Australian National University

STAT3011
Graphical Data Analysis

Course Description
This course introduces the principles of data representation, summarisation and presentation with particular emphasis on the use of graphics. The course will use the R Statistical Programming Language in a modern computing environment. Topics to be discussed include: Data representation; examples of good and bad graphics; principles of graphic construction; some pitfalls to be avoided; presentation graphics. Graphics environments; interactive graphics; windows; linked windows; graphics objects. Statistical graphics; stem and leaf plots, box plots, histograms; smoothing histograms; quantile-quantile plots; representing multivariate data; scatterplots; clustering; stars and faces; dynamic graphics including data rotation and brushing. Relationships between variables; smoothing scatterplots; simple regression; modelling and diagnostic plots; exploring surfaces; contour plots and perspective plots; multiple regression; relationships in time and space; time series modelling and diagnostic plots.

Mode of Delivery	On campus
Prerequisites	STAT1008 Quantitative Research Methods or STAT1003 Statistical Techniques or STAT2001 Introductory Mathematical Statistics
Incompatible Courses	None
Co-taught Courses	STAT4026, STAT7026 Note: graduate students attend joint classes with undergraduates but are assessed separately
Course Convener	Professor Michael Martin
Office Location:	Rm 4.01, CBE Building 26C; Rm 1.31 Chancelry 10T1
Phone:	6125 4852
Email:	Michael.Martin@anu.edu.au
Consultation hours:	Flexible, by appointment - send e-mail to arrange
Bio and research interests	Michael Martin is Professor of Statistics. He has a PhD in Statistics from ANU, and has been lecturing at ANU since 1994. His research interests include bootstrap and resampling methods and applications of statistics to problems from a wide range of disciplines. He is also interested in Statistical Education, and higher education in general, and is Principal Fellow of the Higher Education Academy (PFHEA).
Administrators	Adriana Longhitano / adriana.longhitano@anu.edu.au 02 6125 6189 Office: Level 4 CBE Bldg 26C

SEMESTER 2
2018

http://wattle.anu.edu.au

COURSE OVERVIEW

Course Learning Outcomes

On successful completion of this course students should have an understanding of and be able to understand and apply the principles of data representation, summarisation and presentation with particular emphasis on the use of graphics. To achieve an understanding of and facility in:

1. Working knowledge of the R statistical computing language, particularly graphical capabilities
2. Represent and compare distributions of data
3. Summarise and analyse relationships between a response variable and a covariate
4. Summarise and analyse relationships between a response variable and several covariate
5. Summarise and analyse time- dependent data using basic time series models
6. Effectively communicate statistical analyses graphically, numerically and in written reports

The learning outcomes are open-ended in the sense that students are expected to develop the skills to analyse data without necessarily being told beforehand what tools of analysis will be needed during the analysis.

Assessment Summary

Details about assessment may change during the first two weeks of semester. Please ensure that you check whether there have been changes with your lecturer. Changes to the assessment requirements will be posted on the course Wattle site. Assignments and specific instructions are available from Wattle on the dates specified.

There is no midsemester or final examination. As a result, it is imperative that you work hard on your assignments and that you do not leave starting your project too late. **It is also important that you discuss your work on your project with no-one, not even the family pet (dogs can be surprisingly helpful if you are nice to them; cats less so)!!**

Due dates for assessment given here are provisional at this stage, but are expected to be close to the final due dates. Firm due dates will be advised at the top of the assignments when they are made available, and due dates stated on assignments take precedence over those advertised here.

Assessment item	Value (%)	Due date	Est. Date for return of assessment	Description and detail of the assignment	Linked Learning Outcomes
Assignment 1	20%	22 August 2018	17 Sept 2018 (after break)	Representing and comparing distributions	1, 2
Assignment 2	20%	17 October 2018	5 November 2018	Time series and dependent data	1, 5
Project	60%	23 October 2018	N/A	Graphical Data Analysis – open-ended	1-6

Research-Led Teaching

Statistics is a discipline informs many other disciplines - basically, any discipline that generates or uses data (of almost any kind) can benefit from Statistics. This makes Statistics very naturally a companion to research-led teaching, and data visualisation is a key element of the way in which Statistics can allow researchers to see structure in high-dimensional data. In this course, we will look at a number of real data sets that address real research problems, and we will see (literally) how statistics can lead to a deeper understanding of the structures underlying the data.

Continuous Improvement

We use feedback from students, professional bodies and staff to make regular improvements to the course. In response to this feedback, design improvements from the previous version of the course include:

- Maintaining the assessment structure with no final exam. This is appropriate for an applied course like this, but it also allows a better assessment spread for students during the semester
- More time spent in classes on discussions of data visualisations

Feedback

Students will be given feedback in the following forms in this course:

- Individual feedback on assignments will be posted to Wattle within the Gradebook. Only you will be able to see feedback on your assignment.
- Summary feedback reflecting patterns in data analyses will be posted to Wattle. This feedback will not identify individuals, but will rather be broad in scope and describe general patterns of response to the analyses of the data.

Technology, Software, Equipment

R (free to download)

The course makes extensive use of the free R software for statistical computing. PC/Mac labs are located at many places around campus: for an exhaustive list, visit https://services.anu.edu.au/information-technology/software-systems

Students should be able to use their student cards to access these computing laboratories and should have a computer account automatically set up for them upon registration for this course. If you have not registered for the course, your card will not allow you access to the lab. To get started with the computing requirements for the course, students should make sure that they read the document "Introductory R Worksheet 1: PC familiarity" (linked to from the R workshop page, a link to which appears on the class home page). This document describes how you can log in to the PC's. This document will also tell you how you can obtain or view other documents that you will find useful to learn about the computing setup for the course.

After this initial handout, all handouts will be available only through the class web page at

http://wattle.anu.edu.au

From the web page, you will be able to print out all of the lecture notes, all of the computer code used in the class, and all the tutorials and solutions. **No handouts will be made available except on the class web page.** A number of data sets will be analysed during lectures, live using R as much as possible. To assist you in understanding the data analyses, the R code used to produce displays discussed in class will be made available to you on the class web page. You are free to use and modify this code in conducting your own analyses.

Prescribed texts/class materials

Printed Lecture Notes (available on Wattle). Class materials, including detailed lecture notes, class lecture demonstrations, tutorials, assignments and other relevant materials, will be made available on the class web page hosted on Wattle at http://wattle.anu.edu.au. To log on to Wattle, you need to have an ANU ID (your student number) and a password (the same as for obtaining your e-mail). In order to access the class web page within Wattle, you will need to be formally enrolled in the course or you will need to have arranged access with me (e.g. if you are an Honours student). The class web page will be updated with new information on a regular basis, and will also contain links to other places of interest (such as an R workshop initially). It is essential that you visit the class web page regularly.

Reference materials

STAT3011 makes extensive use of the R statistical programming language. Learning R can be a daunting task, but there are numerous easy-to-read guides to R on the web. Also, you can use own-paced, interactive lessons within R using the swirl package (http://www.swirlstats.com)

Requisites

STAT1008 Quantitative Research Methods or STAT1003 Statistical Techniques or STAT2001 Introductory Mathematical Statistics. Students should have had a first course in Statistics that covers introductory statistics, hypothesis testing and some regression.

Co-teaching

STAT3011, STAT4026 and STAT7026 share lecture and tutorial material. The two courses have similar content and **but the assessment will differ in that in Part A of the project, STAT4026 and STAT7026 students must hand in and comment on FIVE (5) graphics rather than three**. The three cohorts will be treated separately in grading and scaling.

Student Feedback

ANU is committed to the demonstration of educational excellence and regularly seeks feedback from students. One of the key formal ways students have to provide feedback is through Student Experience of Learning Support (SELS) surveys. The feedback given in these surveys is anonymous and provides the Colleges, University Education Committee and Academic Board with opportunities to recognise excellent teaching, and opportunities for improvement.

For more information on student surveys at ANU and reports on the feedback provided on ANU courses, go to

http://unistats.anu.edu.au/surveys/selt/students/ and
http://unistats.anu.edu.au/surveys/selt/results/learning/

Policies

ANU has educational policies, procedures and guidelines, which are designed to ensure that staff and students are aware of the University's academic standards, and implement them. You can find the University's education policies and an explanatory glossary at: http://policies.anu.edu.au/

Students are expected to have read the Academic Misconduct Rules 2014 before the commencement of their course.

Other key policies include:

- Student Assessment (Coursework)
- Student Surveys and Evaluations

COURSE SCHEDULE

Week Number	Theme / Topic / Module	Activity	Required student preparation (Chapter from Lecture Notes)	Assessment deadlines and Tutorials (not assessed)
Week 1	**Introduction and getting to know R**	**Lectures**	**Chapter 1, R Workshop**	**Tutorial 1**
Week 2	**R, Graphics in R**	**Lectures**	**Chapter 1**	
Week 3	**Representing and comparing distributions**	**Lectures**	**Chapter 2**	**Tutorial 2**
Week 4	**Representing and comparing distributions**	**Lectures**	**Chapter 2**	**Tutorial 3**
Week 5	**Relationships between 2 variables**	**Lectures**	**Chapter 3**	**Tutorial 4, Assignment 1 due**
Week 6	**Relationships between 2 variables**	**Lectures**	**Chapter 3**	
Week 7	**Relationships between 3 and more variables**	**Lectures**	**Chapter 4**	
Week 8	**Relationships between 3 and more variables**	**Lectures**	**Chapter 4**	**Tutorial 5**
Week 9	**Relationships between 3 and more variables**	**Lectures**	**Chapter 4**	
Week 10	**Relationships between 3 and more variables/Time dependent data**	**Lectures**	**Chapter 4/ Chapter 5**	**Tutorial 6**
Week 11	**Time Series and Dependent data**	**Lectures**	**Chapter 5**	**Assignment 2 due, Tutorial 7**
Week 12	**Graphical Construction**	**Lectures**	**Chapter 6**	**Tutorial 8** **Project due**

ASSESSMENT REQUIREMENTS

The assessment for STAT3011 will have two components, apportioned as follows: 40% of your grade will come from two assignments spaced through the semester; and 60% of your grade will come from a project (details below) due on the last day of classes. There is no formal written exam, and no practical exam.

Two Rules:

I. Collaboration on assessable work is **forbidden**. Collaboration means talking (or any other form of communication) with someone else about the assignments or project. It is also forbidden for you to copy work from any other source. The idea is pretty simple: do the assignment by yourself. All the work you hand in for

assessment must come from you alone. Anybody caught violating this rule will receive a zero score. You have been warned!

II. Rule I again. I **really** mean it.

Because there is no formal examination in this course, it is particularly important that you do not collaborate on the assignments and projects. You are not permitted to discuss the work on the project with anybody else!! You may ask me questions about the assignments and the projects, but I will only answer questions that I feel are appropriate (that is, I will not answer questions that show you how to proceed - think of the assessment as you would an exam: I can clarify certain things, but I am unable to help you substantively). By the way, if you do collaborate on assignments and projects it is usually a lot more obvious than you think it is... and it is better to get something slightly wrong and lose a few points than to copy something completely right, learn nothing, and have me give you no points because I caught you cheating. By the way, it's bad karma as well...

THE PROJECT

The project is a vital part of the course. The purpose of the project is to (1) encourage you to be more aware and to examine graphics more critically; and (2) enable you to put the principles and methods of graphical data analysis to work on a substantial problem. The project is intended to be a piece of independent work that is carried out essentially without assistance. The project consists of two parts, both of which are compulsory.

I. Graphic Awareness. A collection of three statistical graphics with written comments on each graphic. These graphics should be collected during the semester from published work. You may not draw your own graphics or ask a friend (or anybody, for that matter) to do so for you. Credit will be given for interesting, carefully chosen graphics which show evidence of reasonable wide searching. In addition to including a copy of the graphic itself, you should document the source of the graphic (title of article, authors, source including title, page numbers etc.) and discuss the graphic. Your discussion may include the reason for the graphic, strengths and weaknesses, etc, and may include redrawn, improved versions of the graphic. The discussion should be brief, relevant and insightful, not longwinded.

II. Graphical Analysis. A few weeks into the course, I will provide a short list of data sets and some documentation for them. You must choose one of these data sets, analyse it, and prepare for submission a concise, well-organized report on your analysis. Your analysis must be appropriate and it must be substantially (though not necessarily exclusively) graphical. Your report should begin with a clear statement of the problem you are addressing and the context in which it arises. You should describe what you have done and why. Relevant graphics and/or output should be included in the report, and all such results should be discussed and interpreted in the text. The entire report must be shorter than 8 pages including graphics and data (any page after the 8th will be ignored), and the written part should not be longer than 4 to 5 pages. Attempts to defy the spirit of the page limit by using unreadable typefaces and so on will be noticed, so please don't do it.

Assignment Details

Due Dates	22 August (A1), 17 October (A2), 23 October (Project)
Value or Weighting (%)	20% (A1), 20% (A2), 10% (Project, Part A), 50% (Project, Part B)
Marks	20 (A1), 20 (A2), 10 (Project, Part A), 50 (Project, Part B)
Length	4 pages (A1); 4 pages (A2); 8 pages (part B of Project – no hard limit on Part A, but a suggested length of 3-4 pages for Part A)

Instructions	Analyse data given and present analysis using R software. The answers should be written in report style in clear language. Hand in both text and graphics as part of your answer. The graphics you hand in should be RELEVANT: that is DO NOT HAND IN EVERY GRAPHIC YOU PRODUCE IN THE PROCESS OF WORKING THROUGH THE ASSIGNMENT. Marks will be deducted if you hand in irrelevant graphics, and if your graphics are not sufficiently adorned with explanatory titles and axis labels and so on. The text part of your answer should be in the form of a report: it is not sufficient to merely annotate the graphics you produce. The text part of your report must be CONCISE and TO THE POINT: answers that are too lengthy may also be penalized. NOTE: There is a page limit on each assignment and Part B of the project, and this limit includes text and all graphics. No pages beyond the page limit will be read.
Purpose	A1: Representing and comparing distributions; A2: Time Series and Dependent Data
Marking Criteria	Marks awarded for correct responses, presentation of analysis and effective layout and relevance.
Submission / Presentation Details	All assignments should be handed in as a single stack of single-sided A4 paper stapled in the top left corner. A cover sheet may be included, but there should be no plastic binders, plastic sleeves or other forms of binding. ONLY a single, stapled stack of paper, printed on one side. Assignments are to be submitted on the due date by dropping completed assignments in the marked boxes outside the Research School of Finance, Actuarial Studies and Applied Statistics main office.
Return of Assignments	Assignments will be returned in class and uncollected assignments will be placed in the filing cabinets in the fourth floor foyer of Building 26C.
Scaling	Your final mark for the course will be based on the **raw** marks allocated for each assignment or examination. However, your final mark may not be the same number as produced by that formula, as marks may be **scaled**. Any scaling applied will preserve the rank order of raw marks (i.e. if your raw mark exceeds that of another student, then your scaled mark will exceed the scaled mark of that student), and may be either up or down.
Extensions	Each assignment will be due just before the first class that discusses its solutions, so as a general rule, NO LATE ASSIGNMENTS WILL BE ACCEPTED.
Penalties	No late assignments will be accepted without the prior approval of the lecturer for the course.
Information about examinations	There are no examinations in this course.

As a further academic integrity control, students may be selected for a 15 minute individual oral examination of their written assessment submissions.

Any student identified, either during the current semester or in retrospect, as having used ghost writing services will be investigated under the University's Academic Misconduct Rule.

Communication with students

Email	If necessary, the lecturers and tutors for this course will contact students electronically using their official ANU student email address. Information about your enrolment and fees from the Registrar and Student Services' office will also be sent to this email address.

Announcements	Students are expected to check the Wattle site for announcements about this course, e.g. changes to timetables or notifications of cancellations. Notifications of emergency cancellations of lectures or tutorials will be posted on the door to the relevant room.
Course URLs	More information about this course may be found on: • Programs and Courses (http://programsandcourses.anu.edu.au/2015/Catalogue) • the College of Business and Economics website, and • Wattle, the University's online learning environment. Log on to Wattle using your student number and your ISIS password.

Tutorial registration

There are **no** formal tutorials in computer laboratories. Roughly once a fortnight, starting in roughly week 3, there will be full-class tutorials held during the one of the weekly lecture times. The format of these tutorials will be for me to demonstrate the solutions to tutorial problems. It is essential that you attempt tutorial problems by yourself in the computer laboratory before coming to these classes. Remember, R is a computer language that you will have to learn for yourself, and it is definitely not a spectator sport. If you just watch me present solutions without attempting them yourself you will find the assignments extremely difficult or even impossible! Solutions to tutorial problems will be made available online from the class web page.

Privacy Notice

The ANU has made a number of third party, online, databases available for students to use. Use of each online database is conditional on student end users first agreeing to the database licensor's terms of service and/or privacy policy. Students should read these carefully.

In some cases student end users will be required to register an account with the database licensor and submit personal information, including their: first name; last name; ANU email address; and other information.

In cases where student end users are asked to submit 'content' to a database, such as an assignment or short answers, the database licensor may only use the student's 'content' in accordance with the terms of service – including any (copyright) licence the student grants to the database licensor.

Any personal information or content a student submits may be stored by the licensor, potentially offshore, and will be used to process the database service in accordance with the licensors terms of service and/or privacy policy.

If any student chooses not to agree to the database licensor's terms of service or privacy policy, the student will not be able to access and use the database. In these circumstances students should contact their lecturer to enquire about alternative arrangements that are available.

SUPPORT FOR STUDENTS

The University offers a number of support services for students. Information on these is available online from http://students.anu.edu.au/studentlife/