STATS 769
Overview

Paul Murrell

The University of Auckland

Overview

- This course covers topics related to the analysis of large and/or complex data sets.
- In the first half of the course, we will cover a range of data technologies for accessing and processing data.
- The second half of the course will cover data analysis techniques.

2019-07-19 week 1 overview Paul	2019-07-20 week 1 data science workflow Paul	2019-07-21 week 1 Paul	2019-07-22 week 1 Lab Zero Paul	2019-07-23 week 1 Lab Zero Paul
2019-07-26 week 2 surviving linux Paul	2019-07-27 week 2 prospering in linux Paul	2019-07-28 week 2 Paul	2019-07-29 week 2 Lab 1 Paul	2019-07-30 week 2 Lab 1 Paul
2019-08-02 week 3 data formats Paul	2019-08-03 week 3 web scraping Paul	2019-08-04 week 3 Paul	2019-08-05 week 3 Lab 2 Paul	2019-08-06 week 3 Lab 2 Paul
2019-08-09 week 4 large data problems Paul	2019-08-10 week 4 large data solutions Paul	2019-08-11 week 4 Paul	2019-08-12 week 4 Lab 3 Paul	2019-08-13 week 4 Lab 3 Paul
2019-08-16 week 5 code efficiency Paul	2019-08-17 week 5 parallel code Paul	2019-08-18 week 5 Paul	2019-08-19 week 5 Lab 4 Paul	2019-08-20 week 5 Lab 4 Paul
2019-08-23 week 6 HPC Paul	2019-08-24 week 6 HPC Paul	2019-08-25 week 6 Paul	2019-08-26 week 6 Lab 5 Paul	2019-08-27 week 6 Lab 5 Paul
2019-08-30	2019-08-31	2019-09-01	2019-09-02	2019-09-03
2019-09-06	2019-09-07	2019-09-08	2019-09-09	2019-09-10
2019-09-13 week 7 TBA Yong Wang	2019-09-14 week 7 TBA Yong Wang	2019-09-15 week 7 Yong Wang	2019-09-16 week 7 Lab 6 Yong Wang	2019-09-17 week 7 Lab 6 Yong Wang
2019-09-20 week 8 TBA Yong Wang	2019-09-21 week 8 TBA Yong Wang	2019-09-22 week 8 Yong Wang	2019-09-23 week 8 Lab 7 Yong Wang	2019-09-24 week 8 Lab 7 Yong Wang
2019-09-27 week 9 TBA Yong Wang	2019-09-28 week 9 TBA Yong Wang	2019-09-29 week 9 Yong Wang	2019-09-30 week 9 Lab 8 Yong Wang	2019-10-01 week 9 Lab 8 Yong Wang
2019-10-04 week 10 TBA Yong Wang	2019-10-05 week 10 TBA Yong Wang	2019-10-06 week 10 Yong Wang	2019-10-07 week 10 Lab 9 Yong Wang	2019-10-08 week 10 Lab 9 Yong Wang
2019-10-11 week 11 TBA Yong Wang	2019-10-12 week 11 TBA Yong Wang	2019-10-13 week 11 Yong Wang	2019-10-14 week 11 Lab 10 Yong Wang	2019-10-15 week 11 Lab 10 Yong Wang
2019-10-18 week 12 TBA Yong Wang	2019-10-19 week 12 TBA Yong Wang	2019-10-20 week 12 Yong Wang	2019-10-21 week 12 Yong Wang	2019-10-22 week 12 Yong Wang

Course structure

- Lecture Monday 8-9 303.101
- Lecture Tuesday 8-9 303.102
- Lab Thursday 11-1 302.190
- **Zoom** Lab Thursday 2-4
- Lab Friday 11-1 302.G40

Course delivery

- Lectures are recorded so can be (re)watched later and/or remotely.
- Lecture materials consist of lecture slides and code scripts, plus additional readings; lecture sessions contain live coding demonstrations.
- Lab sessions will provide an opportunity to receive assistance with lab work.
- **In addition to** physical lab sessions, there will be a virtual lab session that occurs online via Zoom.

Course delivery

- We will be working a lot on **virtual machines**.
- The virtual machines run **Linux**.
- The virtual machines will not necessarily have all R packages installed (e.g., **no 'dplyr'** and **no 'ggplot2'**).
- **Your code for labs** will have to work on the virtual machines.

Assessment

In the first half of the course ...

- There are **5 labs** (worth marks).
- There is an **Online Term Test**.

- Each Lab will be worth 3% of your mark (for a total of 15%).
- The Online Term Test will be worth 20% (with **plussage**).

- There will be an open-book, online exam worth 50%.
  You must pass the exam in order to pass the course.

The remaining 15% marks for course work will be assessed in the second half of the course.

Assessment

- Handouts for each lab will be made available at the start of the week.
- Lab attendance is not compulsory, but is the best place to receive help and feedback.
- The lab submission is **online** and will be due on **the following Monday** (consult Canvas for exact submission date-times).
- The lab in the first week is worth **zero marks** and is just to make sure that we can all use the virtual machines and to get practice with the lab submission process and format.

Getting assistance

- Lecture sessions and lab sessions provide opportunities to ask questions and receive feedback.
- The best place to ask questions online is via Piazza (on Canvas)

Class Rep

- Volunteer(s) required !