

CIS4020Fall 2020

Data Science

contact

Dr. Daniel Gillis
Pronouns: he/him/his

dgillis@uoguelph.ca
<http://danielgillis.ca>
[twitter:@drdanielgillis](https://twitter.com/drdanielgillis)

dear students

This semester is going to be - well - different. But different doesn't mean that it's going to be bad. It does mean, however, that it is absolutely necessary that we take care of ourselves, our families and friends, and our community.

Many people, including me, are feeling anxious about the pandemic and the seemingly endless list of awful things we are seeing in the news. Many of us are trying to navigate this reality at the same time as managing other significant challenges, such as caring for children or aging parents, job loss, or financial uncertainty. Mix in daily news of racism, civil rights violations, misogyny, homophobia, transphobia, "murder hornets", climate change, biodiversity loss, and more - it's not surprising that we're feeling uneasy.

With that in mind, as we progress through the semester I want you to take care of yourself first. To paraphrase Dr. Max Liboiron, if you are sick, heartbroken, or exhausted - please do whatever you can to get the help you need. Reach out to a friend, or a family member, or a health professional. Your health and wellness are more important than your work. If you are worried how doing what is right for your health and wellness might impact your ability to succeed in CIS4020, please chat with me. You don't need to let me know what the problem is, but I can work with you to figure out a solution that will help you succeed in this course.

If you, your family, or people you live with are showing symptoms of COVID-19, please follow the guidance of our public health professionals. If you can't work or study at this time because you are too sick or because you need to take care of someone else, please let me know as soon as possible. I will work with you to determine the best path through the course.

If you are living in a different time zone or have limited internet, please reach out to me as soon as possible to let me know how this might affect your ability to attend class or to work on a team. I will work with you to make sure you have the materials you need to complete the course.

Finally, the semester may prove to be challenging so let's do our best to practice empathy and patience for ourselves and for others.

Having said all this, I do believe that this semester is going to be amazing. I've done my best to restructure the course for online delivery in a way that is supportive and flexible. Will it be perfect? No. Will I mess something up? Probably. Will we experience some growing pains and bumps along the way? Undoubtedly. But whatever happens, be reassured that I will do my best to make the fall semester as positive and successful as possible.

Sincerely,



Dan Gillis

Click here for the full letter. See the course website for information on code of conduct, academic integrity, accessibility, and our policy on children in the classroom.

CIS4020Fall 2020

Data Science

contact

Dr. Daniel Gillis
Pronouns: he/him/his

dgillis@uoguelph.ca
<http://danielgillis.ca>
[twitter:@drdanielgillis](https://twitter.com/drdanielgillis)

where we learn is important

The Dish With One Spoon Covenant speaks to our collective responsibility to steward and sustain the land and environment in which we live and work, so that all peoples, present and future, may benefit from the sustenance it provides. As we continue to strive to strengthen our relationships with and continue to learn from our Indigenous neighbours, we recognize the partnerships and knowledge that have guided the learning and research conducted in and for this class. We acknowledge that the University of Guelph resides in the ancestral and treaty lands of several Indigenous peoples, including the Attawandaron people and the Mississaugas of the Credit, and we recognize and honour our Anishinaabe, Haudenosaunee, and Métis neighbours. We acknowledge that the work we do here occurs on their traditional lands so that we might work to build lasting partnerships that respect, honour, and value the culture, traditions, and wisdom of those who have lived here since time immemorial.

what is this course about?

Data Science focuses on extracting the important relations in data. The course is intended as a survey of the discipline and focuses on applied computational methods for data analysis. Topics include algorithms, computational and machine learning methods, software tools, and modeling, as they apply to the analysis of and discovery in big data.

what are we going to learn?

By the end of this course, you should be able to:

- 1 **Explain the primary concepts and tools used in data analysis.**
- 2 **Create systems which analyze, capture, and format large data sets.**
- 3 **Apply the appropriate algorithms for analyzing different types of data.**
- 4 **Describe different data types and their characteristics.**
- 5 **Describe the different statistical and learning algorithms for data analysis.**
- 6 **Create software which implements analysis algorithms.**
- 7 **Integrate existing software tools and libraries into an analysis system.**

CIS4020Fall 2020

Data Science

contact

Dr. Daniel Gillis
Pronouns: he/him/his

dgillis@uoguelph.ca
<http://danielgillis.ca>
twitter:@drdanielgillis

when are we getting together?

Due to the COVID-19 pandemic, we will meet online for our lectures and labs. While I'd much prefer to have everyone in class, it just doesn't feel safe to do this.

While I want you to attend every class and every lab, I realize that there may be times when this won't be possible. Please discuss any challenges you might have with me, and I'll do my best to help you sort them. To accommodate students who might be in different time zones, our lectures will be recorded and made available to you as soon as possible after each class.

To ensure our time is as productive as possible, I'm going to ask that you do some work in advance of class. This typically involves reading, but might involve sketching, watching videos, or creating interpretive dance routines to demonstrate certain data science topics.

Lecture	Tues/Thurs 8:30 am - 9:50 am	online
Lab	Mon 8:30 am - 9:20 am	online

what skills should we have before we take this course?

Students enrolled in CIS4020 are expected to have the skills and knowledge covered in CIS2750, STAT2040, and MATH1160, as well as strong writing skills, the ability to work in teams, and strong communication skills.

Prerequisite	Software Systems Development and Integration	CIS2750
Prerequisite	Statistics I	STAT2040
Prerequisite	Linear Algebra I	MATH1160

what textbooks will we be reading

All reading materials will be provided, but if you want to dive further into any of the topics, feel free to check out these references:

Suggested	Doing Data Science <i>Cathie O'Neil & Rachel Schutt</i>	2014
Suggested	Data Visualization: A Practical Introduction <i>Kieran Healy</i>	2018
Suggested	Statistics for Data Science <i>James D. Miller</i>	2017
Suggested	Statistics for Machine Learning <i>Pratap Dangeti</i>	2017

CIS4020Fall 2020

Data Science

contact

Dr. Daniel Gillis
Pronouns: he/him/his

dgillis@uoguelph.ca
<http://danielgillis.ca>
twitter:@drdanielgillis

where can we find course materials?

There are a lot of things to cover in this course! But never fear - all course material, news, announcements, and grades will be regularly posted on our CIS4020 website. You can find it here. Please be sure to check the website regularly. And while other tools (such as Slack, Discord, Trello, etc.) might be used and are super helpful in their own right, definitely refer to the course website for course information, or ask the teaching team if you are still unsure.

Community Engagement Materials

This year you will be tasked with developing a data science dashboard that supports issues related to social justice and/or public health. With that in mind, you will be required to work with a community partner to ensure you are answering questions using data science methods that are relevant to our community, that cause no harm, and that can help make a difference. To build a relationship with your community partner, you will find materials describing community-engaged projects on our website. It's important for all of us to review these materials before we begin working with a community partner. And remember that community partners are volunteering their time to support us as we learn about data science.

Challenge Materials

In addition to the regular course materials, I will post extra articles or other data about social justice and/or public health issues. Please use these materials to bring yourself up to speed on any given challenge, as this will help you build a better solution and build a rapport with your community partner.

Lecture Materials

Lectures will be presented synchronously. Lectures will be recorded, however, we as a class will determine which parts of class will be included in those recordings. This is to protect student privacy. Recordings will be made available online as soon as possible. Slides will also be made available.

Labs

Labs will be presented asynchronously; however, we will set up time for each team to meet with a TA to clarify issues or to receive advice pertaining to their project. Some labs will require that you or your team submit work for grading. You will have at least a week to complete and submit the materials. Lab materials will be made available through the course website.

Assignments

Assignment descriptions and rubrics will be made available on the course website. You should submit all of your assignments through the course website.

Quizzes

Instead of a final exam, the course has 3 short quizzes. Each quiz will be available on the website for approximately 2 weeks. If you have any issues or need extra time to complete a quiz, please let me know as soon as possible.

CIS4020Fall 2020

Data Science

contact

Dr. Daniel Gillis
Pronouns: he/him/his

dgillis@uoguelph.ca
<http://danielgillis.ca>
twitter:@drdanielgillis

what types of assessments are there?

The course has been broken down into a bunch of small deliverables. Each deliverable has a due date range instead of a single due date. If you are having difficulty with an assignment, please chat with the teaching team as soon as possible. Additionally, if you need more time to get something done, please chat with me sooner rather than later.

Assignments

1. Due between October 1st at 8:30 am & October 7th at 4:00 pm [10%]
2. Due between October 29th at 8:30 am & November 4th at 4:00 pm [10%]
3. Due between November 26th at 8:30 am & December 3rd at 10:00 am [10%]
4. Due December 4th at 4:00pm [15%]

All assignments are to be completed as a team. Assignment 3 is an in-class team presentation. Assignment 4 will also include an individual contribution.

Quizzes

1. Due between September 23rd at 8:30 am & October 6th at 4:00 pm [10%]
2. Due between October 21st at 8:30 am & November 3rd at 4:00 pm [10%]
3. Due between November 11th at 8:30 am & November 24th at 4:00 pm [10%]

You will have approximately 2 weeks to complete a quiz. Quizzes are not to be completed as a team.

Labs

1. Due between September 21st at 8:30 am & September 28th at 4:00 pm [5%]
2. Due between September 28th at 8:30 am & October 5th at 4:00 pm [5%]
3. Due between October 19th at 8:30 am & October 26th at 4:00 pm [5%]
4. Due between November 2nd at 8:30 am & November 9th at 4:00 pm [5%]
5. Due between November 9th at 8:30 am & November 16th at 4:00 pm [5%]

While there are 8 labs, you will need to submit modules to be graded for only 5 of them. Some of the modules will be completed by you and your team, others will be completed by you.

CIS4020Fall 2020

Data Science

contact

Dr. Daniel Gillis
Pronouns: he/him/his

dgillis@uoguelph.ca
<http://danielgillis.ca>
[twitter:@drdanielgillis](https://twitter.com/drdanielgillis)

what happens if something is missed or late?

Missed Labs: if you are going to miss a lab, please let your team know (and if possible - me too). Together we should be able to work around your absence.

Missed Assessments: if you can't complete an assignment or quiz or lab due to medical, psychological, or compassionate reasons, please chat with me.

Accommodation of Religious Obligations: if you are unable to meet an in-course requirement due to religious obligations, please let me know **within two weeks of the start of the semester** to make alternate arrangements.

Late Assignments/Lab Demos/Quizzes: we will begin grading course deliverables shortly after the last moment they are due. If you have not submitted a course deliverable on time, it will be considered late. You may submit any course deliverable late, however, your final grade for that deliverable will be reduced by 10% per day. If there are extenuating circumstances that could explain why you need more than three days extension, let me know before the last due date of the deliverable. We will work to identify a new due date, and the expected penalty for the late submission.

Regrades: if you feel your assignment has been graded incorrectly, please present your case (via email) to the instructor. Be specific about what you believe was graded incorrectly. All regraded material will be *completely* regraded. This could result in your grade being reduced.

CIS4020Fall 2020

Data Science

contact

Dr. Daniel Gillis
Pronouns: he/him/his

dgillis@uoguelph.ca
<http://danielgillis.ca>
[twitter:@drdanielgillis](https://twitter.com/drdanielgillis)

what will we learn during lectures?

WEEK	TOPICS COVERED†
Intro	Intro to Data Science & Biases
Week 1	The Data Science Process Exploring Data
Week 2	Preliminary Data Analysis Statistical Distributions
Week 3	Hypothesis Testing Confidence Intervals
Week 4	Simple Linear Regression Multiple Linear Regression
Week 5	Logistic Regression
Week 6	Poisson Regression K Nearest Neighbours
Week 7	Naive Bayes Classifier Support Vector Machines
Week 8	Decision Trees Neural Networks
Week 9	K Means Mean-Shift Clustering
Week 10	Gaussian Mixture Models Agglomerative Hierarchical Clustering
Week 11	Density-Based Spatial Clustering In Class Presentations
Week 12	In Class Presentations

CIS4020Fall 2020

Data Science

contact

Dr. Daniel Gillis
Pronouns: he/him/his

dgillis@uoguelph.ca
<http://danielgillis.ca>
[twitter:@drdanielgillis](https://twitter.com/drdanielgillis)

what will we learn during labs?

WEEK	TOPICS COVERED†
Week 0	No Lab
Week 1	Ethics & Do No Harm
Week 2	Literature Reviews
Week 3	Types of Data & Simulations
Week 4	Data Visualization
Week 5	No Lab
Week 6	Dashboards in R & Python
Week 7	Science Communication
Week 8	Critiquing Science Communication
Week 9	Critical Reviews
Week 10	No Lab
Week 11	No Lab
Week 12	No Lab

CIS4020Fall 2020

Data Science

contact

Dr. Daniel Gillis
Pronouns: he/him/his

dgillis@uoguelph.ca
<http://danielgillis.ca>
twitter:@drdanielgillis

what if we aren't feeling well?

While health and wellness is always important, it is particularly important this semester. As previously stated, I want you to put yourself first this semester. We need to do whatever we can to support each other, as well as our family, friends, and community. With that in mind, we need to work together, practice patience and empathy, and remain honest about our needs. Only then can we foster and promote a safe, supportive environment, as well as good physical, emotional, spiritual, cultural, and mental health and wellness for everyone.

If you are sick, heartbroken, or exhausted, get rest, reach out to someone, take whatever steps necessary to get well. Work is not more important than your health. -Dr. Max Liboiron

If you are experiencing any challenges, please do not hesitate to contact me, and know that there are resources on campus set up to help you out.

Medical concerns? Student Health Services at x52131

Threats of violence, personal safety? Campus police at x2000

Psychological or emotional concerns? Counselling services at x53244

Accessibility concerns? SAS at x56208

Sexual assault? Campus police at x2000, or counselling services at x53244

Mental Health concerns? Please see the Mental Health Resources page here.

Other sources of help can be found at the following links:

- Student Health Services, Monday to Friday, 8:30am-4:30pm, x52131, J.T. Powell Building
- Counselling Services, Monday to Friday, 8:15am-4:15pm, x53244, Level 3, University Centre
- Wellness Education Centre, Monday to Friday, 8:30am-4:30pm, x53327, J.T. Powell Building
- Student Support Network, Monday to Friday, 12:00pm-10:00pm, Raithby House
- Campus Community Police, 24/7, x2000, Trent Building
- Good2Talk, 1.866.925.5454
- Here 24/7, 1.844.437.3427