

Overview

Big Data has gained significant attention within the last decade. It permeates many aspects of everyday life, from making online searches to choosing a restaurant and shopping for groceries. Big Data has had implications for science, technology, business and even the humanities and social sciences. But what really is Big Data? Does it just mean, “having a lot of data”? Or can we reach a more precise definition to help us understand the uses of Big Data, the implications it has and its relation to the great data controversies of the century.

This course will look at the epistemology and central methods of Big Data. The focus is on digital communication and technologies that impact how society is organized. The content is both theoretical and practical, first delving into the concept of Big Data, including how to conceive data-driven approaches and potential solutions, and then exploring simple ways to use the abundance of digital data sources to analyze, visualize and tell stories with data. Part of the course also involves knowing the limitations of Big Data, critically assessing data-driven methods and gaining familiarity with current concerns about ethics, surveillance and privacy.

The target audience includes anyone who is interested in an introduction to Big Data both as a concept and a research method. No prior technical skills are required, though an interest in working with data, software and simple statistics is an advantage.

Learning Outcomes

By the end of this course, students will have:

1. Familiarity with the latest theories about Big Data including how Big Data methods are different from traditional methods, and be able to see the possibilities in a variety of data sources and conceive data-driven solutions.

2. Knowledge to critically assess the employment of Big Data methods and what implications they may have for society in general.
3. The ability to practically handle different digital data sources and tell stories with data visualizations.

Teaching staff

COURSE COORDINATOR

Dr Christopher Thomson
Karl Popper Rm 604
Email: christopher.thomson@canterbury.ac.nz

LECTURER

Jakob Kristensen
Email: jakob.kristensen@pg.canterbury.ac.nz
Office hour: Wednesday 3:00 - 5:00PM and Thursday 3:00 - 5:00PM

Course Delivery and Workload

Students are expected to attend all lectures and workshops in order to complete the course successfully. DIGI210 is a 15 point, 200 level course. You should expect to spend approximately 150 hours to complete the course, broken down as follows:

Activity	Time
Classes	36 hrs
Reading / Preparation / Revision	36 hrs
Quiz	8 hrs
Presentation	20 hrs
Test	20 hrs
Final essay	30 hrs
Total	150 hrs

Class times

Please see the [Course Information Page](#) for class times and locations. Please note these can be subject to change and it is recommended that you check before each class in the first couple of weeks of term.

Activity	Date	Time	Duration	Location
<u>Week 1</u>				
Lecture 1	19-Nov-18	10:00AM - 11:45AM	2 Hours	TBA
Lecture 2	20-Nov-18	10:00AM - 11:45AM	2 Hours	TBA
Lecture 3	21-Nov-18	10:00AM - 11:45AM	2 Hours	TBA
Lecture 4	22-Nov-18	10:00AM - 11:45AM	2 Hours	TBA
Tutorial 1	20-Nov-18	1:00PM - 2:00PM	1 Hour	TBA
Tutorial 2	21-Nov-18	1:00PM - 2:00PM	1 Hour	TBA
Tutorial 3	22-Nov-18	1:00PM - 2:00PM	1 Hour	TBA
<u>Week 2</u>				
Lecture 5	26-Nov-18	10:00AM - 11:45AM	2 Hours	TBA
Lecture 6	27-Nov-18	10:00AM - 11:45AM	2 Hours	TBA
Lecture 7	29-Nov-18	10:00AM - 11:45AM	2 Hours	TBA
Lab 1	28-Nov-18	1:00PM - 3:00PM	2 Hours	TBA
Lab 2	29-Nov-18	1:00PM - 3:00PM	2 Hours	TBA
Lab 3	30-Nov-18	1:00PM - 3:00PM	2 Hours	TBA
Workshop 1	27-Nov-18	1:00PM - 5:00PM	4 Hours	TBA
<u>Week 3</u>				
Lecture 5	4-Dec-18	10:00AM - 11:45AM	2 Hours	TBA
Lecture 6	5-Dec-18	10:00AM - 11:45AM	2 Hours	TBA
Lecture 7	6-Dec-18	10:00AM - 11:45AM	2 Hours	TBA
Lab 4	3-Dec-18	1:00PM - 3:00PM	2 Hours	TBA
Tutorial 4	4-Dec-18	1:00PM - 2:00PM	1 Hour	TBA
Tutorial 5	5-Dec-18	1:00PM - 2:00PM	1 Hour	TBA

Learn

All course information will be available to you via Learn, the University's eLearning platform. This will include assigned readings, which are essential preparation for each week's classes and must be read in advance.

You will submit your assignments through Learn. These must be presented in a .docx or.rtf format. Portable Document Format (PDF) is not acceptable, and we may ask you to resubmit work if it is in this format. All assessed work will be checked by the Turnitin plagiarism detection software when you upload it to Learn.

Assessment

Assessed work for this course will be as follows:

ITEM	VALUE	DUE DATE	LEARNING OUTCOMES
Quiz	10%	End of first week	LO 1
Group Presentation	20%	End of second week	LO 2
In-class test	30%	End of third week	LO 1, 3
2000 word Essay	40%	End of fifth week	LO 1, 2, 3

Course Expectations

Lectures

- Attend, listen and take your own notes
- Read the required readings before each lecture to fully appreciate the content
- Be respectful of others: arrive on time, stay until end, minimize disruptions

Tutorials

- Attend and participate - tutorials involve engaging learning environments with group-based activities such as small group and class discussions
- Read the required readings before each tutorial (same as the lecture required readings for that day). The focus of the tutorial will usually be the same as the lecture it follows

Labs

- Attend and engage – Labs involve doing fixed exercises and also provide a space for working on a group project of own choosing.

Workshop

- During the second week of the course a workshop will be held with the purpose of introducing Tableau, the main data analysis software. The aim of the workshop is for everyone to get familiar with the standards for preparing data, importing it to the software and understanding basic operations.

Essay Format:

Your essay should:

- be in Microsoft Word (when submitted to Learn)
- include a cover page with your name, title and word count
- include your name and essay question number in the title of the electronic file (e.g. save as “Jane Doe – {title}”)
- use double-line spacing
- be in 12-point font
- number pages

Essay Marking Criteria

The essay assesses your research, writing, organization, and critical thinking skills. The essays will be marked based on the following criteria:

STRUCTURE	Is the essay clear, coherent and well organized?
COVERAGE	Does the essay cover the main theories in the literature that are relevant for the question?
ARGUMENTATION	Does the essay incorporate the key arguments that are relevant for the question, while also presenting a clear thesis?

ANALYSIS	Does the essay offer a detailed examination of the arguments put forward, separating arguments into their constituent elements?
EVIDENCE	Are the arguments well supported with various forms of evidence?
ORIGINALITY	Does the essay show evidence of original thinking?
EVIDENCE OF READING	Is there evidence of engagement with the relevant required readings and at least three other relevant readings found by the student?
PRESENTATION/STYLE	Is the essay well written and presented?
SPELLING AND GRAMMAR	Is the essay free of errors?
CITATION	Are sources correctly and consistently cited (using Harvard style)? Is there a proper bibliography (using Harvard style)?

Schedule and Readings

Lecture 1: What is Big Data? Definitions and Epistemology

This Lecture covers some of the most popular and influential definitions, which have been proposed during the last 10 years. Why are people talking about Big Data now? What are the methodological implications? What does “the end of theory” mean? The point of the lecture is not to define Big Data, but gain familiarity with some of the most common characteristics.

Required readings:

- Anderson, C., 2008. The end of theory: the data deluge makes the scientific method obsolete. *Wired Magazine* 16.07.
- Kitchin, R., 2014. *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage. - Chapter 4

- Kitchin, R., 2014. Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), p.2053951714528481.
- Mayer-Schönberger, V. and Cukier, K., 2014. Big Data: A Revolution That Will Transform How We Live, Work, and Think. – Chapter 1

Optional readings:

- Lagoze, C., 2014. Big Data, data integrity, and the fracturing of the control zone. *Big Data & Society*, 1(2), p.2053951714558281.
- Burns, R., 2015. Rethinking big data in digital humanitarianism: Practices, epistemologies, and social relations. *GeoJournal*, 80(4), pp.477-490.
- Kitchin, R. and McArdle, G., 2016. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3(1), p.2053951716631130.

Lecture 2: Datafication, Correlation and Data-driven Methods

We delve further into how data is appropriated and used for research and why the practices are changing in the Big Data age. The main purpose is to contrast data-driven methods and theory-driven methods.

Required readings:

- Kitchin, R., 2014. *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage. – Chapter 1
- Madsen, A.K., 2015. Between technical features and analytic capabilities: Charting a relational affordance space for digital social analytics. *Big Data & Society*, 2(1), p.2053951714568727.
- Qiu, L., Chan, S.H.M. and Chan, D., 2018. Big data in social and psychological science: theoretical and methodological issues. *Journal of Computational Social Science*, 1(1), pp.59-66.

Optional readings:

- Mayer-Schönberger, V. and Cukier, K., 2014. Big Data: A Revolution That Will Transform How We Live, Work, and Think. – Chapter 5
- Cows, J. and Schroeder, R., 2015. Causation, correlation, and big data in social science research. *Policy & Internet*, 7(4), pp.447-472.
- Holtzhausen, D., 2016. Datafication: threat or opportunity for communication in the public sphere?. *Journal of Communication Management*, 20(1), pp.21-36.
- Mahmoodi, J., Leckelt, M., van Zalk, M.W., Geukes, K. and Back, M.D., 2017. Big Data approaches in social and behavioral science: four key trade-offs and a call for integration. *Current Opinion in Behavioral Sciences*, 18, pp.57-62.

Lecture 3: Sources, Technologies and Applications

This lecture is about how Big Data is collected, with what purposes and how it is applied in both science and business. The intention is to make a connection between the theoretical concepts and the more physically grounded infrastructures that make up the Big Data environment. The lecture covers topics such as open data initiatives, increased computing power, cheap sensor technology, application programming interfaces, machine learning and AI.

Required readings:

- Kitchin, R., 2014. *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage. - Chapter 5
- Clegg, B., 2018. *Big Data: How the Information Revolution Is Transforming Our Lives*. Icon Books Ltd. - Chapter 3

Optional readings:

- Pentland, A., 2014. *Social Physics: How Good Ideas Spread - The Lessons From a New Science*. The Penguin Press. Chapter 8 - 9
- Géron, A., 2017. *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. "O'Reilly Media, Inc.". Chapter 1
- Murthy, D. and Bowman, S.A., 2014. Big Data solutions on a small scale: Evaluating accessible high-performance computing for social research. *Big Data & Society*, 1(2), p.2053951714559105.
- Sormanen, N., Rohila, J., Lauk, E., Uskali, T., Jouhki, J. and Penttinen, M., 2016. Chances and challenges of computational data gathering and analysis: The case of issue-attention cycles on Facebook. *Digital Journalism*, 4(1), pp.55-74.

Lecture 4: Critical Questions for Big Data Methods

Big Data make a lot of promises, but not all of them are necessarily fulfilled. When taking advantage of Big Data it is important to remain critical about the interpretations that can be derived, especially when they are being used to make claims about society.

Required readings:

- Boyd, D. and Crawford, K., 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5), pp.662-679.

Optional readings:

- Freelon, D., 2014. On the interpretation of digital trace data in communication and social computing research. *Journal of Broadcasting & Electronic Media*, 58(1), pp.59-75.
- Couldry, N. and Powell, A., 2014. Big data from the bottom up. *Big Data & Society*, 1(2), p.2053951714539277.
- Bruns, A., 2013. Faster than the speed of print: Reconciling 'big data' social media analysis and academic scholarship. *First Monday*, 18(10).
- Boellstorff, T., 2013. Making big data, in theory.
- Kitchin, R. and Lauriault, T., 2014. Towards critical data studies: Charting and unpacking data assemblages and their work.
- Curini, L., Ceron, A. and Iacus, S.M., 2016. *Politics and big data: Nowcasting and forecasting elections with social media*. Routledge. Chapter 6

Lecture 5: Exploring and Visualizing Big Data

This lecture is about getting the most out of one's data by exploring and choosing the right visualizations. Emphasis is put on exploratory data analysis in contrast to confirmatory data analysis while also taking into account the limitations of exploratory data analysis. Additionally the sessions includes a review of some basic visualization techniques and how they have been powerfully employed in real life cases.

Required readings:

- Aiden, E. and Michel, J.B., 2013. *Uncharted: Big Data as a Lens on Human Culture* (Riverhead, New York). Chapter 1 - 2
- Tufte, E. and Graves-Morris, P., 2014. *The visual display of quantitative information.*; 1983. Chapter 1
- Hartwig, F. and Dearing, B.E., 1979. *Exploratory data analysis*(Vol. 16). Sage. Chapter 1

Optional readings:

- Kennedy, H., Hill, R.L., Aiello, G. and Allen, W., 2016. The work that visualisation conventions do. *Information, Communication & Society*, 19(6), pp.715-735.
- Rogers, R., 2015. Digital methods for web research. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, pp.1-22.
- O'Neil, C. and Schutt, R., 2013. *Doing data science: Straight talk from the frontline*. " O'Reilly Media, Inc.". Chapter 2

Lecture 6: Data Structures and Data Types

We get into the more gritty hands-on details of actually handling and working with data. Data formats and data types such as integers, strings, floating points, datetimes and geographical data and their uses are covered.

Required readings:

- <https://data-flair.training/blogs/tableau-data-types/>

Lecture 7: From Graphs to Data Narratives

Putting your best data on a chart rarely delivers the answer you want. Real world data often have many facets and it is therefore necessary to combine visualizations that highlight certain aspects in a logical manner in order to draw out useful insights.

Required readings:

- Stolper, C.D., Lee, B., Riche, N.H. and Stasko, J., 2016. Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories. *Microsoft Research, April, 3*, p.2016.

Optional readings:

- Gratzl, S., Lex, A., Gehlenborg, N., Cosgrove, N. and Streit, M., 2016, June. From visual exploration to storytelling and back again. In *Computer Graphics Forum* (Vol. 35, No. 3, pp. 491-500).
- Scope, S., More than Telling a Story: A Closer Look at the Process of Transforming Data into Visually Shared Stories.

Lecture 8: Big Connected Data - Politics, social media and algorithms

We leave digital footprints everywhere. It happens when we seek out information and interact with other people, which are done more and more using online technologies. This lecture takes a look at how Big Data is being leveraged specifically in the intersection between politics, media and our everyday lives. The topics span across political prediction models, social media algorithms and filter bubbles.

Required readings:

- Curini, L., Ceron, A. and Iacus, S.M., 2016. *Politics and big data: Nowcasting and forecasting elections with social media*. Routledge. Chapter 1

- Kristensen, J.B., Albrechtsen, T., Dahl-Nielsen, E., Jensen, M., Skovrind, M. and Bornakke, T., 2017. Parsimonious data: How a single Facebook like predicts voting behavior in multiparty systems. *PloS one*, 12(9), p.e0184562.
- Thorson, K. and Wells, C., 2015. Curated flows: A framework for mapping media exposure in the digital age. *Communication Theory*, 26(3), pp.309-328.

Optional readings:

- Bakshy, E., Rosenn, I., Marlow, C. and Adamic, L., 2012, April. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web* (pp. 519-528). ACM.
- Jensen, K.B., 2013. How to do things with data: Meta-data, meta-media, and meta-communication. *First Monday*, 18(10).
- Hansen, P.G., Hendricks, V.F. and Rendsvig, R.K., 2013. Infostorms. *Metaphilosophy*, 44(3), pp.301-326.

Lecture 9: Big Data Controversies - Ethics, Privacy, Misinformation and Surveillance

Big Data has been at the center of many of the biggest controversies of the decade. The purpose of this lecture is to discuss the controversies that surround cases concerning data leaks, information manipulation and privacy breaches.

Required readings:

- Lyon, D., 2014. Surveillance, Snowden, and big data: Capacities, consequences, critique. *Big Data & Society*, 1(2), p.2053951714541861.
- Panger, G., 2016. Reassessing the Facebook experiment: critical thinking about the validity of Big Data research. *Information, Communication & Society*, 19(8), pp.1108-1126.
- Vosoughi, S., Roy, D. and Aral, S., 2018. The spread of true and false news online. *Science*, 359(6380), pp.1146-1151.

Optional readings:

- Schroeder, R., 2014. Big Data and the brave new world of social media research. *Big Data & Society*, 1(2), p.2053951714563194.
- Kramer, A.D., Guillory, J.E. and Hancock, J.T., 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, p.201320040.

Lecture 10: Recap

This lecture will review some of the main points of the previous lectures and address the expectations for the final essay. There will also be extra time for questions and discussion.

Optional readings:

- Clegg, B., 2018. *Big Data: How the Information Revolution Is Transforming Our Lives*. Icon Books Ltd. – Chapter 7

Dishonest practice

Students found to be engaging in dishonest practices will be dealt with under the University Discipline Regulations. Penalties range from refusal to grade the assignment in question, to exclusion from the University. The Digital Humanities Programme understands ‘dishonest practice’ to include:

PLAGIARISM:

The presentation of any material without adequate acknowledgement of the source. This includes text, data, figures, computer code etc, on any medium (print or online). All sources should be adequately referenced, following accepted scholarly standards. Refer to the latest edition of the MLA Style Manual for more information.

COLLUSION:

Working with another person or persons and presenting it as if it was your work alone. While collaboration and discussion is important in your learning, students need to be careful about acknowledging everyone who has had a significant involvement in producing their assessed work.

GHOST WRITING:

Using another person to write an essay or assignment for you. This applies whether they are paid or not.

Disputes Procedures

Though rare, disputes between students and staff do occur. Should there be a problem, attempt first to resolve it by discussion with the course coordinator, Dr Christopher Thomson. You may wish to make initial contact through and be supported by

the student rep. Should there still be matters at issue, your next recourse would be to the Head of School, Professor Paul Millar. Outside the School, help can be sought from any member of the Academic Grievance Committee. You can also approach the UCSA's [Advocacy and Welfare team](#).

FAQs

Where do I get course information (e.g. lecture slides)?

This course uses Learn for sharing course information. This is the university's web-based system for course content delivery. You'll need to log on using your usercode and password, and should see each course you're enrolled in listed there. Please make sure to learn how to access and use this system. If you have problems, please contact the IT Help Desk, (ext 6060 or helpdesk@it.canterbury.ac.nz).

How do I learn more about writing essays?

The lecturer will discuss the specific requirements of the essay for this course. However, if you think you need further assistance with the skill required to write an effective essay, it is recommended that you enrol in a course run by the Academic Skills Centre (at the UCSA Building). More details at:

<http://www.uco.canterbury.ac.nz/lsc/index.php>

Where do I hand in assignments?

An electronic copy of your essay must be submitted to Learn for marking before the date and time it is due. No paper copy is required. There will be a link on the Learn home page where you can upload your essay. This link will be available at least one week before the due date. This course uses Turnitin to check for academic integrity (see below) and all essays must initially be sent to Turnitin (via Learn) before they are submitted for marking. Please submit essays in Microsoft Word format, as marking is done using Word's Review function. If you do not have a copy of Word, it is available on most UC computers. Also, most word processing programmes allow you to save in Word, so please save and submit your final document in this format.

How does Turnitin work?

Before your essay is sent for marking on Learn, it must be submitted to Turnitin (through the same Learn link), which will produce a score showing the percentage of your essay that overlaps with other existing sources, such as published material, online sources and other essays. Review these overlaps carefully to ensure you do not inadvertently plagiarise. You don't need to worry about the score as long as everything is referenced correctly (see relevant section below). Also, please note that it may take up to one hour for Turnitin to return a score. Make sure you leave enough time between receiving your Turnitin score and the deadline to review and edit your work before resubmitting.

What if my essay is going to be late?

Late essays will receive a penalty of 2% per day without an approved extension, up to a maximum of 5 days. After 5 days, essays will no longer be accepted and will receive a mark of zero. If you experience an unforeseen event or circumstance that prevents you from completing your essay on time, you may apply for an extension. ToFAQs apply, please email the lecturer before the deadline and request an essay extension. Please be aware, though, that grounds for an extension are narrow: sickness (with a medical certificate), bereavement, or some other serious matter that's prevented you doing your work. Please start your essays early to avoid problems.

Useful Links

Computer and technical support

Using Learn - student help

<http://learn.canterbury.ac.nz/course/view.php?id=2157&home=1>

Information and Communication Technology Services

<http://www.icts.canterbury.ac.nz/for/students.shtml>

Forum and email etiquette and best practices

<http://forums.adobe.com/thread/414764>

<http://www.101emailtippettips.com/>

Academic support

Academic Skills Centre

<http://www.lps.canterbury.ac.nz/lsc/>

Key Dates for Course Changes

[2018 Guide to Enrolment](#) - see Guide to Enrolment PDF, page 5

Special Consideration

<http://www.canterbury.ac.nz/study/special-consideration/>

Scholarships

<http://www.canterbury.ac.nz/scholarships/>

Academic Progress Reviews

<http://www.canterbury.ac.nz/support/academic/progress-reviews/>

Personal support

Student Support

<http://www.canterbury.ac.nz/support/>

Student Counselling Services (Health Centre)

<http://www.canterbury.ac.nz/healthcentre/counselling.shtml>

Student Services and Communications

<http://www.canterbury.ac.nz/ssac/>

Disability Resource Service

<http://www.canterbury.ac.nz/disability/>

Māori Students

<http://www.canterbury.ac.nz/support/akonga-maori/>

Pasifika Students

<http://www.canterbury.ac.nz/support/pasifika/>

International Students

<http://www.canterbury.ac.nz/international/>