



Foundations of Data Science Course Syllabus

Course	Foundations of Data Science	Faculty	Philosophy
Course Code	NCHAI759	Course Leader	Dr Alexandros Koliouis
Credit Points	15	Teaching Period	Hilary
FHEQ Level	Level 7	Date Approved	June 2020
Compulsory/Optional	Compulsory	Date Modified	
Pre-requisites	None		
Co-requisites	None		

WELCOME

This course introduces students to advanced data programming and analysis techniques to process data sets and derive meaningful, statistically sound observations (e.g., summary statistics).

DESCRIPTION

Data Science is about extracting generalisable, useful and meaningful knowledge from data in a systematic manner. The course covers the emerging field of Data Science at breadth, honing on both programming and data analytics skills.

Students will learn to work with tensors (i.e., multi-dimensional arrays) and apply linear algebra transformations using, e.g., the `numpy` Python library; load, integrate and process structured and unstructured data from multiple sources using `pandas`; apply statistical and machine learning analysis algorithms using `scikit-learn`; and visualise results using `matplotlib`.

AIMS

The aim of the course is to set strong foundations in modern data science technologies and methods. More specifically, the course aims to:

- Develop familiarity with interactive and integrated development environments in Python
- Understand arrays and vectorised computation

- Understand data set representations of data transformations
- Develop ability to load and clean data sets, conduct descriptive, meaningful statistics and visualise results

LEARNING OUTCOMES

On successful completion of the course, students should be able to:

KNOWLEDGE AND UNDERSTANDING

- K1d demonstrate critical awareness and knowledge of basic data science concepts
- K2d demonstrate comprehensive knowledge of feasible operations on data and transformation on data
- K3d show knowledge and understanding of plotting and visualising data
- K4d demonstrate sound knowledge of applying techniques to a data case-study

SUBJECT-SPECIFIC SKILLS

- S1d engage critically in the theory behind the concepts taught in the class
- S2d demonstrate originality in applying the data transformation techniques in an appropriate manner to the chosen dataset
- S3d identify and apply the correct choice of appropriate data transformation techniques

TRANSFERABLE SKILLS

- T1d demonstrate initiative in working independently, effectively, and to deadlines
- T2d identify, transform, critically evaluate and plot accordingly from the dataset
- T3d produce clear and concise and well documented code

LEARNING AND TEACHING PLAN

Teaching and learning strategies for this course include:

- 10 × 1.5 hours of full-cohort lectures
- 10 × 1.5 hours of lab-based tutorials
- 2 office hours per teaching week

Course information and supplementary materials are available on the College's Virtual Learning Environment (VLE).

At the end of Hilary term, students will attend *Collections*, formal meetings where they receive comprehensive and collated feedback about their performance over the term.

Students are required to attend and participate in all the formal and timetabled sessions for this course. Students are also expected to manage their directed learning and independent study in support of the course.

FEEDBACK

Students receive feedback in a variety of ways:

- Written, including via email correspondence¹
- Oral – during office hours, on an *ad hoc* basis, and during lectures and lab sessions
- *Collections* (see above)

Feedback is provided on both formative and summative assignments (normally, within a week of submission for formative and within two weeks for summative assignments) and it is made available on the College's VLE.

TEACHING SCHEDULE

Hilary Term		
Week	Topic	Reading ²
1	Program design in Python • Introduction • Expressions • Objects • Control flow statements – conditionals & iterations • Data structures – lists, tuples & dictionaries • Program structure – functions & packages	Handbook: • §4.0-4.11 Scratch book: • Ch. 1 • Ch. 2 • Ch. 3
2	Data representations • Vectors, arrays & tensors • Algebraic operations • Pandas – series & data frames	Handbook: • Ch. 2 • §3.0 • §3.1 • Ch. 4 Scratch book: • Ch. 4
3	Data analytics	Handbook: • §3.2

¹ **Email policy.** You are welcome to reach me, the course leader, via email and I aim to respond in a timely manner. Please ensure your email subject starts with "FODS:", followed by your subject. This way, I can identify and group your messages appropriately.

² "Handbook" refers to the book *Python Data Science Handbook*. "Scratch book" refers to the book *Data Science from Scratch*. See [ESSENTIAL READING](#) section.

Hilary Term		
Week	Topic	Reading ²
	- Relational algebra – selection, projection, join & aggregation operators - Data streams - MapReduce (data parallelism) - Data visualisation	§3.3 §§3.5-3.9 §3.11 §3.12 - Ch. 4 Scratch book: - Ch. 25
4	Data-driven applications (A and Ω) - Data collection - Web scraping - Data storage & retrieval - Data pre-processing - Data visualisation	Handbook: §3.4 §3.10 - Ch. 4 Scratch book: - Ch. 9
5	Statistics - Random variables - Probability distributions - Gaussian distribution & Mixtures of Gaussians t-tests & Analysis of Variance (ANOVA) - Degrees of freedom	Scratch book: - Ch. 5 - Ch. 6 - Ch. 7
6	Regression - Curve fitting - Linear regression - Regularisation - Support Vector Machines - Clustering – k-nearest neighbours & k-means clustering - Principal Component Analysis	Handbook: §§5.0-5.2 §5.6 §5.7 §5.9 §5.11 Scratch book: - Ch. 12 - Ch. 14 - Ch. 20
7	<i>Reading Week</i>	–
8	Classification - Bayesian models - Decision trees & random forests - Logistic regression - Neural networks	Handbook: §5.5 §5.8 Scratch book: - Ch. 13 - Ch. 17 - Ch. 18
9	Machine learning applications (A to Ω) - Feature engineering & selection - Model selection & model validation - Confusion matrix	Handbook: §5.3 §5.4

Hilary Term		
Week	Topic	Reading ²
	- Hyper-parameter tuning - Classification reports	
10	Text Analytics - Strings – text as data - Text pre-processing - Sentiment analysis	Handbook: §3.10 Scratch book: - Ch. 21
11	Selected Topics in Data Science - Graph analytics - Practical ethical considerations	
12	<i>Collections</i>	–

Please refer to your CELCAT timetable for exact dates and times of lectures, seminars and tutorials.

ASSESSMENT

Please refer to the “Formative and Summative Assessment Planner” for the **submission dates** of your formative and summative assignments. All assignment briefs can be found on the on the course page on the College’s VLE.

FORMATIVE

There is formative work every teaching week. Students will be given a mark for all formative work – not necessarily a precise numerical mark, but at least a qualitative indication of where a piece of work lies. The mark will be communicated to students at the time feedback is given, for it constitutes an element of that feedback:

Qualitative mark	Numerical range
Distinction	72 and above
Merit	62-68
Pass	52-58
Fail	42-48

SUMMATIVE

Students will be formally assessed with **two, equally weighted assignments**:

No.	Assignment Type	Weight (%)	Details	Length
-----	-----------------	------------	---------	--------

AE1	Coding assignment	50	Online	Code and ~2,500 words
AE2	Coding assignment	50	Online	Code and ~2,500 words

READING

REQUIRED READING



Python Data Science Handbook

By Jake VanderPlas

2016

O'Reilly Media, Inc.

ISBN: 9781491912058

The book is available online on [GitHub](#). It is also available for download from O'Reilly via Northeastern's University Library.



Data Science from Scratch, 2nd Edition

By Joel Grus

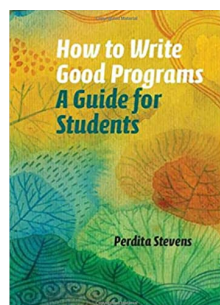
2019

O'Reilly Media, Inc.

ISBN: 9781492041139

The book is available online from O'Reilly via Northeastern's University Library ([link](#)).

FURTHER READING



How to Write Good Programs: A Guide for Students

By Perdita Stevens

2020

Cambridge University Press

ISBN: 9781108789875

The book is available online for download from the publisher via Northeastern's University Library ([link](#)).

CHALLENGING READING

MapReduce: simplified data processing on large clusters • J. Dean and S. Ghemawat • Communications of the ACM • January 2008 • [DOI](#)

A relational model of data for large shared data banks • E. F. Codd • Communications of the ACM • June 1970 • [DOI](#)

SciPy Lecture Notes – One document to learn numerics, science and data with Python • [link](#)

Finding the frequent items in streams of data • G. Cormode and M. Hadjieleftheriou • Communications of the ACM • October 2009 • [link](#)

APPENDIX A: COMMON ASSESSMENT MARKING SCHEME

Distinction		Merit		Pass		Fail	
100	Highest possible distinction	68	High merit	58	High pass	48	High fail
90	Extremely high distinction	65	Mid merit	55	Mid pass	42	Mid Fail
85	Very high distinction	62	Low merit	52	Low pass	35	Clear Fail
80	High distinction					20	Fail
75	Mid distinction					5	Almost no attempt
72	Low distinction					0	No attempt or late Submission

APPENDIX B: GENERIC GRADE CRITERIA - LEVEL 7

KNOWLEDGE AND UNDERSTANDING			
100/90/85/80/75/72	68/65/62	58/55/52	48/42/35/20/5/0
Exceptional analysis of key issues/ concepts/ethics with very clear originality and autonomy.	Outstanding levels of accuracy, technical competence, organisation, and expression.	Shows a reasonable understanding of the major factual and/or theoretical issues involved.	Little development of factual or theoretical issues to demonstrates no knowledge or understanding related to the question set to failure to answer the question or develop a relevant argument.
Demonstrates independence of thought and a very high level of intellectual rigour and consistency.	Very high levels of creativity, originality and independence of thought.	Shows evidence of planning and selection from appropriate sources.	Evidence of misunderstanding to scripts contain clear factual errors or misunderstandings.
Exceptional development of argument and the ability to make informed judgements to develops a sophisticated and intelligent argument.	Shows strong evidence of critical insight and critical thinking.	Demonstrates some knowledge of the literature.	Some evidence of planning is demonstrated, but irrelevant material or arguments are included to demonstrates short answers and incoherent argument.
Shows clear evidence of wide and relevant reading and an engagement with the conceptual issues.	Shows a detailed understanding of the major factual and/or theoretical issues and directly engages with the relevant literature on the topic.	Shows, in places, examples of a clear train of thought or argument.	Demonstrates no knowledge of the key issues in the relevant literature.

Shows a rigorous use and a sophisticated understanding of relevant source materials, balancing appropriately between factual detail and key theoretical issues and are evaluated directly, and their assumptions and arguments challenged and/or appraised.	Develops a focussed and clear argument and articulates clearly and convincingly a sustained train of logical thought.	Introduced and concludes appropriately.	Demonstrates little to no evidence of critical thought or analysis.
Shows original thinking and a willingness to take risks.	Shows clear evidence of planning and appropriate choice of sources and methodology, and ability of synthesis under exam pressure.		

SUBJECT SPECIFIC			
100/90/85/80/75/72	68/65/62	58/55/52	48/42/35/20/5/0
Very significant ability to plan, organise and execute independently a research project or coursework assignment to significant ability to plan, organise and execute independently a research project or coursework assignment.	Ability to plan, organise and execute independently a research project or coursework assignment.	Ability to plan, organise and execute a research project or coursework assignment.	Limited ability to plan, organise and execute a research project or coursework assignment to a very limited ability to plan, organise and execute a research project or coursework assignment to no demonstrable ability to plan, organise and execute a

			research project or coursework assignment.
Very significant ability to evaluate literature and theory critically and make informed judgements and very significant ability to analyse data critically.	Strong evidence of critical insight and thinking.	A reasonable understanding of the major factual and/or theoretical issues involved.	Some awareness and understanding of the literature and of factual or theoretical issues, but with little development to clear conceptual or factual errors or misunderstandings to Little or no knowledge or understanding related to the assessment.
Very high levels of creativity, originality and independence of thought to significant ability to plan, organise and execute independently a research project or coursework assignment.	A detailed understanding of the major factual and/or theoretical issues and directly engages with the relevant literature on the topic.	Evidence of some knowledge of the literature with correct referencing.	Limited ability to analyse data to fragmentary evidence of critical thought or data analysis to no evidence of critical thought or data analysis.
Very significant ability to evaluate critically existing methodologies and suggest new approaches to current research or professional practice.	Clear evidence of planning and appropriate choice of sources and methodology with correct referencing.	Shows examples of a clear train of thought or argument.	Incomplete referencing to not engaging with the relevant literature or demonstrate a knowledge of the key issues to Little or no knowledge of the relevant literature to major errors in referencing.
Outstanding levels of accuracy, technical competence, organisation, expression with capacity to develop a	Ability to analyse data critically.	Ability to analyse data.	Limited ability to present a clear and coherent argument to a failure to develop a coherent argument that

sophisticated and intelligent argument.			relates to the research project or assignment to Incoherent argument.
Clear evidence of wide and relevant reading, referencing and an engagement with the conceptual issues.	Capacity to develop a focussed and clear argument and articulate clearly and convincingly a sustained train of logical thought.	Introduced and concludes appropriately.	
Rigorous use and a sophisticated understanding of relevant source materials, balancing appropriately between factual detail and key theoretical issues. Materials are evaluated directly, and their assumptions and arguments challenged and/or appraised			
Original thinking and a willingness to take risks.			

TRANSFERABLE SKILLS			
100/90/85/80/75/72	68/65/62	58/55/52	48/42/35/20/5/0
Exceptional presentation and organisation of work and fluent communication in all contexts.	Presentation and organisation of work appropriate to context and purpose, communication clear to very high-level of communication skills in a range of complex contexts, and ability	Satisfactory organisation and presentation of work, communications mostly appropriate to the context/purpose.	Organisation and presentation of work and communications adequate in most contexts, with some mistakes/irrelevancies to elements of

This work is of publishable quality, with only very minor amendments, and would be likely to receive that judgement if submitted to a peer-reviewed journal. Work is of such a quality that the student is clearly highly capable of doctoral research in the discipline and, in principle, should be prioritised for a postgraduate research grant.	to write at publishable standard. Demonstrates very effective communication in a range of complex and specialised contexts.	Demonstrates capabilities to support effective communication in a range of complex and specialised contexts.	disorganisation/poor presentation/poor communication or expression to communications being too brief or rambling, inappropriate to context or purpose, with many errors /omissions, inadequately expressed/presented.
Exemplary referencing/citation to referencing/citation comprehensive.	Referencing/citation consistent and accurate.	Referencing/citation largely consistent/accurate.	Some errors in referencing/citation to errors/omissions in referencing/citation, or substantial errors/omissions in referencing/citation, or none.
Work demonstrates considerable initiative and autonomy in setting challenging objectives and taking accountability for outcomes to work demonstrating initiative and autonomy in setting challenging	Work demonstrates some initiative and autonomy in setting objectives beyond those given and taking accountability for outcomes to demonstrating the independent learning ability required for	Work demonstrates satisfactory independence in addressing objectives (some beyond those given) and taking responsibility for outcomes.	Work demonstrates adequate independence in addressing mainly given objectives and taking responsibility for outcomes to demonstration of insufficient independence in attempting to address given objectives and taking

objectives and taking accountability for outcomes.	continuing professional development.		responsibility for outcomes to lacking independence, does not address objectives and fails to take responsibility for outcomes.
Evidences advanced team-working and leadership skills to evidence of excellent team-working and leadership skills.	Evidences a high level of team-working and leadership skills.	Evidences team-working and some leadership skills.	Significant weaknesses evident in key areas such as communication, problem-solving and project management. Inability to adapt and to work flexibly, independently and/or as part of a team.
Critical reflection and self-evaluation sustained and exceptionally insightful to Reflection and self-evaluation consistently critical and insightful. Demonstrates a confidence with self-direction and originality in tackling and solving demanding problems.	Sustained reflection and self-evaluation generates a number of critical insights.	Satisfactory self-evaluation and reflection with some critical insights.	
Confidently acts autonomously in planning and implementing tasks at a professional or equivalent	Shows a very high level to high level of employability skills, including team	Shows a consistently good level of employability skills, including team working,	Demonstrates generally effective employability skills, including communication and

level. Demonstrates the skills and attitudes needed to advance own knowledge and understanding, and to develop new skills.	working/leadership, project management, IT/computer literacy, creativity and flexibility. Demonstrates autonomy and notable originality in tackling and solving demanding problems and demonstrates the independent learning ability required for continuing professional development.	project management, IT/computer literacy, creativity and flexibility Demonstrates ability to advance own knowledge and understanding, and to develop new skills. Demonstrates the independent learning ability required for continuing professional development	problem-solving, but with some problematic areas of weakness to limited ability to adapt to inability to work flexibly, independently and/or as part of a team, but with areas of weakness.
--	--	--	--



New College
of the Humanities