consent?

§ James H. Ware, 1989, ‘Investigating Therapies of Potentially Great Benefit: ECMO’, *Statistical Science*.

§ Donald A. Berry, 1989, ‘Comment: Ethics and ECMO’, *Statistical Science*.

· 10 Inequity

*Understand procedures for Equity, Diversity, and Inclusion Training*

o Undertake Equity, Diversity & Inclusion (EDI) training.

· 11 Professional skills

*Effectively communicating data work in presentations; managing data projects, including planning scopes, timelines, and deliverables; basics of information and data security*

(de Graaf, 2019)

o Presentations

o Presenting work in R

o Project management

· 12 Industry case study

# Assessment

### Formative

| Assessment Idea | Pedagogical Reasoning |
| --- | --- |
| Multiple-choice placement/self-assessment quiz at the start of the module | Assess prerequisite knowledge and preexisting, but not required, skills.  Assess confidence in ability to learn programming (affective measure).  Provide students and instructors with immediate feedback. |
| Small independent exercises, including short multiple choice questions asking students to reflect on how they did.  (embedded in the decks and included in the lesson plans) | LO 1, 2  Completing the “I do, we do, you do” strategy (following worked examples) to reinforce what has been demonstrated.  Providing immediate feedback to students and instructors (succeeded in task/did not succeed in task).  Including affective measures. |
| Problem-solving exercises  (One per submodule: Data in R, Manipulation, Wrangling, Programming, Visualisation, and Shiny)  · Given a code segment that generates an error:  o Work independently to identify the problem. Note methods and outcomes.  o In small groups, discuss what did and didn’t work. | LO 3  Deliberately including real errors to develop problem-solving skills and reinforce that errors are part of the process. |
| Small group discussion and report back activities  (One per submodule: Work practices, Ethics, Professional Skills, Inequity) | LO 4, 5  Peer learning for increased engagement |

### Summative

| Assessment Idea | Pedagogical Reasoning |
| --- | --- |
| Summative project  (Assessing content from Data in R, Manipulation, Wrangling, Programming, Visualisation, Shiny, and Professional Skills) | LO 1, 2, 3  Performance measure to evaluate practical/applied skills.  Linking concepts and practices learned in separate submodules. |
| Written reflections  (One per submodule: Work practices, Ethics, Inequity) | LO 4, 5  Measuring knowledge and affect where appropriate. |