

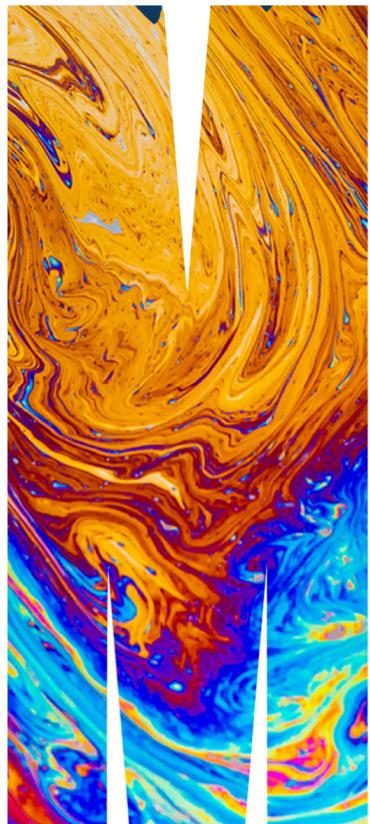
ETC5521: Exploratory Data Analysis

Getting started

Lecturer: *Di Cook*

✉ ETC5521.Clayton-x@monash.edu

CALENDAR Week 1 - Session 0



Di Cook

Professor of Statistics
Monash University

🌐 <https://dicook.org/>

✉ ETC5521.Clayton-x@monash.edu

🐦 @visnut@aus.social

About your instructor

- She has a PhD from Rutgers University, NJ, and a Bachelor of Science from University of New England
- She is a Fellow of the American Statistical Association, elected member of the R Foundation and International Statistical Institute, Editor of the Journal of Computational and Graphical Statistics, and the R Journal.
- My research is in data visualisation, statistical graphics and computing, with application to sports, ecology and bioinformatics. She likes to develop new methodology and software.
- Her students were primarily responsible for producing the tidyverse suite, R Markdown and knitr, plotly, and many other R packages we regularly use.

Jayani Lakshika

PhD student

Monash University

 <https://github.com/JayaniLakshika>

- She is from Sri Lanka.
- Her research is on models for high-dimensional data.
- Loves programming, especially in R.
- Talk to me about sports (cricket, tennis, soccer) - I will stay up all night watching!

Thomas Nguyen

MBAt 2022 graduate

Monash University

 <https://github.com/Thanh-8213>

- He is from Vietnam.
- Tutored ETC5512 in S1 2023. Also tutoring ETC1010 and ETC2420 in S2 2023.
- Talk to me about magic, cooking, board games and badminton!

3/15

Got a question or a comment!

 You can ask, unmute yourself in the chat or raise your hand in the lecture room.

 If watching later, please use the moodle (ED) forum.

4/15

Welcome to ETC5521 Exploratory data analysis!

Before modelling and predicting, data should first be explored to uncover the patterns and structures that exist. Exploratory data analysis involves both numerical and visual techniques designed to reveal interesting information that may be hidden in the data. However, an analyst must be cautious not to over-interpret apparent patterns, and to properly assess the results of a data exploration.

5/15

Unit Structure

- 2 hour lectures  Thu 08.00AM-10.00AM
- 1.5 hour face-to-face tutorial  Thu 15:00-16.30PM CL_Anc-19.LTB_132
- 1.5 hour zoom tutorial  Thu 16.30-18:00PM CL_Anc-19.LTB_132

Attendances are recorded.

6/15

ETC5521 Learning Objectives



1. learn to use modern data exploration tools with real data to uncover interesting structure and unusual observations
2. understand how to map out appropriate analyses, and to define what we would expect to see in the data
3. be able to compute null samples in order to test apparent patterns, and to interpret the results of visual inference

7/15

Resources

⌂ **Course homepage:** this is where you find the course materials (lecture slides, tutorials and tutorial solutions)

<https://eda.numbat.space/>

✉ **Moodle:** this is where you find discussion forum, zoom links, assignments and marks

<https://lms.monash.edu/course/view.php?id=141370>

8/15

Assessment Part 1/2

Active engagement is expected but not assessed

- Lecture and tutorial participation will contribute to this.
- If you have some clashes for lecture or tutorial, there are other opportunities to show your engagement, including:
 - posting in the Moodle/ED discussion forum,
 - github commits (more on this later in the course),
 - peer reviewing (more on this later in the course), and
 - online submissions (see Tutorial Q3 for example).
- There should be **at least 3 measurable and meaningful engagement activities each week**. Think about posting a question or a comment in the discussion forum each week!

9/15

Assessment Part 2/2

Weekly quizzes 10%

There will be a weekly quiz provided through Moodle. These are a great chance to check your knowledge, and help you prepare for the tutorial and to keep up to date with the weekly course material. Your best 10 scores will be used for your final quiz total.

Assignment 1, 2 (25% each): through GitHub classroom

Due dates are in Moodle and on <https://eda.numbat.space/index.html>. These are individual.

Assignment 3, parts 1 and 2 (20% each): through GitHub classroom

Due dates are in Moodle and on <https://eda.numbat.space/index.html>. We'll be working on a buddy system. For each part you will be randomly grouped with another student, and you will work together to provide a solution. Different partner each time. The work needs to be conducted using Github, so that your efforts on the team work will be assessed from your commits.

10/15

Expectations Part 1/2

- Lectures are recorded but you are expected to have either attended the lecture, or watched the recordings fully, prior to the tutorial for the week.
- Tutorials attendance is expected. You should send an email ETC5521.Clayton-x@monash.edu if you expect to be absent in any week.
- Questions related to the course should be raised at Moodle/ED discussion forum.
- For personal or private administrative issues, the email contact is: ETC5521.Clayton-x@monash.edu
- If you miss a lecture/tutorial, it is your responsibility to catch up with missed material, learn about due dates for material to be turned in, and getting assigned to a group for team work, as necessary.
- All times are given in AEST (Melbourne time).

11/15

Expectations Part 2/2

- **ETC5510 is prerequisite.** Some of you may have exemption (due to evidence of equivalent knowledge to ETC5510).
- It is expected that you have already basic skills at data wrangling, visualisation, simple modeling, and comfortable with reproducible reporting using Rmarkdown.
- The priority of the ETC5521 teaching team is to support you in ETC5521 material.
- It's essential in this course that you have a [GitHub](#) account. It is free to register.

12/15

GitHub Classroom

We are going to use GitHub Classroom (etc5521 2023: Exploratory Data Analysis) to distribute assignment templates and keep track of your assignment progress.

1. Clone the test assignment by clicking on the link given in Moodle in the first tutorial.
2. Once you have accepted it (note: some browsers do not work well with GitHub Classroom (so use Chrome or Firefox)), you can find your repo here:
<https://classroom.github.com/classrooms/140034520-etc5521-2023-exploratory-data-analysis>
3. If you don't recall how to connect this repo locally, check [this guide](#).
4. In the first tutorial you will be expected to upload a file to your repo. We will check to make sure you have demonstrated ability to use GitHub.

13/15

Questions?

14/15



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).

Lecturer: *Di Cook*

✉ ETC5521.Clayton-x@monash.edu

🗓 Week 1 - Session 0

