



# Module Specification

Key Information			
Module title	Data Science		
Level	6	Credit value	15
Member Institution	Goldsmiths	Notional study hours and duration of course	150
Module lead author/ Subject matter expert	Aaron Gerow		
Module co-author			

Rationale for the module
Data science is a significant subfield in computer science. Data science has many application areas ranging from medicine to climate science and business analytics. This module builds on several topics covered in earlier parts of the computer science programme including mathematics, databases, programming and graphics. It provides the skillset required to gather, analyse and present data.

Aims of the module
By taking this module, you will gain a data science skillset. With these skills, you will be able to write computer programs that can read, process and analyse textual and numerical data. You will be able to generate plots and interactive visualisations of data. You will understand how to apply statistical methods to the interpretation of results. You will be able to use data analysis in the decision-making process. You will also learn about a range of application domains for data science.

## Topics covered in this module:

The topics listed here are an approximation of what will be covered. The topics presented may be slightly revised to ensure currency and relevance. Students will be advised of any changes in advance of their study

1. Introduction and development environment
2. Working with different types of data
3. Correlation and regression
4. Distributions and significance
5. Processing text data
6. Decision making based on data analysis: from correlation to causation
7. Introduction to data visualisation
8. Time-series data visualisation
9. Scientific data visualisation
10. Case studies: different contexts for data analysis

Approximately 10-12 hours of study will be required per topic. The remaining study time is intended for coursework and examination preparation.

## Learning outcomes for the module

Students who successfully complete this module will be able to:

1. Write computer programs that can read, clean and analyse textual and numerical data.
2. Evaluate the significance of data analysis results using appropriate statistical tests.
3. Write computer programs that can generate plots and visualisations of data.
4. Explain how decisions can be made based on data analysis and evaluate potential issues that may arise.
5. Analyse and critique examples of data analysis being applied in different contexts.
6. Use a data science development environment to write data analysis software and describe the features of the environment.

## Assessment strategy, assessment methods

### Summative and Formative Assessments

The module will contain a range of summative and formative assessments. Summative assessments are assessments which contribute directly towards your final grade. Formative assessments do not count directly towards your final grade. Instead, they provide you with opportunities for low stakes practice, and will often provide some sort of feedback about your progress. For example, a practice quiz might provide you with feedback about why a particular answer was wrong.

### Assessment Activities

The table below lists the assessment activity types you might encounter taking the module. It also states if that type of assessment can be automatically graded. For example, multiple choice quizzes can be automatically graded, and so can some programming assignments. It also states if that type of assessment will be found in the summative coursework and the summative examination. More details about the summative assessments are provided below.

Assessment activity type	Can it be automatically graded with feedback in some cases?	CW	Examination
Quiz	X	X	X
Writing task		X	X
Programming task	X	X	X
Peer review task		X	

### Pass Mark

In order to pass this module, you must achieve at least 35% in each element of summative assessment and an overall weighted average of 40%, subject to the application of rules for compensation. Please refer to the programme regulations for more information.

### Summative Assessment Elements

As this is a module that has a significant amount of theory it is assessed as a theory-based module. This means that the summative assessment is composed of two elements, whose weightings are listed in the table below.

Summative Assessment Component	Percentage of final credit	Deadline
Coursework	50%	Mid session
Examination	50%	End of session

The coursework comprises a variety of practical exercises and quizzes which in total will take up to 25 hours of study time to complete. The examination will be two hours long, and consist of written answer and multiple choice questions.

## Learning resources

The module will draw on a number of different, largely web-based, public resources as well as the resources produced as bespoke material for this module. The standard text book(s) for the module will be:

Python Data Science Handbook: Tools and Techniques for Developers,  
Jake VanderPlas, O'Reilly Media 2016

Edward R Tufte, The Visual Display of Quantitative Information  
[http://www.edwardtufte.com/tufte/books\\_vdqi](http://www.edwardtufte.com/tufte/books_vdqi)

Ben Fry, Visualizing Data: Exploring and Explaining Data with the Processing Environment  
<http://shop.oreilly.com/product/9780596514556.do>