



# Data Science and Statistics in Research: unlocking the power of your data

## Session 1.1:

### Using data and statistics in your research

# OUTLINE

Presenters

Aims of the course

Course website

Timetable

# Presenters

# AIMS OF THE COURSE

- ▶ Gavin Shaddick
- ▶ Daniel Simpson
- ▶ Matthew Thomas
- ▶ Aoibheann Brady
- ▶ Robbie Peck
- ▶ Adwaye Rambojun

# Aims of the course

# AIMS OF THE COURSE

- ▶ The ability to extract information from data is increasingly important in all areas of research and is an essential part of evidence based decision making in the world of big data.
- ▶ The aim of the course is provide an interactive experience for researchers from a variety of disciplines to enable them to use data science and statistical techniques to answer questions in their own research.

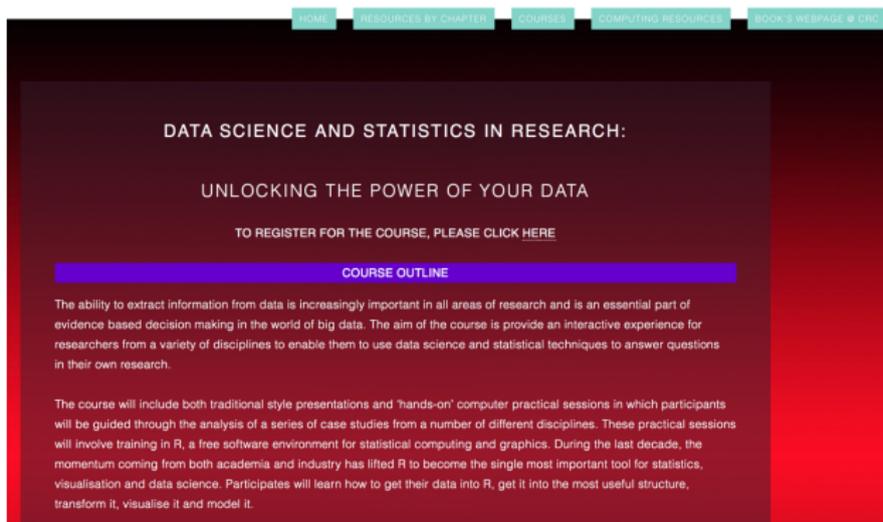
# AIMS OF THE COURSE

- ▶ The course will include both traditional style presentations and 'hands-on' computer practical sessions in which participants will be guided through the analysis of a series of case studies from a number of different disciplines.
- ▶ These practical sessions will involve training in R, a free software environment for statistical computing and graphics.
- ▶ During the last decade, the momentum coming from both academia and industry has lifted R to become the single most important tool for statistics, visualisation and data science.
- ▶ Participants will learn how to get their data into R, get it into the most useful structure, transform it, visualise it and model it.

# Course website

# COURSE WEBSITE

- ▶ [https://www.stat.ubc.ca/~gavin/STEPIDBookNewStyle/course\\_num.html](https://www.stat.ubc.ca/~gavin/STEPIDBookNewStyle/course_num.html)



The screenshot shows a website with a dark red background. At the top, there is a navigation bar with five teal-colored buttons: "HOME", "RESOURCES BY CHAPTER", "COURSE", "COMPUTING RESOURCES", and "BOOK'S WEBSITE @ CRC". The main content area features the following text:

**DATA SCIENCE AND STATISTICS IN RESEARCH:**

**UNLOCKING THE POWER OF YOUR DATA**

TO REGISTER FOR THE COURSE, PLEASE [CLICK HERE](#)

**COURSE OUTLINE**

The ability to extract information from data is increasingly important in all areas of research and is an essential part of evidence based decision making in the world of big data. The aim of the course is provide an interactive experience for researchers from a variety of disciplines to enable them to use data science and statistical techniques to answer questions in their own research.

The course will include both traditional style presentations and 'hands-on' computer practical sessions in which participants will be guided through the analysis of a series of case studies from a number of different disciplines. These practical sessions will involve training in R, a free software environment for statistical computing and graphics. During the last decade, the momentum coming from both academia and industry has lifted R to become the single most important tool for statistics, visualisation and data science. Participants will learn how to get their data into R, get it into the most useful structure, transform it, visualise it and model it.

# Timetable

# TIMETABLE

## Tuesday, November 22

- ▶ 09:30–09:45 Using data and statistics in your research
- ▶ 09:45–10:15 An introduction to R
- ▶ 10:15–10:30 Coffee
- ▶ 10:30–12:00 Using R with your data
- ▶ 12:00–13:00 Lunch
- ▶ 13:00–13:45 Data and variables
- ▶ 13:45–14:45 Initial data analysis
- ▶ 14:45–15:15 Coffee Break
- ▶ 15:15–16:00 Visualising data
- ▶ 16:00–17:00 Is your data ready to analyse?

# TIMETABLE

## Wednesday, November 23

- ▶ 09:30 - 10:00 Discussion and recap from previous day
- ▶ 10:00 - 10:15 Coffee
- ▶ 10:15 - 11:00 Hypothesis testing I
- ▶ 11:00 - 12:00 How significant are your results?
- ▶ 12:00 - 13:00 Lunch
- ▶ 13:00 - 13:45 Hypothesis testing II
- ▶ 13:45 - 14:45 Choosing the right tests for your data
- ▶ 14:45 - 15:15 Coffee
- ▶ 15:15 - 16:00 Non-parametric statistics
- ▶ 16:00 - 17:00 When and how to use non-parametric tests

# TIMETABLE

## Thursday, November 24

- ▶ 09:30 - 10:00 Discussion and recap from previous day
- ▶ 10:00 - 10:15 Coffee
- ▶ 10:15 - 11:00 Linear regression
- ▶ 11:00 - 12:00 Fitting a model to your data
- ▶ 12:00 - 13:00 Lunch
- ▶ 13:00 - 14:00 Clustering
- ▶ 14:00 - 15:00 Clustering time series / Calculating health indices

# Any Questions?