

Lecture 01a – This is ENVX1002

ENVX1002 Introduction to Statistical Methods

Januar Harianto

The University of Sydney

Jan 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

About 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

The following are optional (but highly recommended):

- **Drop-in sessions**: additional help and support, mostly on Zoom
- **Ed discussion**: online forum for questions and discussions

Timetable

Lectures (recorded)

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

Tutorials (recorded)

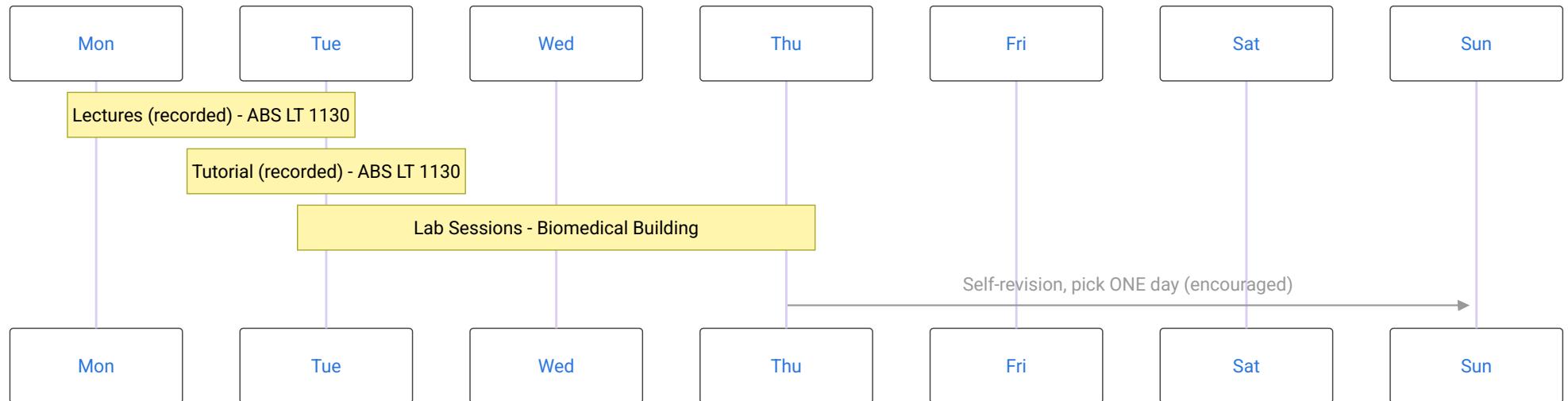
- Tuesday 10am–11am, ABS Lecture Theatre 1130
- 1-hour tutorial directly following your lecture

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)
- [ENVX-resources](#) – GitHub repository for our open-source materials
- [Ed Discussion](#) – main platform for **ANNOUNCEMENTS** and Q&A

Where are the Labs?

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:

Week	Assessment	Description
3	Early Feedback Quiz (individual 5%)	In-person - 15 minutes
5	Project 1: Exploring data (individual 10%)	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 45%)	MCQ + SAQ Questions - 2 hours

- **Week 3:** The **early feedback quiz** is a chance for us to gauge your understanding and provide feedback
- **Week 8:** Coding and data skills evaluation covers R data manipulation and analysis
- Final exam will NOT require you to write or interpret code – focus on understanding concepts and interpreting results

Software and tools

The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.

– John Tukey (1915 – 2000)

Baby steps...

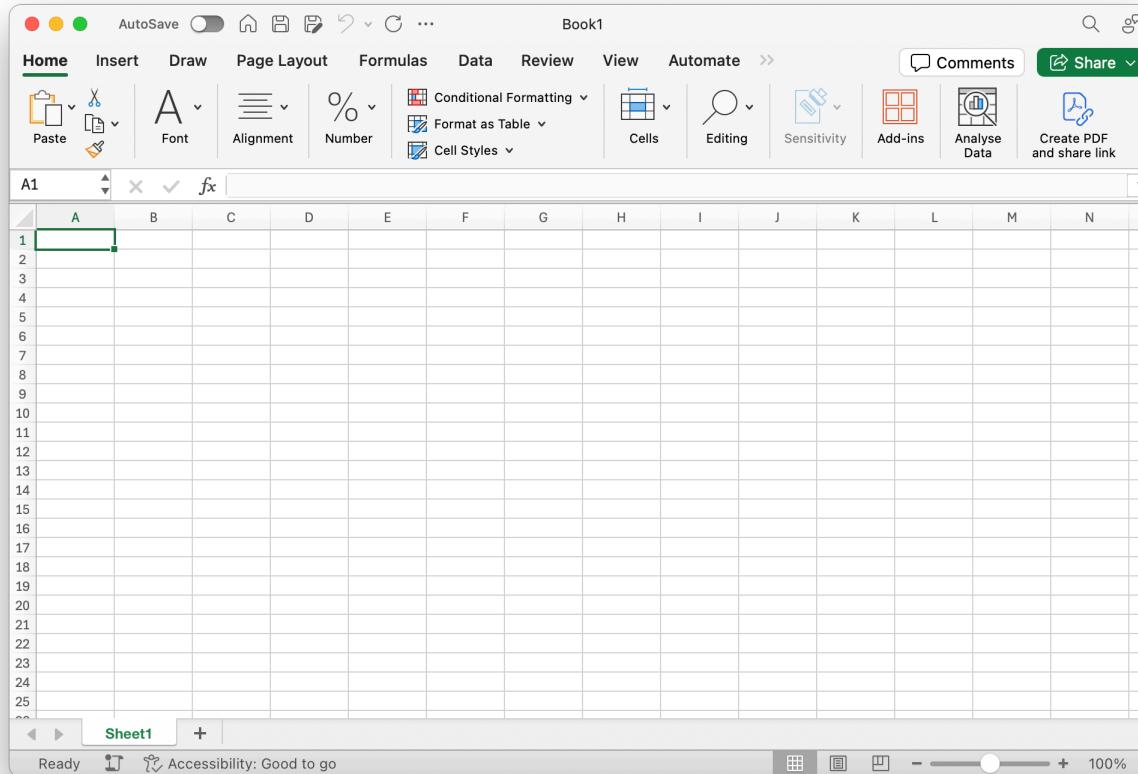
- This unit is designed for beginners - no prior statistics or programming required
- We start with basics – pace increases after **week 4**
- Focus on understanding *concepts* first, then tools
- We provide plenty of support – more on this later

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, but highly recommended*

MS Excel

- **Widely used** for data entry and basic analysis
- A standard tool in many industries, including science, often to store data
- Can be a useful complement to R for data cleaning and simple calculations
- A stepping stone to more advanced tools?



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**

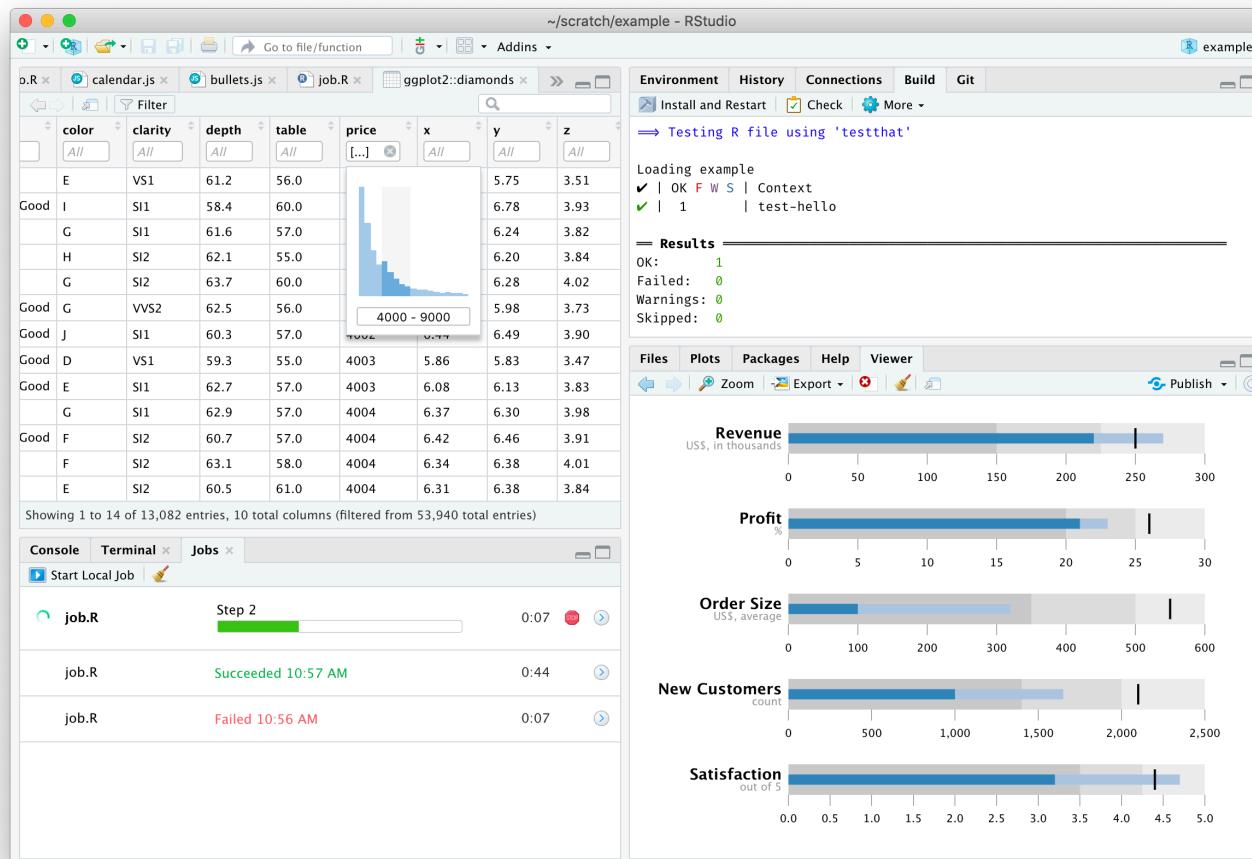
Why R?

1. Built for beginners
2. Makes your work reproducible
3. Powerful yet accessible

- **Importantly** – the skills you learn are *highly* transferable to other tools and languages.
- 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

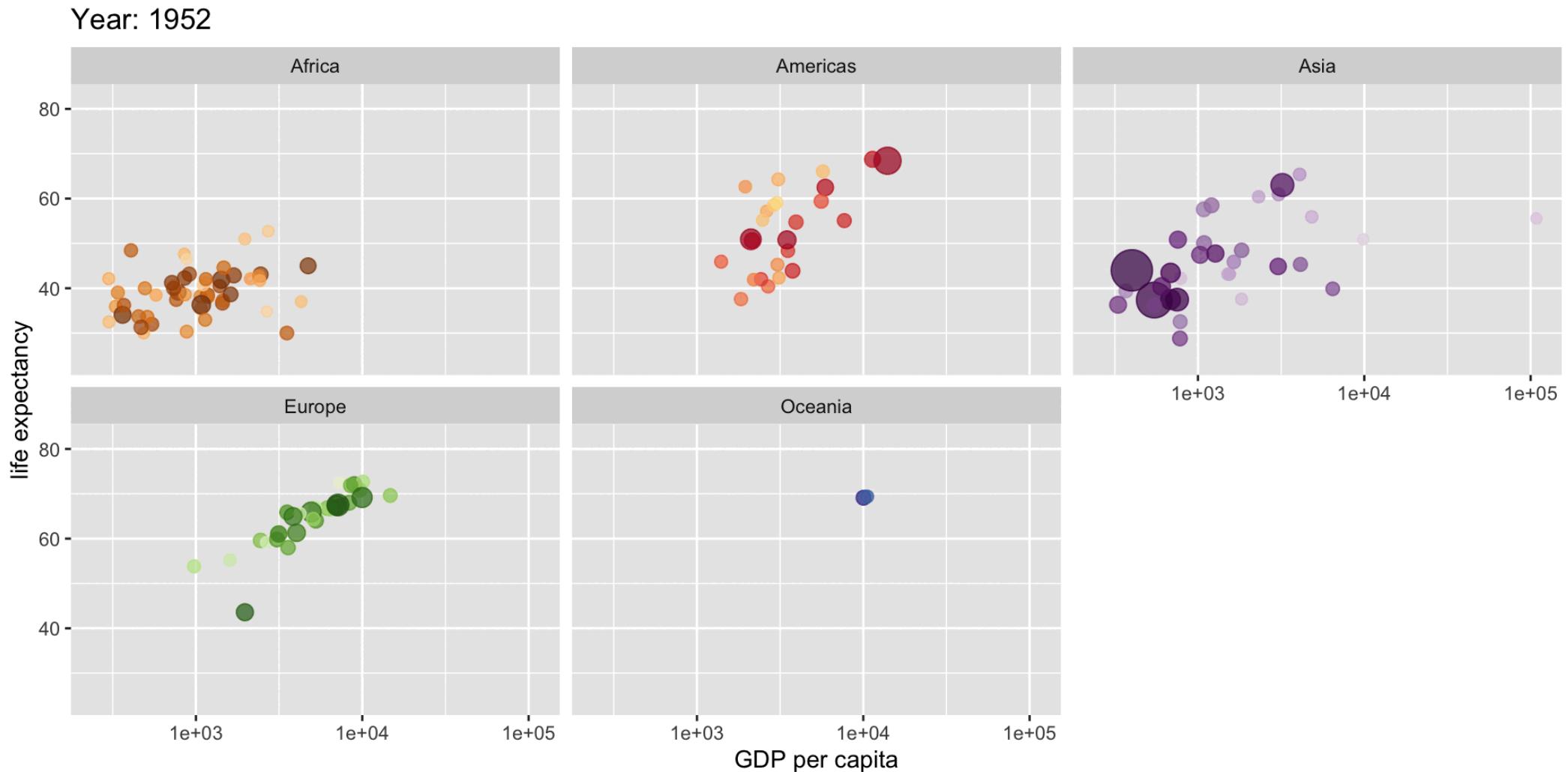


Starting with R

- 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



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 - Marjority of our resources are built using [**Quarto**](https://quarto.org/) -- a markdown-based document 1
4 - Lecture slides
5 - Tutorials
6 - Lab exercises
7 - Quarto makes everything **reproducible** - what does it mean?
8 - Free and open source, available on the [ENVX resources](https://github.com/ENVX-resources) GitHub repository
```

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



GitHub Copilot



- Slightly different from your usual AI tools
- Acts as a *pair programmer* – suggests code as you type, based on context
- **Double-edged sword – can be a crutch if you rely on it too much**
- We will start using this tool in **Week 4**

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!**

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

- Quinn & Keough (2002). Sections 1.1-1.2, pages 1-7.