

# DEPARTMENT OF ECONOMICS

2022

## APPLIED ECONOMICS 388



## DATA SCIENCE FOR ECONOMISTS

Dr DL Shepherd  
CGW Schumann 614  
[debrashepherd@sun.ac.za](mailto:debrashepherd@sun.ac.za)

## **1. OVERVIEW OF THE MODULE**

This course is aimed at introducing programming and computational tools useful for future careers as economists. The course covers manipulation of data, data visualisation and data analysis, including basic machine learning techniques and numerical methods useful in economics.

## **2. LECTURING SCHEDULE**

This module usually entails four lectures per week. Two of these will be computer-lab based, and the remaining two for formal lectures.

The normal lecture schedule and venue allocation is as follows:

Monday: 13:00-14:00	Computer lab	Fharga E.01 (Neelsie)
Monday: 14:00-15:00	Computer lab	Fharga E.01 (Neelsie)
Wednesday: 15:00-16:00	Formal lecture	Van der Sterr 2118
Thursday: 9:00-10:00	Formal lecture	Van der Sterr 2118

Students are expected to stay up to date, by working through the relevant material posted on SUNLearn each week.

## **3. COMPUTER LAB SESSIONS**

Apart from the formal lectures, there will also be weekly computer lab based classes. In these classes, students will learn how to:

- use the Stata software package;
- wrangle and manipulate data;
- explore and visualise data to answer empirical questions;
- use machine learning and econometric techniques

Students with access to a computer should please install Stata on this computer. Instructions on how to do this will be provided in the first week of classes. Students can then follow these classes and complete exercises on their own computers.

## 4. **ASSESSMENT**

The assessment of this portion of the module will consist of two term tests and one take-home assignment. Your mark for "Introductory Data Science for Economists" contributes 50% to your total Economics 388 mark, and is calculated in the following way:

- Test 1 (online) – 15%
- Test 2 (written) – 20%
- Practical assignment – 15%

There will also be formative assessment opportunities.

## 5. **SOURCES:**

The prescribed text for this course is:

Gutman, A.J., & Goldmeier, J., 2021. Becoming a Data Head: How to Think, Speak, and Understand Data Science, Statistics, and Machine Learning (1<sup>st</sup> ed.), John Wiley & Sons Inc.

Written by award-winning data scientists, this book will give you the language and tools necessary to talk and think critically about data science. Unlimited access (i.e. read online or download) to an electronic version of this text is provided through the Stellenbosch University library. You can also make use of the link:

<https://ebookcentral.proquest.com/lib/sun/detail.action?docID=6561823>

Secondary texts include:

- Bruce, P., Bruce, A., & Gedeck, P., 2021. Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python (2<sup>nd</sup> ed.), O'Reilly Media, Inc.
- Bueno de Mesquita, E., & Fowler, A., 2021. Thinking Clearly With Data (1<sup>st</sup> ed.), Princeton University Press.
- Imai, K., & Bougher, L.D., 2021. Quantitative Social Science: An Introduction in Stata (1<sup>st</sup> ed.), Princeton University Press.

Additional readings (e.g. journal articles), class slides and tutorial notes will be provided on SUNLearn.

## 6. **WORK PROGRAMME**

### 6.1 THINKING DATA SCIENCE

- What is data?
- Thinking statistically
- Correlation vs causation

- Data wrangling
- Data management
- Data manipulation

## **6.2 SPEAKING DATA SCIENCE**

- Data exploration
- Data visualizations
- Sampling and distributions
- Statistical inference
- Hypothesis testing

## **6.3 DATA SCIENCE TOOLBOX**

- Regression for describing
- Regression for prediction
- Classification
- Hidden groups
- Basic machine learning