

# Lecture 01a – This is ENVX1002

ENVX1002 Introduction to Statistical Methods

**Floris Van Ogtrop (coordinator) & Januar Harianto**

*The University of Sydney*

Feb 2025



THE UNIVERSITY OF  
**SYDNEY**

# Welcome

# About us...

## Floris van Ogtrop - Unit Coordinator

- Room 306, Level 3, Biomedical Building, Australian Technology Park, Eveleigh
- Ph: 02 8627 1024
- Email: floris.vanogtrop@sydney.edu.au

## Teaching schedule



**Januar Harianto**  
Weeks 1 – 4, Lecturer



**Floris van Ogtrop**  
Weeks 5 – 8, Unit Coordinator



**Si Yang Han**  
Weeks 9 – 12, Lecturer

# ENVX1002 Learning outcomes

- **LO1.** Implement basic **reproducible research practices** – including consistent data organisation, documented code, and version-controlled workflows so that statistical analyses and results can be readily replicated and validated by others.
- **LO2.** Demonstrate proficiency in **utilising R and Excel** to effectively explore and describe life science datasets.
- **LO3.** Apply parametric and non-parametric **statistical inference** methods to experimental and observational data using RStudio and effectively **interpret and communicate** the results in the context of the data.
- **LO4.** Be able to put into practice **both linear and non-linear models** to describe relationships between variables using RStudio and Excel, demonstrating creativity in developing models that effectively represent complex data patterns.
- **LO5.** Be able to articulate statistical and modelling results clearly and convincingly in both **written reports and oral presentations**, working effectively as an individual and **collaboratively** in a team, showcasing the ability to convey complex information to varied audiences.

# Delivery format

All lectures and tutorials are held in [ABS Lecture Theatre 1130](#). Lab sessions are held in the Biomedical Building, Australian Technology Park, Eveleigh.

- **Lectures (recorded)**: deliver content, provide context, and introduce new concepts
- **Tutorials (recorded)**: practice and apply concepts from lectures, prep for labs
- **Labs**: hands-on practice with R and data analysis, with demonstrators to help you
- **Ed discussion**: online forum for questions and discussions
- **Drop-in sessions**: additional help and support as needed on Zoom or in person

# Timetable

## Lectures

- Monday 12pm–1pm, ABS Lecture Theatre 1130
- Tuesday 9am–10am, ABS Lecture Theatre 1130

## Tutorials

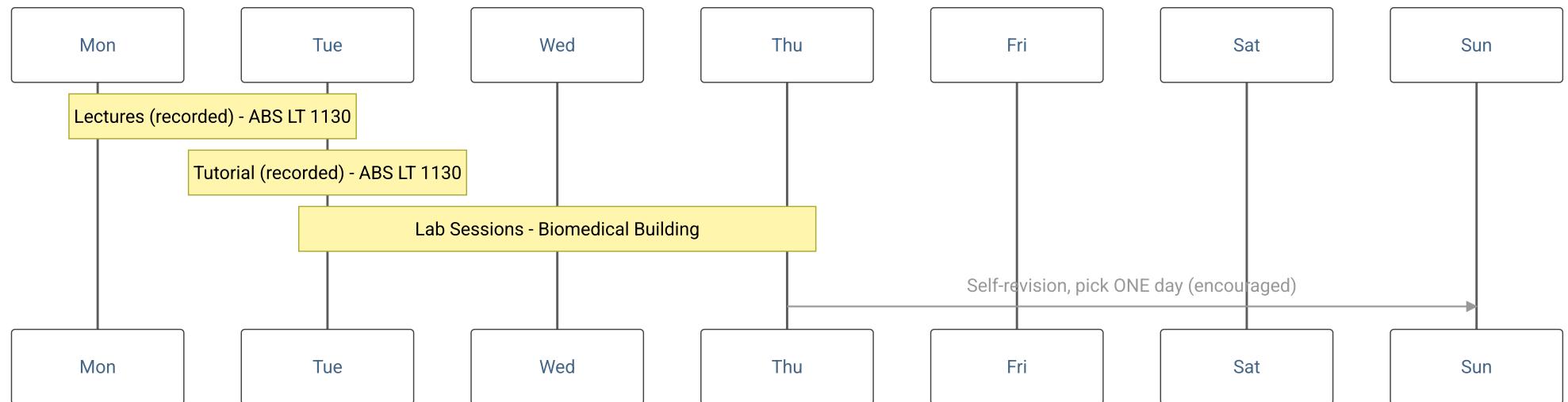
- Tuesday 10am–11am, ABS Lecture Theatre 1130
- 1-hour tutorial directly following your lecture

## Computer Labs

- 2-hour in-person lab session with demonstrators
- Biomedical Building, Australian Technology Park, Eveleigh
- See timetable for your allocated time

# Schedule at a glance...

## ► Code

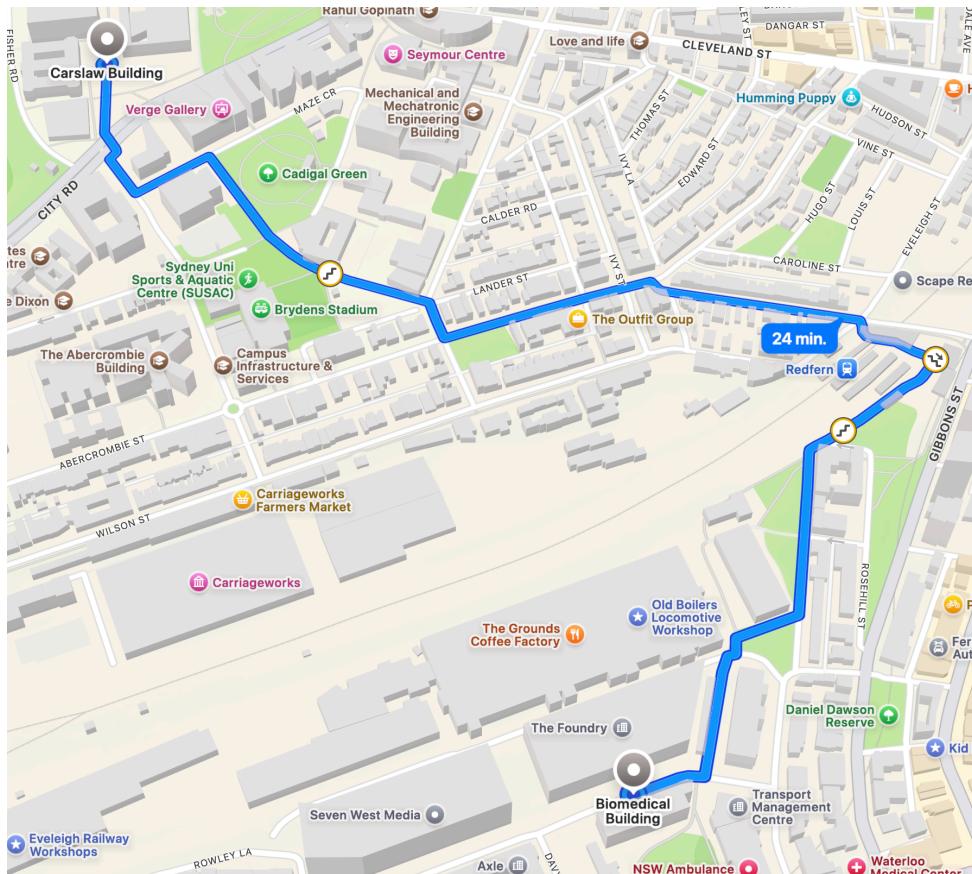


# Resources

- Canvas (of course)
- Ed Discussion – main platform for **ANNOUNCEMENTS** and Q&A
- ENVX-resources – GitHub repository for our open-source materials

# Where are the Labs?

- Lab sessions include extra time (30 minutes) for travel – already programmed in the timetable (so clashes are avoided)
- A free shuttle service is available between campus and the labs, but the schedule is very limited
- Take advantage of the **new community access gates at Redfern Station**: saves 5 minutes



# Content & assessments

# Topic outline

- Week 01 - **Data:** Reproducible science
- Week 02 - **Data:** Introduction to statistical programming
- Week 03 - **Data:** Exploring and visualising data
- Week 04 - **Data:** The Central Limit Theorem
- Week 05 - **Inference:** 1-sample tests
- Week 06 - **Inference:** 2-sample tests
- Week 07 - **Inference:** Non-parametric tests 1
- Week 08 - **Inference:** Non-parametric tests 2
- Week 09 - **Modelling:** Describing relationships
- Week 10 - **Modelling:** Linear functions
- Week 11 - **Modelling:** Linear functions – multiple predictors
- Week 12 - **Modelling:** Non-linear functions
- Week 13 - **Revision:** Past exam questions and review

# Assessments

## ► Code

The most up to date (and slightly more comprehensive) information for **2025** is [here](#). In a nutshell:

### ENVX1002 Assessments

Week	Assessment	Description
3	Early Feedback Quiz (individual 5%)	In-person - 15 minutes
5	Project 1: Exploring data (individual 15%)	Written report, 500 words
8	Coding and data skills evaluation (individual 15%)	In-person - 50 minutes
13	Project 2: Modelling (10% + Peer assessment 5%)	Group presentation - 5 minutes
Exam	Final exam (individual 50%)	MCQ + SAQ Questions - 2 hours

Any questions?

# Software and tools

## Baby steps...

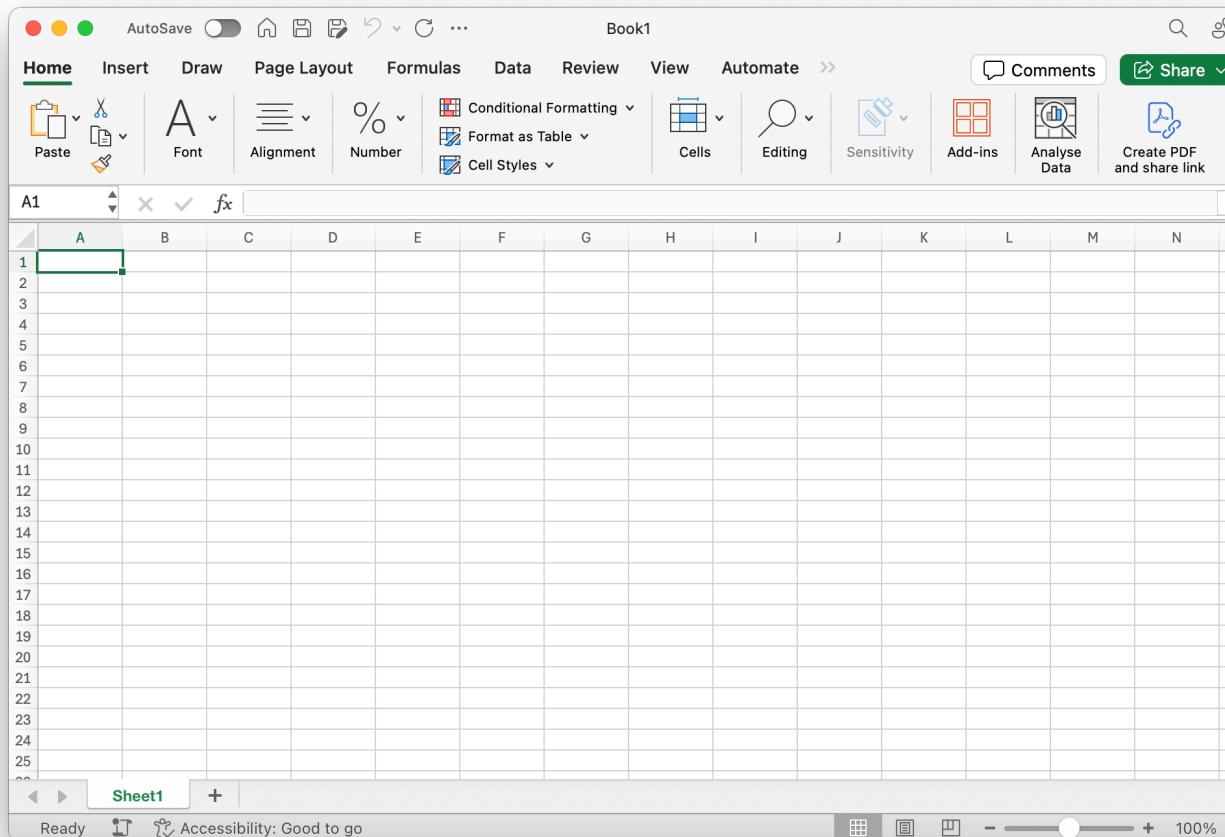
- This unit is designed for beginners - no prior statistics or programming required
- We start with basics – pace increases after **week 4**
- Focus on understanding *concepts* first, then tools
- We provide plenty of support – more on this later

# Our tech stack

1. **MS Excel** – for data entry and basic analysis
2. **R** – a programming language for data analysis
3. **RStudio** – an integrated development environment (IDE) for R
4. Quarto (**Markdown**) – a key platform for reproducible reports and documents
5. **GitHub Copilot** – AI-powered code completion tool. *Optional*

# MS Excel

- **Widely used** for data entry and basic analysis
- Versatile and user-friendly – most of you are already familiar with it
- Great data exploration tool



R



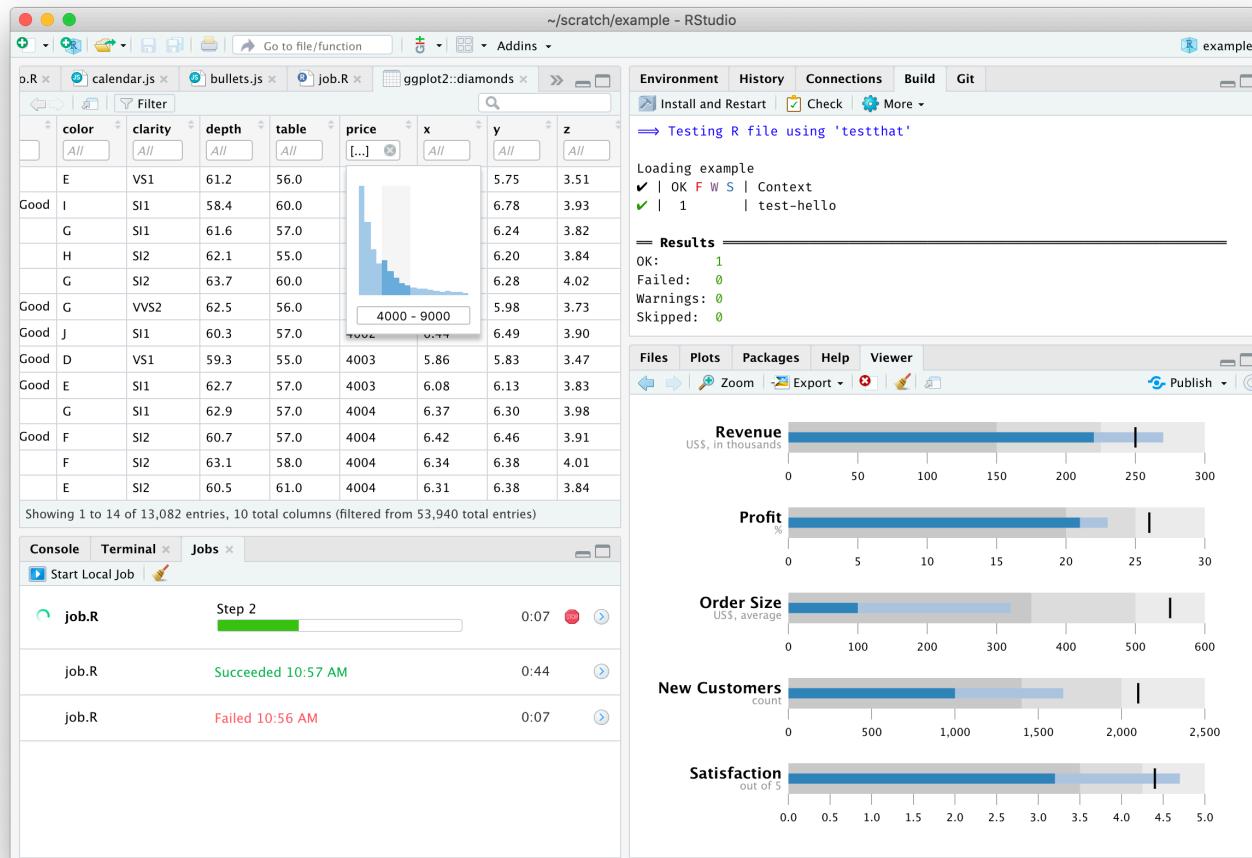
- A free, open-source programming language
- Widely used for data analysis and statistics
- **Standard tool in scientific research**
- Extensive collection of packages for data science
- Strong support for creating publication-quality graphics
- **Large, active community for help and resources**

# Why R?

1. Beginner friendly (works without needing to know a lot of programming)
  2. Makes your work reproducible
  3. Powerful yet accessible
- **Importantly** – the skills you learn are *highly* transferable to other tools and languages such as Python and Java.
  - Most easily integrated with generative AI tools – *more on this soon*
  - Well-documented and discussed online (so you can find help easily)

# RStudio

- NOT the same as R – it's an **integrated development environment (IDE)**
- Runs R (...and Python, and SQL, and more)
- Makes it easier to write and run R code by providing a *significantly* more user-friendly interface

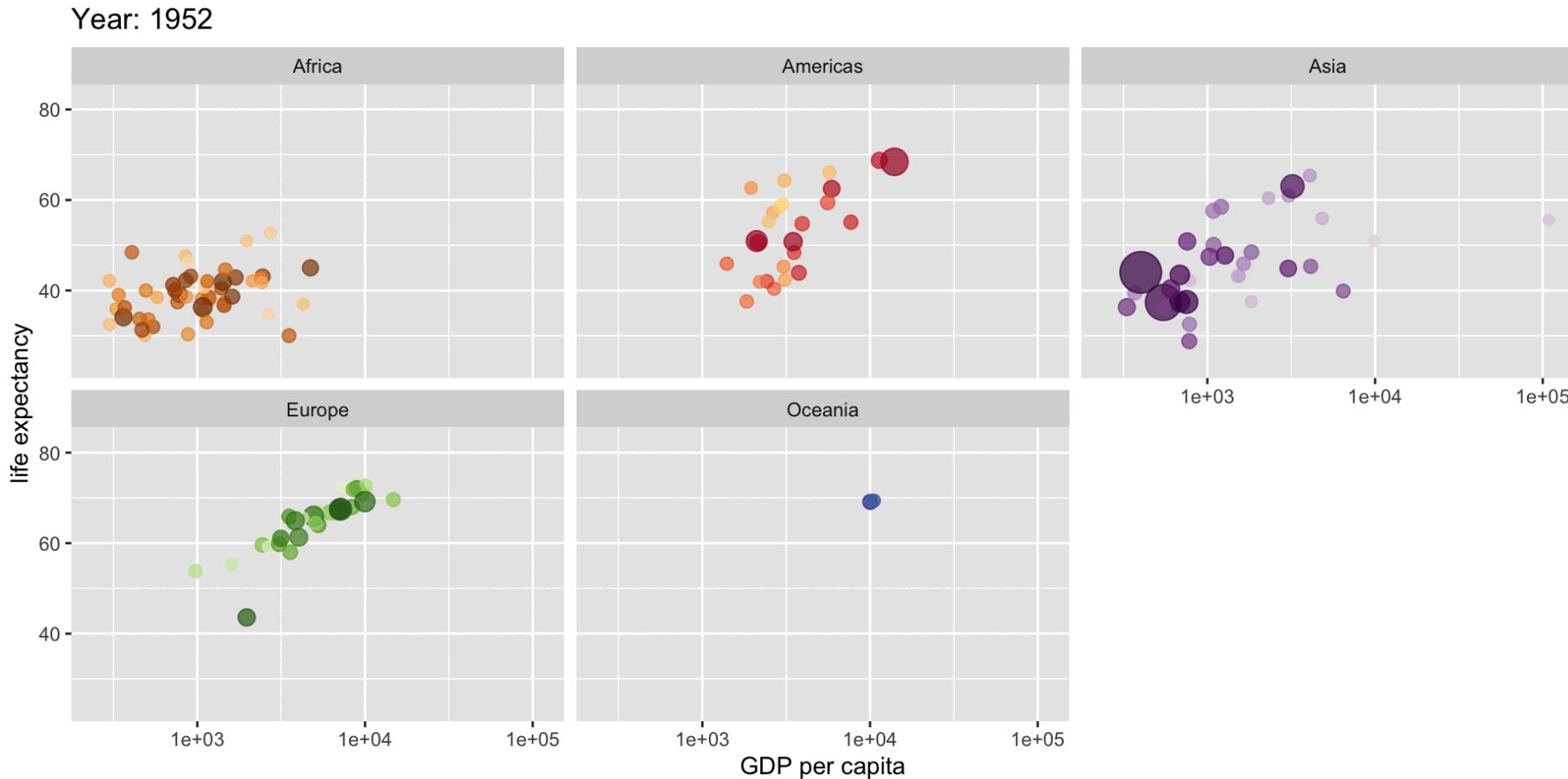


# Starting with R

- It's normal to feel overwhelmed at first
- We'll learn step by step
- Practice is key - a little bit each day helps
- Don't hesitate to ask questions!

# Satisfying when it works

► Click to see the code for this animation



Note: we will revisit this plot in the next lecture

# Quarto

- Majority of our resources are built using **Quarto** – a markdown-based document format that **you will learn to use** in this unit
  - ➡ Lecture slides
  - ➡ Tutorials
  - ➡ Lab exercises
- Quarto makes everything **reproducible** - what does it mean?
- Free and open source, available on the [ENVX resources](#) GitHub repository – re-use and modify as you wish (but follow [CC BY 4.0](#))

```
1 ## Quarto
2
3 - Majority of our resources are built using [**Quarto**](https://quarto.org/) --
4 a markdown-based document format that **you will learn to use** in this unit
5 - Lecture slides
6 - Tutorials
7 - Lab exercises
8 - Quarto makes everything **reproducible** - what does it mean?
9 - Free and open source, available on the [ENVX resources](https://github.com/ENVX-resources)
10 GitHub repository -- re-use and modify as you wish
11 (but follow [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/))
```

# R, RStudio, Quarto!?

- Again, it's normal to feel overwhelmed at first
- **These technologies are complementary – everything is implemented in RStudio**
- The **tutorials** and **labs** will guide you through the process



If you need help...

# Seek help early

- You are not alone, and **you need to learn to ask for help**
- We provide a LOT of support in various forms:
  - ➡ Face-to-face (in a group setting): tutorials, labs
  - ➡ Face-to-face (one-on-one): consultations – book a time with us (email)
  - ➡ Online (collaborative): Ed for general questions
  - ➡ Online (private): use **private posts** on Ed
- From time to time, we will organise **drop-in sessions** for additional help

NOTE: **We cannot help you if you don't ask!**

# Advice from 2024 ENVX1002 students!

- “Practicals are important.”
- “Ed discussion is your saviour (but check if someone has asked the same q before posting).”
- “The lab notes are really helpful for the projects - so turn up and pay attention.”
- “Start work on the assignments early so that if you need help, you can get it.”
- “Complete the tutorials before the prac as it makes the prac a lot more meaningful”
- “Watch the lectures.”
- “Taking step by step”
- “Keep track of the code and what it means.”
- “Do not ignore the practicals and tutorials, they teach the most.”
- “Don’t be closed minded.”
- “work with others to make notes for codes.”
- “find a stats buddy in your prac to talk things over with!”
- “Ask questions, speak up if labs are moving too fast passed concepts and ask how things are relevant to out into practice.”
- “The experience that teachers and demos have makes it far more relatable and helps cement an idea through inference.”
- “Stick with it, focus in the pracs and all is fine.”
- “Keep up with tutorial quizzes, code along with the tutorials as well as the practical, the more code you do, the more the content sticks in your head.”
- “Show up to the labs and don’t rely on copilot or chat gpt.”
- “Study after practicals to ensure you understand code.”

# Thanks!

**Tomorrow:** Lecture (1h) and then Tutorial (1h) – see you there!

This presentation is based on the [SOLES Quarto reveal.js template](#) and is licensed under a [Creative Commons Attribution 4.0 International License](#).

# References and Resources

## Core Reading

- Quinn & Keough (2002). *Experimental Design and Data Analysis for Biologists*. Cambridge University Press. Sections 1.1-1.2, pages 1-7.

# Software Resources

- CRAN (The Comprehensive R Archive Network) - R packages repository
- RStudio - Integrated development environment
- Quarto - Scientific publishing system