DATA2001 – Data Science, Big Data, and Data Diversity
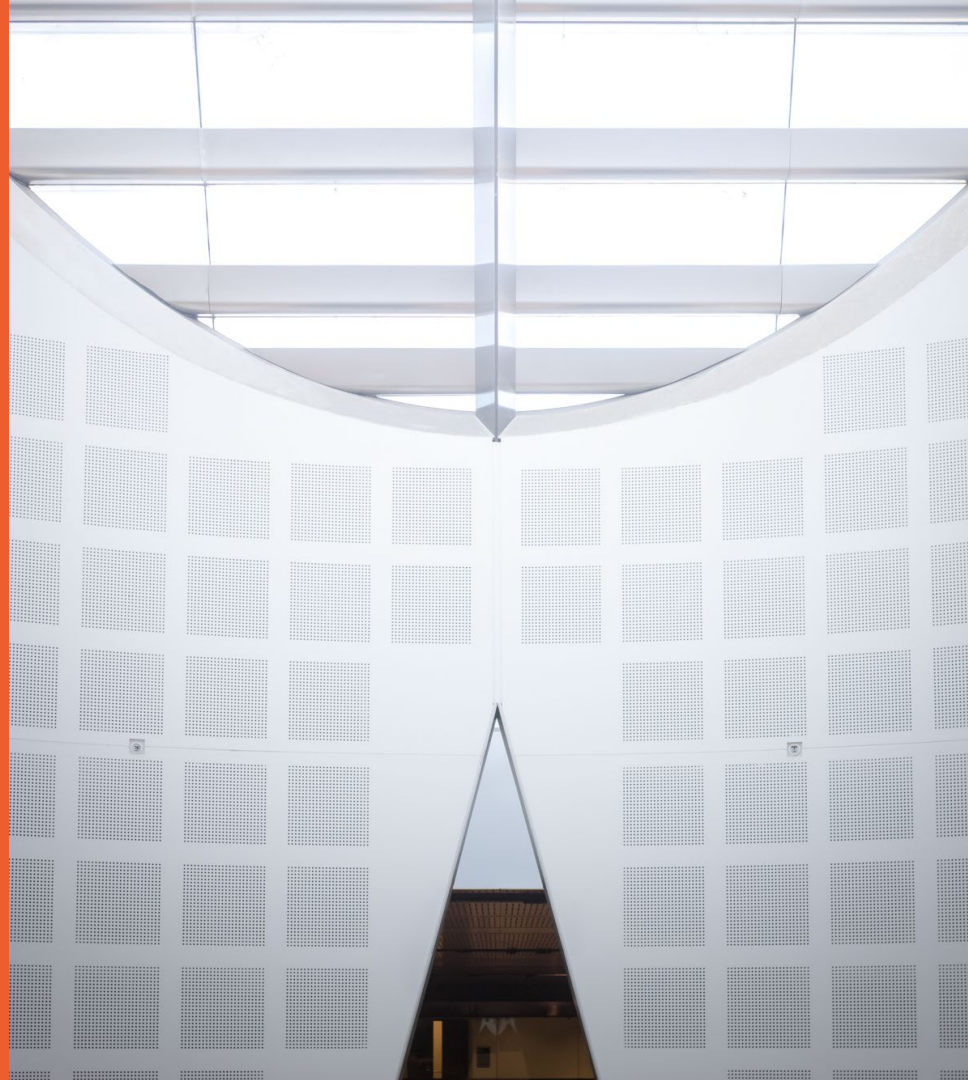## Week 1: Administrativa

Presented by

Dr Ali Anaissi

School of Computer Science

THE UNIVERSITY OF
SYDNEY

Acknowledgement of Country

I would like to acknowledge the Traditional Owners of Australia and recognise their continuing connection to land, water and culture. I am currently on the land of the Gadigal people of the Eora Nation and pay my respects to their Elders, past, present and emerging.

I further acknowledge the Traditional Owners of the country on which you are on and pay respects to their Elders, past, present and future.

Welcome to DATA2001

– Lecture: Wednesday, 11am –1pm

   + 2h tutorials on Thursdays and Fridays (starting Week 2)

– Lecturer: **Dr Ali Anaissi**
                Consultation Time: By email

– Contact Email: ali.anaissi@sydney.edu.au

– Tutors: Jarrod (TA) plus a team of 11 tutors => cf. Canvas

– Course Web page: **Canvas**

Course Objectives

– Knowledge of the main challenges analysing 'Big Data':
  <u>Data Volume</u>, <u>Variety</u>, *Velocity*, *Veracity*

– Understand the impact of data volume on data processing
  – Indexing
  – Declarative data processing
  – Need for distributed computation

– Ability to perform declarative data analysis with SQL

– Experience with handling datasets of diverse kinds of data
  – Ability to use appropriate Python libraries to ingest, combine and summarise data from a variety of data models

Outline of the Semester

	Week	Topic
Data Volume / Scalability	Week 1	Introduction & Organisation
	Week 2	Data Analysis with Python
	Week 3	Accessing Data in a Database
	Week 4	Declarative Data Analysis with SQL
	Week 5	Scalable Data Analytics
Data Variety	Week 6	Web Scraping Data
	Week 7	Web APIs and NoSQL; mid-semester SQL Test
	Week 8	Text Data Processing *(no tutorials => Good Friday)*
		Easter Break
	Week 9	(Geo-)Spatial Data
	Week 10	Timeseries Data / Health Data
	Week 11	Image Data Processing
	Week 12	Big Data Processing
	Week 13	Unit of Study Review

Prerequisites

– Assumed you have already done **DATA1002** or an equivalent computational course before

  – Procedural Programming

  – Python

– You will be working with PostgreSQL, usually through pgAdmin, and also some online learning platform for SQL

  – Tutorials and assignments

– You will also be expected to do programming tasks in Python

Discussion Forum

–   We will use Ed STEM as discussion forum

–   All announcements will be posted there

–   If you need help from a tutor or lecturer, send email or ask in lecture.

–   Everything else should be asked on discussion forum

–   Ask questions and answer other students' questions

    –   Tutors will monitor, but may wait a couple of days to give students an opportunity to respond

    –   Rate good questions to raise their prominence

Assessment

Assessment Package

Component	DATA2001/2901
Weekly Review Quizzes	10%
SQL Test (week 7)	15%
Assignment (week 12)	20%
Final Exam	55%

- All progressive marks will be consolidated in https://canvas.sydney.edu.au
  - Report any errors/omissions within 10 days
- A pass requires at least:
  a) **>= 40% in exam marks**; and
  b) **>= 50% overall mark.**
- All submitted work must be your own:
  - http://sydney.edu.au/engineering/student-policies/academic-honesty.shtml

Weekly Review Quizzes

- Every Week (starting Week 2), there will be some homework questions about the lecture content in Canvas
  - Combination of short answer questions and **multiple-choice questions** on concepts of the lecture of the week
  - Randomised question sets
- You have one week to work on the set of questions
- There is only a single submission (no re-submission)
  - But we plan(!) to make review quizzes available after the due time where you can re-try a different set of questions for 0pts
- Submission by Tuesday midnight (AEDT) each week in **Canvas**
- **First review quiz in Week 2**

SQL Tutorials / SQL Test

– Online self-learning tutorial for SQL
  – Starting this week
  – Covers all aspects of the SQL database query language
  – Various degrees of difficulty; consists of about eight modules
  – We will be using the **Grok** online learning platform for this
  – **Online-tutor support** will be made available too
    (cf. announcement in Ed and Canvas)

– **Formal SQL Test in Week 7**
  – Your knowledge of SQL will be assessed by means of an in-semester online SQL test
  – Set of SQL questions to be answered in time
  – Conducted with online system, during the scheduled lecture time in Week 7

Practical Assignment

– Groupwork

  – Aim to work in pairs

  – All members must be in the same tutorial

  – Groups to be arranged before Easter break

  – Scenario will be introduced later in the lecture,
    discussion of progress with your tutors throughout the semester

– Submission will be code **and** written report

  – Reports must be submitted via Turnitin, and checked for plagiarism

  – Deadline Week 12

  – Late Submissions: 5% penalty per day late

Final Examination

– Written exam

 – Mostly short answer

 – Possibly some MCQ

– Can cover content encountered anywhere through the unit:

 – Lectures

 – Tutorials

 – any Homework

 – Programming and Assignment (but won't assess coding)

Lateness and Academic Integrity

Special Consideration Policy

– If your performance on assessments is affected by **illness** or **misadventure**

– Follow proper administrative procedures

  – Have professional practitioner sign special USyd form

  – Submit application for special consideration online, upload scans

  – Note you have only a quite short deadline for applying: **within 3 days**

  – http://sydney.edu.au/current_students/special_consideration/

– Also, notify coordinator by email *as soon as anything begins to go wrong*

– There is a similar process if you need special arrangements, e.g., for religious observance, military service, representative sports

Academic Integrity

- <u>Academic integrity</u> refers to behaving honestly, ethically and responsibly in relation to all elements of your study at the university, including assessments.

- Always submit your own work, sit your own tests, and take your own examinations.

- Acknowledge any contributions in your assignment which are not your original thoughts, ideas or words.

- <u>Academic Honesty Education Module</u> – all commencing students must complete by census date. Continuing students can <u>self-enrol</u> at any time.

Strategies for maintaining academic integrity

Planning and time management

Use citations and referencing

Know your strengths and what you need to develop

Know when and where to ask for help

THE UNIVERSITY OF SYDNEY

Academic Integrity (University Policy)

- "The University of Sydney is unequivocally opposed to, and intolerant of, plagiarism and academic dishonesty.
  - Academic dishonesty means seeking to obtain or obtaining academic advantage for oneself or for others (including in the assessment or publication of work) by dishonest or unfair means.
  - Plagiarism means presenting another person's work as one's own work by presenting, copying or reproducing it without appropriate acknowledgement of the source." [from site below]

  http://sydney.edu.au/elearning/student/EI/index.shtml

- Submitted work is compared against other work (from students, the internet, etc)
  - TurnitIn for textual tasks (through Canvas), other systems for code
- Penalties for academic dishonesty or plagiarism can be severe
- Complete self-education AHEM1001 (required to pass INFOxxxx)

Example: Allowed Sources of Help

Source of Help

Lecturer	Teaching Assistants / Tutors	Classmates	Private tutors	Online forums/ Online tutors	Students outside course/UoS	Hired coders Tutorial Company outside University	Relatives	Other

- Individual & group assessment
- A student needs to gain an understanding of fundamental knowledge/skills
- A student needs to gain skills to find, evaluate and apply existing knowledge/solutions

🟩 Encouraged
🟨 Attribution required
🟥 Not acceptable
⬜ Ask Lecturer/Coordinator

Example: Acceptable Types of Help

Types of Help

Understanding General Concepts	Explained using similar material (not assignment)	Sharing approach/concept to derive assignment solution	Designing code/solution	Implementing code/solution

- Individual & group assessment
- A student needs to gain an understanding of fundamental knowledge/skills
- It is important to master the knowledge/skills themselves
- Students are encourage to obtain help through relevant teaching material and practices

■ Encourage
■ Attribution required
■ Not acceptable
■ Ask Lecturer/Coordinator

Health and Safety ; Housekeeping

Help keep our campus COVID safe

- The University is following NSW Government and NSW Health guidance as a minimum standard in our response to the COVID-19 pandemic.
- Please follow these safety precautions:
  - ✓ Stay home if you are sick
  - ✓ Wash hands regularly
  - ✓ Avoid physical greetings
  - ✓ Wear a mask
  - ✓ Cough or sneeze into your elbow or a tissue
  - ✓ Keep 1.5m away from others.

Feeling unwell?

- Stay at home
  - if you are feeling unwell with any COVID-19 symptoms
  - If you have been directed to self-isolate
- Get tested
  - If you are feeling unwell with COVID-19 symptoms, please get tested as soon as possible
- Did you test positive?

  Yes? If you have visited campus within the last 72hours you must advise the University via:
  - email covid19.taskforce@sydney.edu.au, or
  - call +61 2 9351 2000 (select option 1)
- Stay informed
  - Monitor the list of confirmed COVID case locations on campus page to check for potential exposure and follow NSW Health isolation and testing requirements.

Keeping our community safe

- Most large lectures will be delivered online and accommodations will be made for international students who have not yet returned to Australia.

- If you become infected with COVID-19 during the semester, or need to isolate, please notify your unit of study coordinator, as with any unexpected absence.

- If COVID-19 isolation or illness impacts assessment, use the usual mechanisms including simple extensions and special consideration to arrange reasonable adjustments. Visit https://www.sydney.edu.au/covid-19/students/study-information/test-exams-assessment.html#consideration.

- Further information on student support can be found on the University website at https://www.sydney.edu.au/covid-19/students/support-wellbeing.html

- Other helpful study information can be found on the website at https://www.sydney.edu.au/covid-19/students/study-information.html

General Housekeeping – Use of Labs

- Keep work area clean and orderly
- Remove trip hazards around desk area
- No food and drink near machines
- No smoking permitted within University buildings
- Do not unplug or move equipment without permission

Emergency Procedures (on campus)

- In the unlikely event of an emergency, we may need to evacuate the building.

- If we need to evacuate, we will ask you to take your belongings and follow the green exit signs.

- We will move a safe distance from the building and maintain physical distancing whilst waiting until the emergency is over.

- In some circumstances, we might be asked to remain inside the building for our own safety. We call this a lockdown or shelter-in-place.

- More information is available at www.sydney.edu.au/emergency

EMERGENCIES

WHERE IS YOUR CLOSEST SAFE EXIT ?
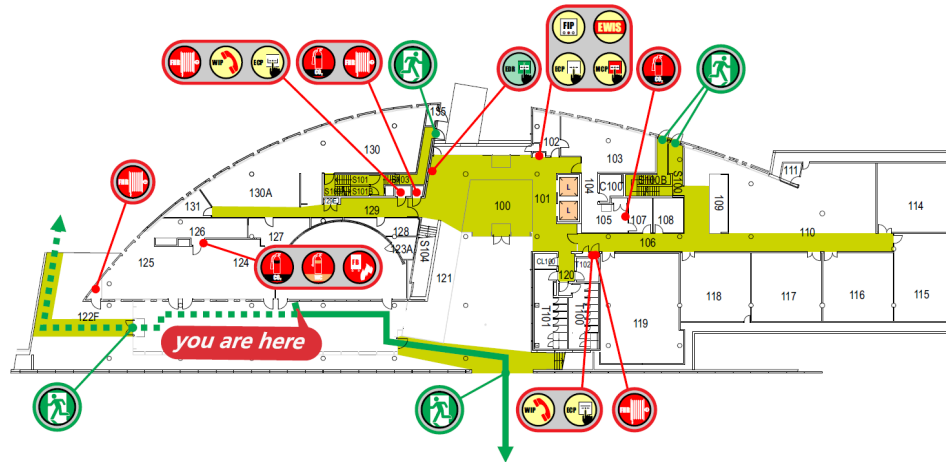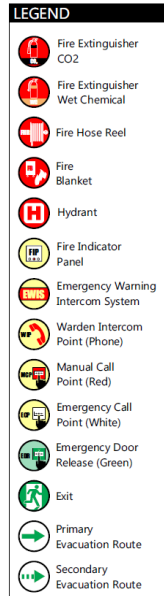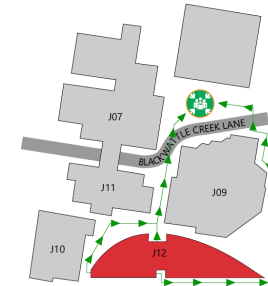
EMERGENCIES

Evacuation Procedures

ALARMS

🔊 **BEEP…BEEP…** - Prepare to evacuate
1. Check for any sign of immediate danger
2. Shut down equipment & processes
3. Collect any nearby personal items

🔊 **WHOOP…WHOOP…** - Evacuate the building
1. Follow the Exit signs
2. Escort visitors & those who require assistance
3. Do not use the lifts
4. Proceed to the Assembly Area

EMERGENCY RESPONSE

1. Warn anyone in immediate danger
2. Fight the fire or contain the emergency, if safe & trained to do so

If necessary…
3. Close the door, if safe to do so
4. Activate the 'Emergency Call Point (White)' or the 'Manual Call Point (Red)'
5. Evacuate via your closet safe exit
6. Report the emergency to 0-000 & 9351 3333

MEDICAL EMERGENCY

– If a person is seriously ill/injured:

1. **call an ambulance 0-000**

2. **notify the closest Nominated First Aid Officer**

   If unconscious– send for Automated External Defibrillator (AED)
   AED **locations**.

   NEAREST to CS Building (J12)
   - Electrical Engineering Building, L2 (ground) near lifts
   - Seymour Centre, left of box office
        - Carried by all Security Patrol vehicles

3. **call Security -** 9351-3333

4. **Facilitate the arrival of Ambulance Staff (via Security)**

Nearest Medical Facility
University Health Service in Level 3, Wentworth Building

First Aid kit – SIT Building (J12)
kitchen area adjacent to Lab 110

School of Computer Science Safety Contacts

CHIEF WARDEN

Greg Ryan
Level 1W 103
9351 4360
0411 406 322

FIRST AID OFFICERS

Julia Ashworth
Level 2E Reception
8627 9058

Will Calleja
Level 1W 103
9036 9706
0422 001 964

Cecille Faraizi
Level 2E 237
9351 6060

Orally REPORT all INCIDENTS & HAZARDS to your SUPERVISOR

OR

Undergraduates: to Cecille Faraizi
9351 6060

Coursework
Postgraduates: to Julia Ashworth
8627 9058
or Keiko Narushima
8627 0872

CS School
Manager: Priyanka Magotra
8627 4295

Assistance

– There are a wide range of support services available for students: https://sydney.edu.au/campus-life/health-wellbeing-success.html

– Please make contact, and get help
– You are not required to tell anyone else about this
– If you are willing to inform the unit coordinator, they may be able to work with other support to reduce the impact on this unit
  – E.g., provide advice on which tasks are most significant

DISABILITY SERVICES

Do you have a disability?

– You may not think of yourself as having a 'disability' but the definition under the **Disability Discrimination Act** is broad and includes temporary or chronic medical conditions, physical or sensory disabilities, psychological conditions and learning disabilities.

– The types of disabilities we see include:

  – anxiety, arthritis, asthma, asperger's disorder, ADHD, bipolar disorder, broken bones, cancer, cerebral palsy, chronic fatigue syndrome, crohn's disease, cystic fibrosis, depression, diabetes, dyslexia, epilepsy, hearing impairment, learning disability, mobility impairment, multiple sclerosis, post traumatic stress, schizophrenia, vision impairment, and much more.

– Students needing assistance must register with Disability Services –

  – it is advisable to do this as early as possible.

– http://sydney.edu.au/study/academic-support/disability-support.html

Do you have a disability?

You may not think of yourself as having a 'disability' but the definition under the **Disability Discrimination Act (1992)** is broad and includes temporary or chronic medical conditions, physical or sensory disabilities, psychological conditions and learning disabilities.
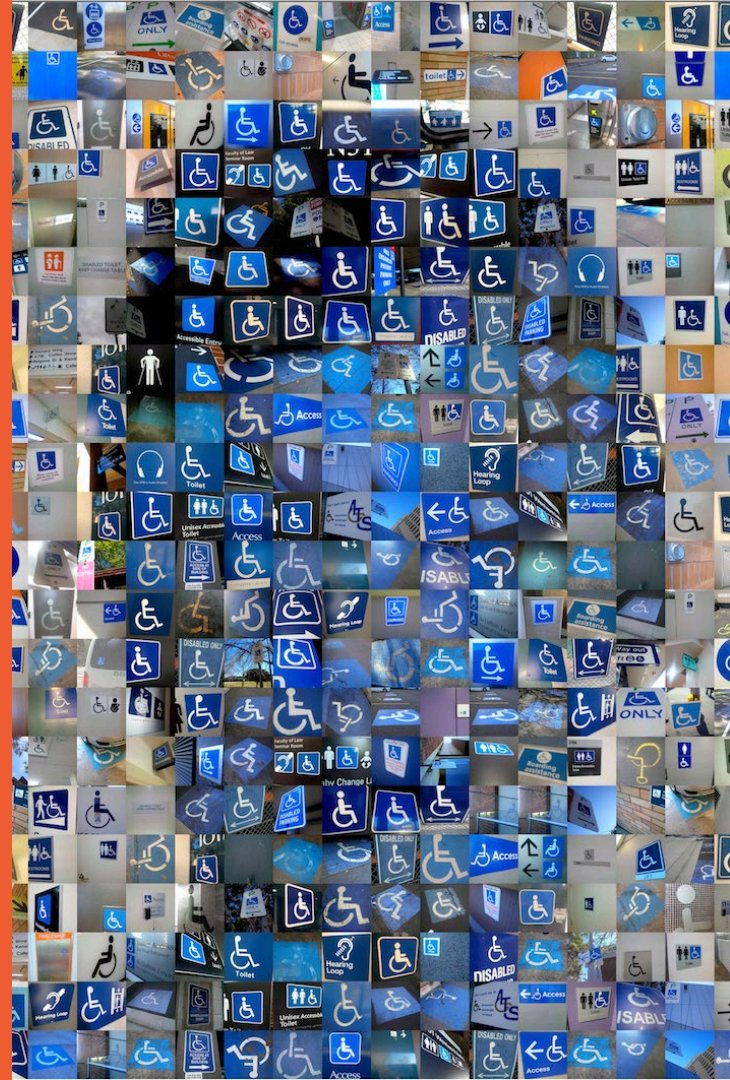
The types of disabilities we see include:
Anxiety // Arthritis // Asthma // Autism // ADHD
Bipolar disorder // Broken bones // Cancer
Cerebral palsy // Chronic fatigue syndrome
Crohn's disease // Cystic fibrosis // Depression
Diabetes // Dyslexia // Epilepsy // Hearing impairment // Learning disability // Mobility impairment // Multiple sclerosis // Post-traumatic stress // Schizophrenia // Vision impairment
and much more.

Students needing assistance must register with Disability Services. It is advisable to do this as early as possible. Please contact us or review our website to find out more.

THE UNIVERSITY OF
SYDNEY

Disability Services Office
sydney.edu.au/disability
02-8627-8422

38

Other Support

– Learning support
  – http://sydney.edu.au/study/academic-support/learning-support.html
– International students
  – http://sydney.edu.au/study/academic-support/support-for-international-students.html
– Aboriginal and Torres Strait Islanders
  – http://sydney.edu.au/study/academic-support/aboriginal-and-torres-strait-islander-support.html
– Student organization (can represent you in academic appeals etc)
  – http://srcusyd.net.au/ or http://www.supra.net.au/
– Please make contact, and get help
– You are not required to tell anyone else about this
– If you are willing to inform the unit coordinator, they may be able to work with other support to reduce the impact on this unit
  – eg provide advice on which tasks are most significant

Advice

- Metacognition
  - Pay attention to the learning outcomes in Canvas
  - Self-check that you are achieving each one
  - Think how each assessment task relates to these
- Time management
  - Watch the due dates
  - Start work early, submit early
- Networking and community-formation
  - Make friends and discuss ideas with them
  - Know your tutor, lecturer, coordinator
  - Keep them informed, especially if you fall behind
    - Don't wait to get help
- Enjoy the learning!

Next Week

- Data Analysis with Python
  - Starting Using Jupyter Notebooks
  - Cf. the following intro slides to Python and Jupyter

Any Questions about the Unit's Organisation?