

# ETC5521: Exploratory Data Analysis

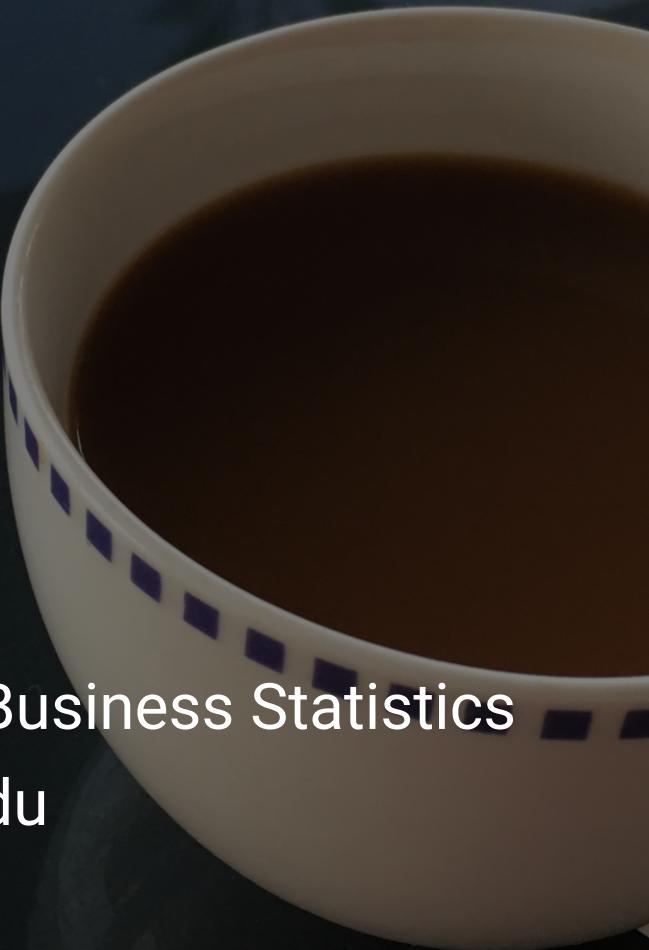
## Getting started

Lecturer: *Di Cook*

Department of Econometrics and Business Statistics

✉ ETC5521.Clayton-x@monash.edu

Week 1 - Session 0



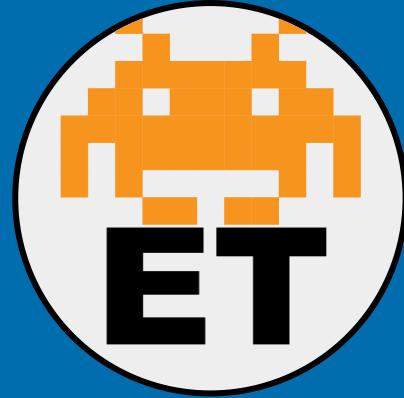


**Di Cook**  
*Professor of Business  
Analytics*  
Monash University

🌐 <https://dicook.org/>  
✉️ ETC5521.Clayton-  
x@monash.edu  
🐦 @visnut

# About your instructor

- 🚀 I moved to Melbourne in Jul 2015 to start my job at Monash
- 🚀 I have a PhD from Rutgers University, NJ, and a Bachelor of Science from University of New England, Armidale 🎓
- 🚀 I am a Fellow of the American Statistical Association, elected member of the International Statistical Institute, and an elected member of the R Foundation. 📈
- 🚀 My main research is in data visualisation, statistical graphics and computing, with application to sports, ecology and bioinformatics. I like to develop new methodology and software. 📊



**Emi Tanaka**  
*Lecturer in Statistics*  
Monash University

🌐 <https://emitanaka.org/>  
✉️ ETC5521.Clayton-x@monash.edu  
🐦 @statsgen

# About your instructor

- 👉 I moved to Melbourne in Jan 2020 to start my job at Monash
- 👉 I teach R at university courses as well as use it daily for research and hobby 🤖
- 👉 If you ever wonder who answers those stackoverflow questions, I'm one of those magic fairies 💫



- 👉 My main statistical application is agriculture and bioinformatics; these days I do more software



**Sayani Gupta**  
*PhD student*  
Monash University

🌐 <https://sayani.netlify.app>  
🐦 @SayaniGupta07

# About your tutor

- ➲ My research focuses on building visualization and statistical tools for analyzing spatially distributed big time series data.
- ➲ I'm passionate about putting to use statistics to solve real world problems.
- ➲ I spend my time learning Indian classical music, and exploring cafes near me.



**Sherry Zhang**  
*PhD student*  
Monash University



<https://huizezhangsh.netlify.app>



@huizezhangsh

# About your tutor

- 🚀 I am a first year PhD student in Monash University
- 🚀 I'm working on the diagnostics of projection pursuit optimisation
- 🚀 I like programming using R

# Introduce yourself in the chat!

- 🚀 What was your undergraduate degree in?
- 🚀 What's your profession?
- 🚀 Where do you live?
- 🚀 What's your hobby?
- 🚀 What's one thing you want everyone to know (about you)?

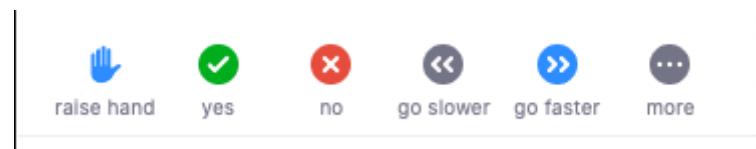


# Help me!

You can write in the chat.

You can ask (unmute yourself).

You can raise hand from zoom under Participants.





# 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.

🚀 2 hour lectures 🚧 Mon 4.00-6.00PM

🚀 1.5 hour tutorial 🔧 Mon 6.00-7.30PM

\*Attendances are recorded.\*



# ETC5521 Learning Objectives

*i*

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



# 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=81234>**

🔔 **Flux for quizzes.** Sign in with your monash email address, to be sure your marks are transferred correctly.

**<https://flux.qa/J974UJ>**

# Engagement (total 10%)

- 👉 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 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** (contributing to 1% of your final mark capped to 10% over the whole semester; in other words, you can miss 2 weeks, or 4 weeks counting mid-semester break although harder to gain engagement marks over that period, without impacting your mark).

# **Weekly quizzes 10%**

There will be a weekly quiz provided through flux, on reading or exercises conducted during the tutorial. These are a great chance to check your knowledge, and help you keep up with the weekly course material.

## **Assignment 1, 2, 3 (20% each): through GitHub classroom**

Weeks 4, 7, 12 are the due dates for these. We'll be working on a buddy system. For each assignment 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 clear.

# Final presentation 20%

In other classes, e.g. ETC5510 you have had the chance to build your EDA skills. This project is designed to help you critically assess a data analysis. The task will involve examining a given data analysis, and critically and positively examining the parts, and providing strategies for improving on it.



# 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 may not be recorded, and attendance is expected.
- ☞ Questions related to the course should be raised at moodle discussion forum.
- ☞ For personal or private administrative issues, the email contact is:  
ETC5521.Clayton-x@monash.edu
- ☞ If you register after the start of the semester or 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).



# Expectations Part 2/2

- ☛ **ETC5513 is prerequisite.** Some of you may have exemption (due to evidence of equivalent knowledge to ETC5513) or gone through the bridging course instead.
- ☛ It's expected that you know how to use **git**, **github** and basic **R Markdown** (as taught in ETC5513).
- ☛ 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.

# GitHub Classroom

We are going to use GitHub Classroom to keep track of your repos and distribute templates.

1. Start with the test assignment by clicking on the link given in [Moodle](#) **before Fri 7 Aug 11PM** and make sure you identify yourself by connecting your name to the roster.
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://github.com/etc5521-2020>
3. If you don't recall how to connect this repo locally, check [this guide](#).
4. Make some small changes to your repo and make sure these changes are pushed to GitHub.

# That's it, for this lecture!



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

Lecturer: Di Cook

Department of Econometrics and Business Statistics

✉ ETC5521.Clayton-x@monash.edu