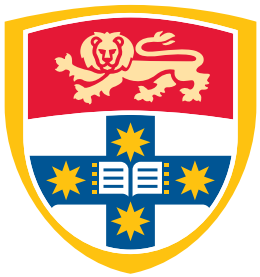
Honours in Data Science

Detailed Guide for the 2025 academic year



THE UNIVERSITY OF SYDNEY

School of Mathematics and Statistics

Contents

1 Entry requirements

Preliminary entrance into the honours program is through the Faculty of Science application portal. The Faculty requirements which must be met include:

- have qualified for or be a graduate with a Bachelor of Science degree or equivalent from the University of Sydney or equivalent qualification from another tertiary institution;
- have completed a relevant major (i.e. minimum of 24 credit points of 3000-level units of study) relating to the intended Honours discipline;
- have achieved a Weighted Average Mark (WAM) of at least 65.00 or have a credit average (65.00) in 48 credit points of relevant 2000-level and 3000-level units of study; and
- *securing the agreement of a supervisor.*

Please note that, in some Schools and Disciplines, the minimum WAM requirement is higher than 65.00, particularly where entry is very competitive. Please, check with the Honours coordinator and the potential supervisor about these extra-requirements.

1.1 Formally

The faculty offers three main Honours pathways and it can be confusing:

- Students who are currently enrolled in a bachelor's degree at the University of Sydney, see our advice on the current student website.
- If you've completed a qualifying bachelor's degree and are either returning after graduation or transferring from another institution:
  - Our appended honours degree, the Bachelor of Science (Honours), is an option for students who did not complete two majors in their single bachelor's degree. If you completed a Bachelor of Liberal Arts and Science degree, you can apply for the Bachelor of Liberal Arts and Science (Honours).

1.2 It's important to note that:

- All acceptances into Honours (including in cases where the School's requirements are not met) are ultimately at the discretion of the School. However, a student meeting all of the above criteria (or the equivalent from another institution) should be confident of acceptance.
- The Faculty of Science Honours **application deadlines** are
  - 15 January 2025 for Honours commencement in Semester 1, 2025;
  - 25 January 2025 for Honours commencement in Semester 2, 2025.

No application will be accepted after the deadline.

2 Structure of Honours

An Honours year in Data Science involves

- four 6CP courses (worth 50% of the final mark) and
- and a project (worth 50%).

2.1 The Honours project (or thesis) (50%)

The Honours project centres around an essay/thesis consisting of roughly 60 pages[1] written on a particular topic from your chosen area. It does not need to contain original research (although it might) but it should clearly demonstrate that you have understood and mastered the material. The assessment of the Honours thesis is based on the scientific and quantitative content and its exposition, including the written english. The thesis is due at the end of your second semester, specifically at 5pm on Monday of Week 13.

Toward the end of the second semester (Friday weeks 9-10), each student gives a 25 minutes talk on their thesis project. The aim of the talk is to explain to a broader audience the purpose and nature of the project. The talk is followed by 5 minutes dedicated to questions from the audience which includes staff members and fellow students.

2.1.1 Writing proficiency

As mentioned above your essay is also assessed based on the quality of the writing. This does not mean we look for the next Shakespeare however you should make sure you express your ideas in an organized manner using a clear and grammatically correct English. The University of Sydney offers several resources that can help you achieve this goal. The Learning Hub offers workshops for students that need help with extended written work, and a trove of online resources for improving your writing skills is also available. Make sure you make use of these resources as early as possible as writing skills develop slowly over time and with much practice.

2.2 Course work (50%)

The Honours program in *Data Science* specifies different combinations of courses that can be taken including courses offered by the School of Mathematics and Statistics and the School of Computer Science.

A list of courses that will be offered in 2025 is available in the *Data Science* degree structure document which outlines the combinations of courses that can be taken for credit. Please, be careful that this list may change each semester (and some units may be offered some years and not others).

If in doubt, contact the Honours Coordinator.

[1] This page number is a very rough guideline and should not be taken as binding.

3 Important course work information for all students

3.1 Selecting your courses

It is a requirement to **select your courses after consulting the Honours supervisor and the Honours coordinator!**

3.2 AMSI courses

Students are welcomed to check the courses offered in January at the AMSI Summer School and also courses available via the Advanced Collaborative Environment (ACE). These courses can possibly be taken for credit (by enrolling in the unit AMSI4001), but this can only be done in consultation with the student's supervisor and with the approvals of the specific honours coordinator as well as the School's Honours coordinator, Prof. Laurentiu Paunescu.

Start the procedure early and before enrolling, to be sure that you will get approval before the start of the Summer School.

4 Program Administration

The Data Science Honours coordinator is

Dr. Clara Grazian,
Carslaw Building, Room 822
Email: clara.grazian@sydney.edu.au

The Co-director of Teaching (Statistics & Data Science) is

A/Prof. John Ormerod,
Carslaw Building, Room 815, Phone 9351 5883,
Email: jennifer.chan@sydney.edu.au

The Honours Coordinator is the person that students should consult on all matters regarding the Honours program. In particular, students wishing to substitute a course from another Department, School or University must get prior written approval from the Honours Coordinator. Matters of ill-health or misadventure should also be referred to the Honours Coordinator.

5 Honours courses in Data Science

The Bachelor of Advanced Studies (Honours) (Data Scence) requires 48 credit points from this table including:

- 12 credit points of 4000-level and above Honours coursework selective units from List 1, and
- 12 credit points of 4000-level and above Honours coursework selective units from List 1, List 2, List 4 or List 5 with a maximum of 6 credit points of units from List 5, and
- 24 credit points of 4000-level Honours research project units

Note:

- Not all courses are offered every year. Moreover, some courses may have pre-requisites and exclusions. Please, check the Honours page for udpate information about offerings, pre-requisites, and exclusions.
- Some of you may want to enrol in courses which are not available under Sydney Student, but are possible under the rules of the Honours program. In this case you need to ask for a **manual enrolment**.

  To do so, you need to i) have your supervisor's approval and send it to the Honours coordinator, ii) send the Honours coordinator the full list of 4 courses you are taking across the Honours year, iii) send the Honours coordinator your undergraduate transcripts, iv) get the Honours coordinator's approval, v) send it to Sydney Student.

  Instructions will be provided at the beginning of the semester.
- **Under no circumstances, the Honours coordinator can approve enrolment to a course if the pre-requisites are not met without the approval of the Unit Coordinator of the course**. Enrolment in courses is subject to check of the pre-requisites. In the case the pre-requisites are not met, it is responsibility of the Unit Coordinator of the course to approve or reject the request to enrol in the course.
- Consider that Honours students can enrol in a very limited number of level 5000 courses out of the School of Mathematics and Statistics (List 5).

6 Project

6.1 General information on projects

Students can work on a project of their own topic provided they secure in advance the supervision of a member of the University of Sydney (from any School) and provided they receive the approval of the Honours Coordinator.

Each student is expected to have made a choice of a project and supervisor well before the beginning of the first semester (or the beginning of the second semester for students starting in July). Students are welcome to consult on this matter with the Honours coordinator. The Honours coordinator must be informed of the choice of supervisor before the start of the program.

Work on the project should start as soon as possible but no later than the start of the semester. The break between the semesters is often an excellent time to concentrate on your research but you should make sure you make continuous progress on your research throughout the year. To ensure that, students should consult their appointed supervisor regularly, in both the researching and writing of the work.

Prospective students interested in any of these topics are encouraged to discuss them with the named supervisors as early as possible.

The Honours project in Data Science may include synthesising and generalising results from the statistical literature, developing novel methodologies or attacking a problem in applied statistics in an innovative way.

One electronic copy of the thesis must be submitted to the Honours coordinator before the beginning of the study vacation at the end of your last semester. The exact date will be made known.

It is recommended that you go through the following checklist before submitting your thesis:

- Is there an adequate introduction?
- Have the chapters been linked so that there is overall continuity?
- Is the account self-contained?
- Are the results clearly formulated?
- Are the proofs correct? Are the proofs complete?
- Have you cited all the references?

6.2 Proposed DARE/DATA61 Student projects

Under the Data Science Honours project, it is possible to do a project jointly between the DARE Centre and Data61, a business unit of CSIRO.

Enrolment in these projects is by application, and it is competitive: several students may apply for the same project. You need to apply for these projects before application for the Honours program: please, contact the Honours coordinator early if interested in any of them.

1. **Hierarchical Implicit Networks for Prediction and Uncertainty Quantification for Hydrological and Agricultural Time Series**
   Data61 Supervisors: Dan Pagendam, Joel Dabrowski and Kelly Trinh
   Deep neural models are powerful tools for learning complex patterns in data and have great potential in forecasting to aid agriculture under a changing climate. We propose to assess the utility of hierarchical implicit networks (HINs) as tools for time series forecasting in hydrology and agriculture. A HIN involves the use of two neural networks: (i) a network that makes the prediction; and (ii) a network that predicts the parameters of the first network. By configuring the second network to produce parameter predictions that are samples from a distribution (i.e., as a probabilistic generative model), the first network will be able to provide ensembles of predictions that quantify predictive uncertainty; an important consideration when using a forecast to make decisions in agricultural and environmental management. Specifically, we propose to apply HINs to forecasting problems involving hydrological time series (e.g. groundwater levels; freely available from the Bureau of Meteorology) and to simulated agricultural time series available from APSIM.

2. **Multimodal Large Language Models for Climate Adaptation**
   Data61 Supervisors: Sarvnaz Karimi, Maciek Rybinski, Necva Bolucu, TBC
   Research in Large Language Models, in particular in the field of Natural Language Processing (NLP), has shown promises to make information presented in documents in all forms of text, tabular data, and images accessible through search or question answering systems. The state-of-the-art however for such multimodal data from scientific reports and literature is lagging behind other more generic sources. This is even exacerbated when low-resource languages are involved. We seek to advance this area by improving NLP and machine learning algorithms that represent these multimodal multilingual resources to improve their accessibility, especially in the application area of climate science.

3. **Connecting Natural Disasters and their Impacts to Climate Processes**
   Data61 Supervisors: Nandini Ramesh and Mahesh Prakash
   Natural disasters of meteorological origin, such as floods or bushfires, are generally short-term, rapid events spanning a few days; however, they are made more or less probable through the occurrence of slower variations on seasonal or longer timescales. This project will aim to identify these relationships between short-term and long-term events. Using a newly-available natural disaster dataset along with climatological data from the Bureau of Meteorology and international agencies, this project will analyze these relationships in Australia and the Indo-Pacific region using a range of spatial statistical methods, including Poisson regression and Gaussian processes, in order to develop maps of disaster probabilities under varying climatic conditions.

4. **Scaling EvaSim - Revolutionizing Emergency Response with Advanced Traffic Simulation**

Data61 Supervisors: Hanna Grzybowska, Dhirendra Singh, Vincent Lemiale
The main objective of this project is to expand our innovative simulation-based tool, EvaSim, designed to support first responders in managing traffic loads during bushfire evacuations. EvaSim, part of the SEEKER platform, offers pre-planned and alternative evacuation routes, focusing on minimizing clearance times from multiple zones to arrival points (e.g. shelters). The tool employs a spatial-queue-based dynamic traffic simulation model, ready to incorporate behavioural models and data depicting evolution of bushfires. Our next step is to scale up EvaSim for large-scale simulations, capable of handling thousands and millions of agents on a state and national level. This involves optimizing and enhancing the simulation's speed and efficiency. We have already implemented the simulation kernel in efficient C++, and the project aims to further optimize and scale the tool for broader applications. This project will make a significant impact on emergency response planning through cutting-edge technology and strong C++ expertise.

5. **Preserving extremes in future climate data through archetypal analysis**
Data61 Supervisors: Nikhil Garg, Raymond Cohen, Geoff Lee
Climate model outputs require bias corrections to ensure that these models are representative of recorded climate observations. However, this calibration may introduce changes to climate extremes; for example, rainfall extremes can be underpredicted in future climate projections in certain regions, limiting their efficacy in decision-making. Archetypal clustering is a technique used to identify groupings that are most different from each other and allows a probabilistic interpretation of how similar data points are to a given cluster. We propose to apply this novel technique to quantify the extremes before and after bias correction and devise techniques to preserve climate extremes through the calibration process.

6. **Quantifying changes in natural hazard likelihoods under future climate variability**
Data61 Supervisors: Raymond Cohen, Geoff Lee
Natural hazard risk assessments require detailed hazard consequences and corresponding likelihoods. These are typically obtained through a combination of modelling and historical records. Under future climate conditions, the likelihood of a given hazard (eg coastal inundation, flooding, bushfires, heat stress, cyclones etc) will change, but the relationship between changes in likelihood of different climate parameters and hazards needs to be better understood. This project aims to develop and validate novel data-driven techniques that establish these required relationships. This will enable us to perform risk assessments that account for future climate variability at a national scale, facilitating improved decision-making.

7. **Data-driven models for aerosol transport in the atmospheric boundary layer**
DARE Supervisors: Matt Cleary, Clara Grazian
Plume models are fused with measured data (e.g. from drone-mounted sensors) to predict aerosol concentrations over broad areas with application to cloud formation or toxic contaminant dispersion in a city street canyon (e.g. following a terrorist attack).

8. **Data-driven models for combustion in energy and propulsion systems**
DARE Supervisors: Matt Cleary
Statistical models for combustion are fused with sensor data to better predict propulsion and work output.

9. **Quantifying biomass production and quality using drones and image analysis**
   DARE Supervisors: Willem Vervoort, Floris Van Ogtrop
   Capturing pasture production at large scales is complex due to spatial and temporal variation. Drones capture hyperspectral imagery across a range of scales, but conversion of images to production is still underdeveloped. Using drones, you will develop new ways to quantify pasture growth and quality.

10. **Improving irrigation efficiency using digital twins**
    DARE Supervisors: Willem Vervoort, Floris Van Ogtrop
    Digital twins (simulation models) are an effective way to evaluate irrigation strategies. This project will use an extensive database of field data to develop digital twins for border check irrigation in Uruguay. The outcomes will highlight a modelling approach for scenarios in irrigation efficiency.

11. **Groundwater Knowledge Graph Modelling**
    DARE Supervisors: Willem Vervoort, Monica Bian
    The project aims to combine the existing knowledge of hydrogeologists, other domain experts and data scientists to construct a knowledge graph representation of a groundwater system. Scope of the project involves combining the measured quantitative time series data with qualitative and descriptive conceptual data on groundwater rock systems. The resulting groundwater knowledge graph contains known facts (entities and their links) extracted from the data and provides a structured representation of the knowledge. Link prediction approaches are proposed to predict new links for the knowledge graph given the existing links among the entities, thus allowing prediction of unknown relations and behaviours.

12. **Understanding the impact of environmental and climate changes on groundwater resources**
    DARE Supervisors: Willem Vervoort, Monica Bian
    Groundwater is a reliable but limited resource for a range of uses such as mining, irrigated agriculture, and human consumption. A lack of balance between groundwater pumping and replenishment by recharge of the groundwater has led to concerns of overuse. Managing this requires greater understanding into the dynamics of groundwater levels. The project analyses the past and current groundwater data as training and validation data for groundwater level predictions. Using the state-of-the-art embedding and change detection algorithms, the project aims to provide better understanding of the groundwater level changes under the influence of varying environmental and climate factors.

7 Assessment

7.1 The honours grade

The student's honours grade is based on the average mark achieved by each student, over the 4 courses and the project. Courses account for 50% of the assessment and the project for the remaining 50%.

According to the Faculty of Science guidelines, the grade of Honours to be awarded is determined by the honours mark as follows:

Grade of Honours	Faculty-Scale
First Class, (possibly) with Medal	90–100
First Class	80–89
Second Class, First Division	75–79
Second Class, Second Division	70–74
Third Class	65–69
Fail	0–64

The Faculty has also given the following detailed guidelines for assessing of student performance in Honours.

95–100 Outstanding First Class quality of clear Medal standard, demonstrating independent thought throughout, a flair for the subject, comprehensive knowledge of the subject area and a level of achievement similar to that expected by first rate academic journals. This mark reflects an exceptional achievement with a high degree of initiative and self-reliance, considerable student input into the direction of the study, and critical evaluation of the established work in the area.

90-94 Very high standard of work similar to above but overall performance is borderline for award of a Medal. Lower level of performance in certain categories or areas of study above.

Note that in order to qualify for the award of a university medal, it is necessary but not sufficient for a candidate to achieve a SCIWAM of 80 or greater and an honours mark of 90 or greater. Faculty has agreed that more than one medal may be awarded in the subject of an honours course.

The relevant Senate Resolution reads: "A candidate with an outstanding performance in the subject of an honours course shall, if deemed of sufficient merit by the Faculty, receive a bronze medal."

80-89 Clear First Class quality, showing a command of the field both broad and deep, with the presentation of some novel insights. Student will have shown a solid foundation of conceptual thought and a breadth of factual knowledge of the discipline, clear familiarity with and ability to use central methodology and experimental practices of the discipline, and clear evidence of some independence of thought in the subject area.

Some student input into the direction of the study or development of techniques, and critical discussion of the outcomes.

75-79 Second class Honours, first division – student will have shown a command of the theory and practice of the discipline. They will have demonstrated their ability to conduct work at an independent level and complete tasks in a timely manner, and have an adequate understanding of the background factual basis of the subject. Student shows some initiative but is more reliant on other people for ideas and techniques and project is dependent on supervisor's suggestions. Student is dedicated to work and capable of undertaking a higher degree.

70-74 Second class Honours, second division – student is proficient in the theory and practice of their discipline but has not developed complete independence of thought, practical mastery or clarity of presentation. Student shows adequate but limited understanding of the topic and has largely followed the direction of the supervisor.

65-69 Third class Honours – performance indicates that the student has successfully completed the work, but at a standard barely meeting Honours criteria. The student's understanding of the topic is extremely limited and they have shown little or no independence of thought or performance.

0-64 The student's performance in fourth year is not such as to justify the award of Honours.

7.2 The coursework mark

Students are required to attend 4 courses of 6CP during the academic year and the coursework mark is a simple average of the courses they took.

Student performance in each honours course is assessed by a combination of assignments and examinations. The assignment component is determined by the lecturer of each course and the examination component makes up the balance to 100%.

7.3 The project mark

The project's mark is split 90% for the essay and 10% for the student's presentation. The presentation mark is determined by the stats staff attending the presentation.

The essay is assessed by three members of staff (including the supervisor). The overall final mark for the essay is a weighted average of all three marks awarded. A weighting of 50% is attached to the supervisor's original mark, while a weight of 25% is attached to each of the two marks awarded by the other examiners.

The criteria which the essay marks are awarded by each examiner include:

- quality of synthesis of material in view of difficulty and scope of topic, and originality, if any;
- evidence of understanding;
- clarity, style and presentation;
- mathematical and/or modelling expertise and/or computing skills.

The student's supervisor will also consider the following criteria:

- Has the student shown initiative and hard work which are not superficially evident from the written report?

- Has the student coped well with a topic which is too broad or not clearly defined?

7.4 Procedures

All assessable student work (such as assignments and projects) should be completed and submitted by the advertised date. If this is not possible, approval for an extension should be sought in advance from the lecturer concerned or (in the case of honours projects) from the Honours Coordinator. Unless there are compelling circumstances, and approval for an extension has been obtained in advance, late submissions will attract penalties as determined by the Board of Examiners (taking into account any applications for special consideration).

Appeals against the assessment of any component of the course, or against the class of Honours awarded, should be directed to the Head of School.

Note: Students who have worked on their projects as Vacation Scholars are required to make a declaration to that effect in the Preface of their theses.

8 Seminars

Mathematical Statistics and Data Science seminars are usually held every week on Friday afternoons. These seminars are an important forum for communicating ideas, developing critical skills and interacting with your peers and senior colleagues. Seminars are usually given by staff members and invited speakers. All Honours students are encouraged to attend these seminars. Keep in mind that attending these seminars might help develop your presentation skills.

9 Entitlements

Honours students enjoy a number of privileges, which should be regarded as a tradition rather than an absolute right. These include:

- Office space and a desk in the Carslaw building.
- A computer account with access to e-mail and the internet, as well as LaTeX and laser printing facilities for the preparation of projects.
- Photocopy machine for any of your work related material.
- After-hours access to the Carslaw building.
- A pigeon-hole in room 728 — please inspect it regularly as lecturers often use it to hand out relevant material.
- Participation in the School's social events.
- Class representative at School meetings.

10 Scholarships, Prizes and Awards

University of Sydney Honours Scholarships
These $6,000 Honours Scholarships are awarded annually on the basis of academic merit and personal attributes such as leadership and creativity.

University Medal
Awarded to Honours students who perform outstandingly. The award is subject to Faculty rules, which require a mark of at least 90. More than one medal may be awarded in any year. Notice that received an Honours mark equal or higher than 90 does not imply receiving a University medal.

11 Life after Fourth Year

Students seeking assistance with post-grad opportunities and job applications should feel free to ask lecturers most familiar with their work for advice and written references. The Head of Statistics Programme, the Program Coordinator and the course lecturers may also provide advice and personal references for interested students.

Students thinking of enrolling for a higher degree (MSc or PhD) should direct all enquiries to the Director of Postgraduate Studies:

`pg-director@maths.usyd.edu.au`

Students are also strongly encouraged to discuss potential research topics with individual staff members.

Students who do well in their Honours studies may be eligible for postgraduate scholarships, which provide financial support during subsequent study for higher degrees.

Last but not least, there is a number of jobs for people with good statistical knowledge. Have a look here.

Every year the NSW Branch of the Statistical Society of Australia organises an Career Event in September (in 2024, it is on the 19th of September) that may be of interest.