

ECSC10038 Data Science in Ecology and Environmental Science

Thursdays 13:10 – 16:00; Room: Teaching Studio G.09, Murchison House

Course Github Page: <https://github.com/EdDataScienceEES/DataScienceHub2025/>

Course Organiser:	Dr Hannah Wauchope (hannah.wauchope@ed.ac.uk)	Other Key Staff:	Dr James Watt (james.watt@ed.ac.uk)
Course Administrator:	Francesca Finnegan Tel: (0131 6)50 4842 Email: francesca.finnegan@ed.ac.uk	Course location:	Kings Buildings & Online
Credits available:	20	SCQF Level:	11

Summary

We live in a data driven world, and knowing how to manage, visualise, analyse and interpret data is a key skill for ecological and environmental scientists. This is a crash course in the kinds of skills a student will need to take off in a data driven career.

Course description

We focus our teaching around ecological and environmental science problems, but much of what we cover will be applicable in a wide range of disciplines. See the weekly breakdown of topics below. We place a heavy emphasis on the importance of critical thought - our aim is to help you learn how to think about numbers, and to not be scared by statistical representations. This critical thinking will extend to AI – we will discuss when it is and isn’t useful, and consider best practice for AI assisted coding. You are allowed to use AI to support assessment, but our in-class sessions are an AI free zone – consider it gym training for your brain. We teach mostly in ‘R’ – the preferred coding language of ecologists and environmental sciences.

Course Material and Flipped Classroom Approach

The course will be hosted online through the GitHub platform, used by professional data scientists, allowing for version control, reproducible workflows and submission of coding-based assignments (<https://github.com/>).

Each week, class is taught in one three-hour session, consisting of an introductory lecture, then workshop time (sometimes individual, sometimes small groups) for you to engage in the content and practice their own work, supported by the teaching team.

We teach a flipped classroom approach – each week you will have tutorials and readings to do which we expect you to have worked on *before you come to class*. It is very important you engage with this content – the less you do, the more you will struggle. These materials will be clearly communicated through GitHub (and Learn for Week 1).

Learning Outcomes

- Understand key quantitative skills in the disciplines of ecology and environmental sciences including data management, data visualization, programming and statistical analysis
- Use data science tools to address research questions and challenges in ecology and environmental sciences.
- Critically evaluate statistical analyses and reports for reliability and bias
- Implement version control to back up work and write reproducible workflow reports.
- Practice teaching quantitative skills and develop an online tutorial.

Course Schedule

Week	Date	Lecture Topic	Lecturer	Assessment
1	18 th Sept	Welcome + Functional Programming	Wauchope + Watt	
2	25 th Sept	Version Control	Watt	1. Clean-That-Code Set
3	2 nd Oct	Data Wrangling	Watt	
4	9 th Oct	Data Visualisation	Wauchope	1. Clean-That-Code Due – 10th Oct
5	16 th Oct	Building Linear Models	Wauchope	2. WWF Report Set
6	23 rd Oct	Building Generalised and Hierarchical Linear Models	Wauchope	
7	30 th Oct	Interpreting Model Output (Model fit to Data)	Wauchope	
8	6 th Nov	Interpreting Model Output (Model representation of reality)	Wauchope	2. WWF Report Due – Friday 7th Nov 3+4. Tutorial + Reflective Piece Set
9	13 th Nov	Google Earth Engine	Watt	
10	20 th Nov	Dashboarding	Watt	
11		NO CLASS		3+4. Tutorial + Reflective Piece Due – Friday 28th Nov

Assessment

Data Science has one formative assessment and 3 summative assessments.

REMEMBER for summative assessments we often mark both your Learn Submission **AND** the assessment Github Repo. You **MUST** ensure your repo is finalised by the deadline, when you will be locked out AND submit the relevant material to Learn. **IF YOU DO NOT SUBMIT TO LEARN YOU WILL FAIL THE ASSIGNMENT.**

- **Assessment 1 (FORMATIVE): 'Clean-That-Code', DUE DATE FRIDAY 10th OCTOBER, 12pm**
 - This is an assignment to test and develop your coding skills. It would be *very* easy to complete this assessment with AI, but we challenge you not to – the point here is improve your coding skills and understanding, so that you have the knowledge to sense check AI when you *do* use it.
- **Assessment 2 (SUMMATIVE – 40%): WWF Report DUE DATE FRIDAY 7th NOVEMBER, 12pm**
 - You will use the statistical models taught in weeks 5-8 to assess the how the population size of a species of your choice is changing. You will produce a report as if for WWF (the charity) detailing the trajectory of the species, how you have chosen to model the data to reach this conclusion, and how confident you are in your findings. Submit this report to learn, we will mark this *and* your github repo.
 - *AI Usage Statement:* Full AI ("AI tools may be used to identify ideas, plan, edit, and generate text. You must acknowledge any AI use in your submission.")
- **Assessments 3 & 4 (SUMMATIVE): Tutorial (40%) + Reflective Piece (20%) DUE DATE FRIDAY 28th NOVEMBER, 12pm**
 - This is an assessment two parter (combined together for reasons of submission logistics).
 - *Assessment 3* is a Tutorial: you will produce a website tutorial on a data science subject of your choice, aimed at a moderately skilled student at a similar level to you. We ask you to submit a link to the tutorial alongside...
 - *Assessment 4*, a 500-word reflection, detailing how your thinking about data science has changed from taking this course.
 - *AI Usage Statement:* AI Assisted Editing ("AI tools may be used for identifying ideas, planning, and improving the clarity of your writing, but not for content generation. AI use must be acknowledged in your submission.")

Students must attain an overall average of 40% (or above) to pass the course.

See marking rubrics below. You will be given a grade breakdown by each broader category (e.g. 'Model Construction' in the WWF Report)

2. WWF Report Rubric		Outstanding 80-100	Excellent (70-79)	Very good (60-69)	Good (50-59)	Pass (40-49)	Fail (0-39)
Model Construction (25%)	<p>Model Choice and Justification (60%): All four model choices are reasonable and justified</p> <p>Assumptions (40%): Model assumptions are tested and discussed, variance explained by the model is discussed.</p>	All components correctly identified and excellently justified	All components correctly identified and well justified	Most components correctly identified and most correctly justified	Only some components correctly identified and/or justified	At least one component correctly identified and at least one correctly justified	No components correctly identified or none correctly justified
Model Interpretation (25%)	<p>Model Output (50%): Explanatory variables accurately discussed according to sign, strength, standard error and significance</p> <p>Biases and Confounders (50%): The extent to which the model is likely to reflect reality is discussed</p>	Assumptions are tested correctly, and these + model variance are interpreted with depth and nuance	All components correctly described and their meaning for model interpretation explained excellently	Assumptions are tested correctly and they + variance mostly interpreted well	Assumptions are tested, but either not appropriately, or not well interpreted, Variance mentioned but poorly	Assumptions are tested but inappropriately, and interpretations poor. OR these or variance not mentioned	Assumption testing is entirely wrong, or not interpreted
Presentation (25%)	<p>Figures (25%): At least one figure showing raw data and modelled outputs (incl. error), with appropriate caption</p> <p>Table (25%): A table displays model output incl standard error, formatted in markdown, good number of significant figures</p> <p>Text (25%): Report is written in clear, concise scientific language, with introduction and summary statement</p> <p>Report Format/Creativity (25%): The report is formatted to look professional and attractive</p>	Figure(s) are beautifully displayed and accurate with perfect captions, possibly extra e.g. map	Figure(s) are attractively displayed, containing all information we might want with detailed captions	Figures are correct, display most information we might want, captions are good but lacking in detail	All of the above, but there are errors in the data Or, data is correct, but figures are poorly formatted	As before but there are serious errors, or very poor formatting	Either no figures or, for some marks, figures present but both wrong and poorly formatted
Reproducibility (25%)	<p>Repo (25%): Clearly organised with folders, and an informative readme</p> <p>Code (25%): Runs without error and is clearly commented, including explanations of any model testing or decisions</p> <p>Data Citation (5%): Data is appropriately cited in the report</p> <p>AI Statement (45%): A reflective statement detailing and critiquing how AI has been used to shape this work</p>	Repo is perfectly organised with and excellent signposting in the readme.	Repo is well organised with folders, and clear signposting in the readme	Repo is organised into folders, with decent signposting. Occasionally could be clearer.	Repo has has some organisation, and file headings and signposting mostly clear.	Some attempt has been made to organise repo, but it is still hard to find ones way around.	Repo is poorly organised, no editing of readme.

3. Tutorial Rubric		Outstanding 80-100	Excellent (70-79)	Very good (60-69)	Good (50-59)	Pass (40-49)	Fail (0-39)
Topic and Structure (20%)	Topic (50%): Tutorial is of a level appropriate to someone who has taken data science	Tutorial is highly advanced, guiding the learner well beyond skills covered in data science. We as lecturers may learn something new.	Tutorial is advanced, stretching someone who has taken data science to new ways of thinking	At a good level for someone who has taken data science (or as above but not entirely coherent as topic)	Pretty easy for someone who has taken Data Science, doesn't introduce new ways of thinking (or as above but not very coherent as topic)	Basic, likely too easy for someone who has taken data science (or as above but not at all coherent as topic)	Covers something extremely easy, or already covered in class. And/or no coherence at all to what is taught.
	Learning Objectives (50%): 3-5 distinct learning objectives, each taking ~5-10mins for a reader to work through	3-5 learning objectives that are clearly defined, hold together with a logical flow, and are of an appropriate length	As above, but less strong on at least one score (e.g. flow not entirely logical)	As above, but less strong on multiple scores. Length of the minority is inappropriate	Quite poorly done OR either too many or too few OR majority of inappropriate length	As before, but change 'OR' to 'AND'	
Content (40%)	Writing (33%): The tutorial is clear and logical, with a written structure that guides the reader and helps them feel situated in their learning. (Think: intros, wrap ups, next steps etc)	The reader feels carried by the writing, never confused by where the flow is going. Situates the learning of this specific tutorial within a broader picture.	Excellent structure, text flows from one section to the next, just lacking some of the details of above	Very good structure, only occasionally are there logic gaps or moments where the reader isn't clear	Good structure, though at times the reader may feel they're not clear on e.g. what's coming next or why	Some structure, but there are logical inconsistencies or gaps in flow, leading to confusion	Very poor structure, no introductions or summaries, logical inconsistencies.
	Concepts (33%): The concepts being taught are clearly and correctly explained	Concepts are beautifully explained in erudite ways that bring new understanding to the reader, explanations from multiple angles	Concepts excellently explained, broken down where necessary, easy to understand	Concepts are explained well, perhaps just occasional confusion or inconsistencies	Concepts mostly well explained, some confusion or error at times	Attempts are made to explain concepts, but they are confused or wrong at points	No attempt made to explain concepts or why certain things must be done. OR entirely wrong explanations.
	Reproducibility (33%): The reader can correctly follow along at home, accessing and understanding all necessary code, data, etc, with nothing behind pay walls.	All code and data clearly accessible and appropriately cited. Code runs with no errors, and is clearly commented. If repo referenced, it is clearly organised. If no code/repo, other equivalent guides are well used (e.g. screenshots, formulas)	As before, with small inconsistencies (e.g. not ideal citation, 1-2 coding errors)	As before, but multiple inconsistencies	The reader is able to follow along, just, but must sift through errors and confusions at many points	The reader cannot follow along, e.g. code or data inaccessible, insufficient instructions	
Creativity (40%)	Creativity in Teaching (33%): Creative additions are used to assist with teaching, beyond just clear writing/explanations	The web interface is used to full effect with multiple innovative techniques used, including interactive elements	Multiple creative techniques used, just less innovation than above	At least one creative approach to teaching used to effectively help grasp ideas	At least one creative approach used in teaching, but it lacks effectiveness	Some effort at creative teaching, but it only helps a little	No, or entirely wrong, attempts at creative teaching
	Figures (33%): At least two figures are used to demonstrate learning outcomes or explain concepts (NOTE as distinguished from purely aesthetic figures)	More than two gorgeous figures that elucidate concepts/outcomes in innovative ways	At least two figures that are professionally formatted and effective in demonstrating concepts/outcomes	At least two well formatted figures that demonstrate concepts/outcomes	At least two figures, but formatting is a little unpolished, or they are not effective	At least two figures but they are unprofessional or ineffective, OR as above but 1 figure	One or fewer figures, with no redeeming qualities
	Markdown Visuals (33%): The markdown format is used to creative an atractive tutorial including headings, images, banners etc	Tutorial looks like a fully professional website. Multiple aesthetic visuals, plus text formatting used to full effect	As above, just not quite at professional level	Tutorial visually coherent and attractive, multiple visuals and text formatting used, just not always to full effect	As before, though some inconsistencies in look	At least some attempt at making tutorial attractive (images, text formatting), but not visually coherent	No attempt made to make tutorial visually appealing

4. Reflective Piece Rubric	Outstanding 80-100	Excellent (70-79)	Very good (60-69)	Good (50-59)	Pass (40-49)	Fail (0-39)	
Connections (25%) ("What aspects of the course have affected you the most in terms of skills/ways of thinking?")	Does the reflection connect specific course topics (e.g., statistical modelling, data visualisation etc.) to the development of critical skills or new philosophies/ways of thinking (e.g. analytical thinking, data ethics, critical appraisal of stats in public discourse)?	In depth connections are made from multiple angles, to multiple aspects of the course. These are linked to development of specific critical thinking skills and philosophies that demonstrate deep and wide thinking and reading on various topics, beyond clear course engagement.	Deep connections are made to multiple aspects of the course. These are linked to development of critical thinking skills and philosophies that demonstrate the student has engaged deeply with the course	Connections are made to multiple aspects of the course, and there are clear connections to critical thinking skills and philosophies	Connections are made between at least one aspect of the course and the development of some critical thinking skills or philosophies, but depth is lacking	At least some connection is made, but there is a lack of nuance and depth. Evidence of at least some engagement with the course.	Connections are not made, it is clear the student has not engaged with the course.
Relevance (25%) ("Why are these particular connections meaningful to you given past experiences/ways of thinking?")	Does the reflection clearly demonstrate why the learning experiences being reflected on are particularly and personally meaningful to the student, with relevance to past experiences or ways of thinking (i.e. why is what you're talking about personal to you?)?	It is obvious that the connections have been deeply meaningful, with specific examples of why they hold particular resonance for the student. There is a clear narrative of how the course fits in to specific life experiences and past ways of thinking	It is clear that the experiences have been personally meaningful, with specific examples, just perhaps less nuance or narrative than 'Outstanding'	There is evidence that connections have been personally meaningful, at least one specific example is given, but there is not as much depth	There is evidence that connections have been meaningful, with at least one example, but evidence is superficial and/or example is not especially relevant	Some evidence that connections have been meaningful, but no or very poor examples given, and not much depth evident. The piece could have been written by any student, it's not clear what is personal to the author.	No evidence that the learning experiences have been meaningful
Impact (25%) ("How will you use these connections for future actions and ways of thinking?")	Does the reflection demonstrate how the development of new skills/ways of thinking (as detailed in 'connections') will impact their future, either in thinking about datascience, critical thinking, or related subjects?	Deep and considered thinking is shown about how the development of new connections will impact future decisions and ways of thinking, with clear examples demonstrating new approaches from multiple angles	Excellent evidence and examples are given to show thinking about how connections will impact future decisions, with multiple examples	Great evidence is given about how the student will approach situations in new ways or acting or thinking in the future, but there is a lack of depth. At least one good quality example demonstrates this	Good evidence of how connections might lead to new ways or acting or thinking in the future, but there is a lack of depth. At least one example, but also shallow or less relevant	Some attempt to discuss how connections will be used for future actions and ways of thinking, but shallow, and with no or poor examples.	No evidence given about how connections might lead to new ways of acting or thinking in the future.
Clarity (25%) (Writing style, explanation of concepts)	Reflection is well-organized, with clear and logical structure. Abstract concepts are explained accurately for a reader unfamiliar with data science.	Outstanding structure as with other criteria AND the narrative flows seamlessly, and there is a satisfying conclusion	Excellent structure, the reader is always clear about where the piece is going, ideas hold together, and there are clearly delineated paragraphs	Great structure, good use of topic sentences, and most ideas hold together to a common conclusion. Just some points where the flow is not totally clear.	Good structure, but at times it's not clear what specific point paragraphs are addressing, or how the essay will progress next. Some ideas feel out of place.	Some attempt at structure, but the piece is confused, and it is frequently unclear what the topic of a paragraph is or why an idea is being introduced	The piece is very confused, jumps to new ideas with no context. It is hard to understand what is going on or why things are being said.
		Truly excellent, creative, crystal clear explanations of data science concepts: though the piece still contains all appropriate depth, an intelligent 10 year old could grasp the outline of the ideas	Excellent, creative explanation of abstract concepts, that captures the ideas and complexities without getting lost in unnecessary details	Great explanation of data science concepts, but at times ideas still require a moment to understand, especially to someone who has not taken data science	Good explanation of data science concepts, but often ideas require a moment to understand, especially to someone who has not taken data science	Some attempt at explanation of data science concepts, but the reader would likely be confused, or unable to understand, unless they had taken this course	Concepts very confusing, even to someone who has taken data science