ACM Data Science Task Force Course Example

Foundations of Data Science (COMP6235)

University of Southampton, UK
Dr Adriane Chapman, Dr Jian Shi, Prof. Elena Simperl
Adriane.Chapman@soton.ac.uk

.

Knowledge Areas that contain competencies (knowledge, skills, and dispositions) covered in the course

Knowledge Area	**Total Number of Contact Hours**
Big Data Systems (BDS)	12
Data Acquisition, Management and Governance (DAMG)	14
Data Mining (DM)	14
Professionalism (PR)	4

Where does the course fit in your undergraduate Data Science curriculum?
It is used as an introductory course at Masters level.

Is this course from or used in other curricula/majors?
It is used as an introductory course at a number of Masters courses, e.g. in Artificial Intelligence. To quote: 'Although our MSc is offered in the School of ECS, and our assessments across the modules reflect that computational focus, our MSc takes in students from any discipline as long as they can show they have had a solid math module, and can show experience with programming (python, java, etc.). Because of this, we spend a bit of extra time in Foundations trying to make sure the entire cohort has the same base skills with respect to statistics and tools.'

What is covered in the course?
1. Statistics

a. a coverage of (very basic) stats to make sure everyone is on the same page (mostly because this was not available in other modules, and students were coming in from very different backgrounds)
b. tools and libraries that are useful/helpful

2. Data handling
   a. basics and "big data"
   b. how to get data - open data, crowd-sourcing, etc.
   c. cloud technologies
   d. tools and libraries
3. Open Topics and Data Science research e.g. When Data Science goes wrong (fairness); data privacy; crowd-sourcing; applications of data science

What is the format of the course?

Normally two one-hour lectures and two one-hour tutorials each week for 11 weeks in the first semester. There is also related coursework. With Covid this will move to online delivery.

How are students assessed?

There are 3 coursework assessments:

CW1 (30%) Emphasis on statistics set in week 2, to be handed in at week 8:
Tech Report. Students are given a dataset and asked a series of questions they should try to answer (e.g. when is it best to fish). They must apply the stats they've learned, using the tools, and write a report that answers the business questions.

CW2 (30%) Emphasis on Data Management, handed out in week 8, due in week 11
jupyter notebook. They are given a dataset and asked a series of questions. They must use the tools to analyse the data and get answers.

CW3 (40%) A group project, hand in at week 14, followed by talk at week 15.
PowerPoint (and video of PowerPoint presented). Students are put into groups. The groups bid on a topic, and we try to spread the topics. Each group must then research some aspect of that topic, scope it as they want, etc. They then present their research to the class.

Course tools and materials

It is vital that students read up on the topics covered in the lectures. The following list covers some key books for the course.

- Cathy O'Neil and Rachel Schutt, Doing Data Science, O'Reilly, 2014.
- Russell Jurney, Agile Data Science, O'Reilly, 2013.
- Paul Teetor, R Cookbook, O'Reilly, 2011.
- Tom White, Hadoop: The Definitive Guide, O'Reilly, 2015.

Why do you teach the course this way?

Since students typically hold a first degree, they are able to embark on the format described elsewhere.

Body of Knowledge coverage

KA	Sub-domain	Competencies Covered	Hours
BDS	Techniques of Big Data Applications, Cloud Computing, Software Support for Big Data applications	Demonstration of proficiency in the use of various relevant tools, including statistics	12
DG	Data acquisition, Information extraction, Data cleaning	Skills to be addressed through attention to aspects of data handling	14
DM	Data preparation, Information extraction	Data handling provides the context for the demonstration of good practice	14
PR	Continuing professional development, Teamwork, Privacy and confidentiality, Ethical considerations	Professionalism is addressed and has to be demonstrated largely through group work and major project activity	4