Teaching Dossier

Julia Gallucci, HBSc, MSc, PhD (c)
Institute of Medical Science, University of Toronto

Contents

Statement of Teaching Philosophy

My teaching philosophy is shaped by my interdisciplinary background and experiences as a student, teaching assistant, course instructor, and curriculum developer in biostatistics, data science, and psychology/neuroscience at the University of Toronto. I recently co-led UNICE (Undergraduate Neuroscience Initiative for Curriculum Enhancement), a project integrating foundational coding skills into neuroscience courses, and presented this work at the Open Science in Undergraduate Education Symposium; an experience that deepened my passion for evidence-based teaching and pedagogical research. These roles have allowed me to approach pedagogy from multiple perspectives, reflecting on how best to support learning, engagement, and intellectual growth. I view teaching as a dynamic, reciprocal process, evolving with every class, every student question, and every new challenge.

At the core of my approach is the belief that students learn best when they are **actively engaged**, feel **supported**, and see the **relevance** of what they're learning to their broader goals. I structure my teaching around three guiding principles: **active participation**, **collaborative learning**, and **clear, inclusive communication**.

1. **Active participation**: I design learning activities that invite students to explore, apply, and question new ideas, rather than passively absorb information. Whether I'm teaching statistical modeling, coding, or neuroscience, I prioritize opportunities for students to engage directly with the material.

   This might involve live coding exercises in R or Python, case-based problems using real-world datasets, or reading peer-reviewed research articles as a group and analyzing them through a critical lens. In biostatistics and data science courses, I encourage students to code along with me during live demonstrations, helping them build fluency with techniques in real time. In neuroscience and psychology, I often incorporate low-stakes quizzes and group reading aloud of primary research to reinforce key concepts and spark discussion.

   These strategies are grounded in evidence-based practices like retrieval practice, elaboration, and interleaving, and also support confidence. Many students report feeling more capable and motivated when they see themselves solving real problems and working through uncertainty in a supportive environment. For instance, students in my psychology class have noted, this approach "creates space to process ideas," ensures time to clarify confusing concepts, and leaves them "feeling confident in what was taught" because they are actively solving problems in a supportive environment.

2. **Collaborative learning**: I foster a classroom environment where dialogue between students and between students and instructor is central to learning. I intentionally build in activities that require peer interaction, like structured group discussions or paired programming in coding courses. These collaborative moments give students the space to articulate their ideas, confront diverse perspectives, and share knowledge.

Collaboration also helps demystify difficult content. When students explain ideas to each other or work through challenges together, they deepen their understanding. During group work, I often move through the room (or breakout rooms), ask questions to prompt deeper thinking, and encourage quieter students to contribute. In both in-person and virtual settings, this emphasis on dialogue fosters a sense of community, particularly important in technical courses that can feel intimidating at first. When students feel supported by their peers and their instructor, they are more willing to engage.

3. **Clear, inclusive communication**: I strive to communicate concepts and expectations with clarity and accessibility, while staying responsive to the diverse needs of my students. Having taught both undergraduate and graduate learners from varied academic backgrounds, I've learned that clarity is not just about simplifying content; it's about making the learning process transparent and inclusive.

   To support this, I use plain language explanations, visual aids (e.g., annotated code, diagrams), and step-by-step breakdowns of complex material. I also check for understanding frequently, through informal polls, Q&A sessions, and reflective check-ins. When introducing technical terms or methods, I make sure to define them in context and revisit them over time. Students have remarked that this approach "made the information really digestible without long–winded explanations" and that I "consistently went over concepts that were heavy or confusing," which helped them organize their studying and build confidence.

   Flexibility is also key to inclusion. I communicate expectations, grading criteria, and deadlines in multiple formats (e.g., verbally and in writing), and I create space for students to ask questions, formally and informally. As one student noted, this made it "one of the only courses where I would leave classes actually feeling confident in what was taught," underscoring the importance of transparency, patience, and support. I emphasize that learning is a gradual process, encouraging students to embrace the learning curve and be patient with themselves as they navigate new material. From the first day, I always tell students, "I'm here for support," to signal that I care about their success.

At its core, I see teaching as more than just sharing knowledge; it's about helping students build **critical thinking, curiosity, and confidence**. My goal is to support students in becoming thoughtful, capable learners who can apply complex ideas independently and adapt them to real-world challenges. Whether students pursue research, industry, or further study, I want them to leave my courses not only with a solid grasp of coding, neuroscience, and data analysis but also with the practical experience and intellectual habits needed to thrive in their next steps, whatever that may be. By showing **enthusiasm, staying flexible, and demonstrating openness to new ideas and challenges** myself, I aim not just to teach the material but to inspire a lasting capacity for learning, problem-solving, and professional growth.

Highlights of Teaching Experience

Undergraduate and graduate courses, University of Toronto

References for course descriptions and details are provided below.

Ref	Role	Course	Time	# of students	Duties
1	Head Teaching Assistant	HMB300 (Neurobiology of Behavior)	Fall 2025	300	Led weekly tutorials guiding student projects, provided feedback on deliverables, and mediated interactions between students and community partners.
2	Teaching Assistant	LMP2004 (Introduction to Biostatistics)	Fall 2025	20	Led weekly tutorials, prepared materials, graded weekly assignments, midterms, and exams.
3	Head Teaching Assistant	HMB200 (Introduction to Neuroscience)	Summer 2025	120	Designed tutorial materials, oversaw teaching Assistants, planned and delivered coding sessions, and managed email communication with students.
1	Teaching Assistant	HMB300 (Neurobiology of Behavior)	Summer 2025	60	Prepared tutorial materials, led weekly tutorials, marked assignments, and managed email communication with students.
4	**Course Instructor**	**PSY496 (Cognitive Dysfunction in Neurological Disorders)**	**Winter 2025**	**50**	**Designing syllabi and evaluations, planning and delivering lectures, coordinating duties with the course TA, liaising with students online and in-person, submitting final marks to the department for approval.**
3	Head Teaching Assistant	HMB200 (Introduction to Neuroscience)	Winter 2025	270	Designed tutorial materials, oversaw teaching Assistants, planned and delivered coding sessions, and managed email communication with

					students.
5	Head Teaching Assistant	HMB314 (Laboratory in Human Biology)	Winter 2025	150	Conducted weekly labs, led pre-lab meetings with TAs, graded lab reports, monitored discussion board questions, and managed student administrative inquiries (e.g., medical accommodations, accessibility requests).
1	Head Teaching Assistant	HMB300 (Neurobiology of Behavior)	Fall 2024	300	Prepared tutorial materials, led weekly tutorials, marked assignments, and managed email communication with students.
2	Teaching Assistant	LMP2004 (Introduction to Biostatistics)	Fall 2024	20	Led weekly tutorials, prepared materials, graded weekly assignments, midterms, and exams.
3	Head Teaching Assistant	HMB200 (Introduction to Neuroscience)	Summer 2024	120	Led weekly tutorials, prepared materials, graded term tests, and exams.
1	Teaching Assistant	HMB300 (Neurobiology of Behavior)	Summer 2024	60	Prepared tutorial materials, led weekly tutorials, marked assignments, and managed email communication with students.
3	Teaching Assistant	HMB200 (Introduction to Neuroscience)	Winter 2024	270	Led weekly tutorials, prepared materials, graded term tests, and exams.
5	Head Teaching Assistant	HMB314 (Laboratory in Human Biology)	Winter 2024	150	Conducted weekly labs, led pre-lab meetings with TAs, graded lab reports, monitored discussion board questions, and managed student administrative inquiries (e.g., medical accommodations, accessibility requests).
1	Teaching Assistant	HMB300 (Neurobiology of	Fall 2023	300	Prepared tutorial materials, led weekly tutorials, marked

		Behavior)			assignments, and managed email communication with students.
2	Teaching Assistant	LMP2004 (Introduction to Biostatistics)	Fall 2023	20	Led weekly tutorials, prepared materials, graded weekly assignments, midterms, and exams.
3	Head Teaching Assistant	HMB200 (Introduction to Neuroscience)	Summer 2023	120	Led weekly tutorials, prepared materials, graded term tests, and exams.
1	Teaching Assistant	HMB300 (Neurobiology of Behavior)	Summer 2023	60	Prepared tutorial materials, led weekly tutorials, marked assignments, and managed email communication with students.
3	Teaching Assistant	HMB200 (Introduction to Neuroscience)	Winter 2023	270	Led weekly tutorials, prepared materials, graded term tests, and exams.
5	Teaching Assistant	HMB314 (Laboratory in Human Biology)	Winter 2023	150	Conducted weekly labs, graded lab reports.
1	Teaching Assistant	HMB300 (Neurobiology of Behavior)	Fall 2022	300	Prepared tutorial materials, led weekly tutorials, marked assignments, and managed email communication with students.
1	Teaching Assistant	HMB300 (Neurobiology of Behavior)	Summer 2022	60	Prepared tutorial materials, led weekly tutorials, marked assignments, and managed email communication with students.
3	Teaching Assistant	HMB200 (Introduction to Neuroscience)	Summer 2022	120	Led weekly tutorials, prepared materials, graded term tests, and exams.
3	Teaching Assistant	HMB200 (Introduction to Neuroscience)	Winter 2022	270	Led weekly tutorials, prepared materials, graded term tests, and exams.
1	Teaching Assistant	HMB300 (Neurobiology of	Fall 2021	300	Prepared tutorial materials, led weekly tutorials, marked

		Behavior)			assignments, and managed email communication with students.
1	Teaching Assistant	HMB300 (Neurobiology of Behavior)	Summer 2021	60	Prepared tutorial materials, led weekly tutorials, marked assignments, and managed email communication with students.
6	Teaching Assistant	PSYC70 (Advanced Research Methods Laboratory)	Winter 2021 *(Scarborough campus)*	100	Graded weekly assignments, held office hours, and managed email communication with students.
1	Teaching Assistant	HMB300 (Neurobiology of Behavior)	Fall 2020	300	Prepared tutorial materials, led weekly tutorials, marked assignments, and managed email communication with students.

1. A third-year course offered by the Human Biology Department at the University of Toronto. This course examines higher brain functions and the biological mechanisms underlying human and animal behaviour. Topics may include the influence of the gut microbiome on behaviour, the impact of pathogens on neuronal development, and the neurobiological basis of placebo effects. In 2024, I was promoted to Head Teaching Assistant, a role in which I provided leadership and coordination among the TA team, supported course planning, and contributed to the refinement of instructional materials. In Fall 2025, the course shifted to **an experiential learning approach,** which I was heavily involved in implementing as head TA. This included integrating innovative pedagogy into the classroom, managing deliverables from external partners, and ensuring that students' projects provided a valuable contribution to the organizations involved.

2. A graduate-level course offered by the Department of Laboratory Medicine and Pathobiology at the University of Toronto. This course introduces fundamental concepts in biostatistics, emphasizing both theoretical foundations and practical applications. Students learn core statistical techniques relevant to general medicine, pathology, and clinical embryology.

3. A second-year course offered by the Human Biology Department at the University of Toronto. This course introduces students to the development, physiology, and dynamic function of the nervous system in relation to various aspects of human behaviour. Emphasis is placed on the critical evaluation of scientific evidence to support and enrich learning. In this large undergraduate course, I was fortunate to be invited to guest lecture and present my thesis research to smaller groups of approximately 50 students. These sessions were delivered during the terms of Winter 2022, Winter 2023, and Summer 2024. In 2023, I was promoted to Head Teaching Assistant, a role

in which I provided leadership and coordination among the TA team, supported course planning, and contributed to the refinement of instructional materials.

4. A fourth-year course offered by the Psychology Department at the University of Toronto. This course explores the cognitive impairments associated with neurological disorders such as Alzheimer's disease, Parkinson's disease, and schizophrenia, and emphasizes the integration of neuropsychological models with biological mechanisms to better understand the cognitive symptoms of these conditions. This was my first experience as the sole course instructor, and it was a formative opportunity that gave me insight into the full scope of teaching responsibilities, from preparation and content delivery to student communication and course administration. The small class size allowed for meaningful engagement, and I actively sought student feedback to guide continuous course improvement.

5. A second-year course offered by the Human Biology Department at the University of Toronto. This lab-based course introduces students to diverse areas of Human Biology research, with a focus on the body's responses to stressors such as physiological and environmental change. Students explore whole-body, cellular, and molecular responses using techniques ranging from standard clinical measures (e.g., blood pressure) to advanced research methods (e.g., RT-qPCR). In 2024, I was promoted to Head Teaching Assistant, a role in which I provided leadership and coordination among the TA team and managed student administrative inquiries.

6. A second-year course offered by the Psychology Department at the University of Toronto. This course prepares students to transition from consumers to producers of psychological research. Emphasis is placed on scientific reasoning, statistical thinking, and mitigating cognitive biases. Students also gain hands-on experience with R, RStudio, and R Markdown to support reproducible workflows.

Postgraduate certificate, University of Toronto

References for course descriptions and details are provided below.

Ref	Role	Course	Time	# of students	Duties
1	**Course Instructor**	**Linear regression, classification, and resampling**	**Fall 2025**	**125**	**Facilitated content delivery through virtual learning sessions, hosted office hours, coordinated duties with course TAs, and liaised with students online.**
2	Teaching Assistant	Introduction to Python	Fall 2025	125	Graded assignments, held office hours, monitored chat during live sessions, and managed student

					communication via Slack.
1	**Course Instructor**	**Linear regression, classification, and resampling**	**Summer 2025**	**125**	**Facilitated content delivery through virtual learning sessions, hosted office hours, coordinated duties with course TAs, and liaised with students online.**
2	Teaching Assistant	Introduction to Python	Summer 2025	125	Graded assignments, held office hours, monitored chat during live sessions, and managed student communication via Slack.
1	**Course Instructor**	**Linear regression, classification, and resampling**	**Winter 2025**	**125**	**Facilitated content delivery through virtual learning sessions, hosted office hours, coordinated duties with course TAs, and liaised with students online.**
2	Teaching Assistant	Introduction to Python	Winter 2025	125	Graded assignments, held office hours, monitored chat during live sessions, and managed student communication via Slack.
3	Teaching Assistant	Introduction to Building Software: Unix Shell, Git/GitHub & Python	Summer 2024	125	Graded assignments, held office hours, monitored chat during live sessions, and managed student communication via Slack
3	Teaching Assistant	Foundational Skills I: Unix Shell, Git/GitHub, Python, & Building Research Software	Winter 2024	125	Graded assignments, held office hours, monitored chat during live sessions, and managed student communication via Slack.
4	**Course Instructor**	**Introduction to R**	**Fall 2023**	**15**	**Facilitated content delivery through virtual learning sessions, hosted office hours, coordinated duties with course TAs, and liaised with students online.**
4	**Combined Teaching**	**Introduction to R**	**Summer 2023**	**10**	**Facilitated content delivery through virtual learning**

	Assistant and Course Instructor				**sessions, hosted office hours, graded assignments, and liaised with students online.**
5	Teaching Assistant	Estimating, Testing, and Learning	Winter 2023	15	Graded assignments, held office hours, monitored chat during live sessions, and managed student communication via Slack.

1. An intensive 3-week module offered by the Data Science Institute at the University of Toronto. This module introduced participants to the design, implementation, and evaluation of linear regression and classification models, including resampling techniques for model validation. Topics included the distinction between prediction and inference, interpretability, bias-variance trade-offs, and ethical implications of model-based decision-making. As the course instructor, I served as the primary technical facilitator and continuously adapted the pace and complexity of instruction based on student feedback.

2. An intensive 2-week module offered by the Data Science Institute at the University of Toronto. This module provides a comprehensive introduction to Python programming. Participants develop strong proficiency in Python fundamentals, with a particular focus on functions and object-oriented programming. This module also featured an invited industry speaker, offering learners exposure to real-world applications and career pathways in data science.

3. An intensive 2-week module from the Data Science Institute, University of Toronto. This module provides foundational skills in Unix shell and Git version control, emphasizing reproducibility. Participants develop proficiency in shell commands, file navigation, Git repositories, and collaborative workflows.

4. An intensive 2-week module from the Data Science Institute, University of Toronto. This module introduces programming using R, with a focus on data manipulation, visualization, and ethical and professional considerations in data science. As the course instructor, I served as the primary technical facilitator and continuously adapted the pace and complexity of instruction based on student feedback.

5. An intensive 3-week module from the Data Science Institute, University of Toronto, introducing the foundations of statistical learning with applications in R. The module covers key concepts such as linear and logistic regression, classification methods, resampling techniques, and model assessment. Emphasis is placed on the distinction between inference and prediction, model interpretability, and the ethical use of data in decision-making.

Professional Development Activities

Teaching Assistants' Training Program (TATP)
University of Toronto | 2024

TATP is a peer-led pedagogical training program supporting teaching across all three campuses of the University of Toronto. It offers evidence-based workshops and resources that promote inclusive and innovative teaching practices for TAs, course instructors, and future faculty. Completed modules included:

1. Online Technology & Student Engagement: This module supports TAs and instructors in promoting active, inclusive learning in synchronous and asynchronous online settings. It covers strategies for checking understanding, equitable participation, and designing engaging activities. It was key in developing techniques to foster meaningful virtual interaction.

2. Laboratories/Practicals: This module focused on planning and delivering effective lab sessions, covering pre-lab talks, procedure demonstrations, time management, monitoring progress, and formative feedback. It enhanced my skills in managing hands-on learning and connecting experiments to theory through structured teaching.

3. Discussion-Based Tutorials: This module emphasized preparation and structure to facilitate inclusive, meaningful discussions that deepen student learning, countering the belief that spontaneous dialogue alone suffices. It strengthened my ability to lead engaging, well-organized tutorials.

4. Grading Practices: This module covered University of Toronto grading policies, addressing fair and consistent grading across formats, effective feedback, managing expectations, and handling appeals. It improved my approach to equitable grading and clear communication to support student learning.

Attendee, Open Science in Undergraduate Education Symposium
Allen Institute, Seattle, U.S.A. | June 2025

Participated in a two-and-a-half-day symposium focused on fostering open science practices in undergraduate biology education. Sessions featured teaching-focused faculty across diverse disciplines. Highlights included the Integrating Programming into Biology and Neuroscience Courses workshop.

Teaching Innovation & Curriculum Development

Module Development: Data Learning Certificate (University of Toronto)

As a content developer with the Data Science Institute at the University of Toronto, I contributed to the creation and redesign of several modules in the Data Learning Certificate series. My work centered on the design and pedagogical development of modules in statistical and deep learning, with a focus on making complex theoretical material accessible through thoughtfully structured content and applied examples. Prof. Rohan Alexander, Director of Technical Skills Curriculum & Instruction, described my contributions as “invaluable to the development of these modules.” He specifically highlighted my initiative in redesigning the Linear Regression, Classification, and Resampling module, which transformed from “a frequent source of student complaints to one of the most highly reviewed offerings.” He further emphasized my “reliability, responsiveness, and pedagogical expertise,” qualities that underscore the impact of my work in shaping high-quality learning experiences.

Topics in Deep Learning: Healthcare, Medicine, and the Life Sciences

I contributed to the development of a domain-focused module introducing foundational deep learning models with applications to biomedical data. Innovations included designing clear and accessible slide decks that explained theoretical foundations using visual aids and analogies to support diverse learners, alongside domain-specific examples in genomics, medical imaging, and clinical decision-making. The module also integrated model interpretability tools and facilitated discussions on algorithmic bias in healthcare applications.

See GitHub Repository: https://github.com/UofT-DSI/deep_learning_topics

Topics in Statistical Learning: Linear Regression, Classification, and Resampling

I led the design of a 3-week module focused on statistical modelling techniques, including linear regression, classification, clustering, and statistical inference. Innovations included the development of slide decks that presented complex statistical modeling techniques in plain language, supported by visual walkthroughs and minimal formalism; interactive Jupyter notebooks for hands-on model tuning and validation; and scaffolded assignments that balanced theoretical understanding with practical implementation in Python.

See GitHub Repository: https://github.com/UofT-DSI/LCR

Topics in Statistical Learning: Regularization, Splines, and Trees

I led the design of a 3-week module focused on more advanced statistical modelling techniques, including ridge and lasso regularization, natural splines, and tree-based approaches. Innovations included the development of slide decks that presented complex statistical modeling techniques in plain language, supported by visual walkthroughs and minimal formalism; interactive Jupyter notebooks for hands-on model tuning and validation; and scaffolded assignments that balanced theoretical understanding with practical implementation in Python.

See GitHub Repository: https://github.com/UofT-DSI/regularization_splines_trees

Curriculum Development: Coding Integration in Neuroscience Education

As a graduate student, I served as a lead on UNICE (Undergraduate Neuroscience Initiative for Curriculum Enhancement), a project supported by the Learning & Education Advancement Fund from the Office of the Vice-Provost, Innovations in Undergraduate Education, University of Toronto, which aims to address a key gap in the neuroscience curriculum: the lack of structured programming instruction. In collaboration with Drs. Bill Ju and Jimmy Fraigne, this initiative integrates foundational coding skills (e.g., MATLAB, Python, and R) into existing neuroscience courses, including high-enrollment classes such as CJH332, HMB200, HMB300, and HMB310, as well as upper-year seminars and labs. Once fully implemented, it aims to reach up to 1,500 students annually. The rollout of this project includes phased integration of coding assignments and lecture content, instructor support, and collaboration with faculty to align computational tasks with course objectives. Annual evaluations will assess student outcomes and guide refinements, including the potential use of large language models (LLMs) to support coding instruction.

I was invited to present this initiative at the *Open Science in Undergraduate Education Symposium* hosted by the Allen Institute in Seattle, U.S.A., where I gave a poster presentation titled ***"Integrating Publicly Open Datasets and Coding into Neuroscience Education: Strategies for Large Undergraduate Courses."*** This project aims to enhance student readiness for research and industry in a field increasingly shaped by computational approaches.

Evaluation of Teaching Assistance

Feedback on PSYC70, Winter 2021

At the conclusion of the course, my supervising instructor completed a formal TA evaluation form, which provided detailed feedback on my teaching performance. The feedback focused on areas such as preparedness, communication, responsiveness to student needs, and overall contribution to the course. This evaluation was shared with me to support my ongoing development as a teaching assistant and to strengthen my instructional skills.

Scale: 1 - Exceeds expectations 2 – Meets expectations 3 - Does not meet expectations

Area	Evaluation
Reliability:	
Regularity of attendance at course planning/coordinating meetings (if such attendance is a job requirement)	2
Quality of contributions to course planning/coordinating meetings	2
Availability to students during office hours	2
Engagement:	
Demonstration of interest in the course and the class material	2
Proficiency:	
Understanding of the material covered in the course	1
Communication:	
Effective communication with students	1
Judgement:	
Good judgment in dealings with students	1
Grading:	
Accuracy and timeliness of grading written assignments	1
Quality of feedback/comments on written assignments/tests	1
Overall rating of TA effectiveness	1

Qualitative comment

"Julia provided excellent, detailed feedback to the students in her tutorial section for their projects. Grading was always done faster than expected, which helped students develop their weekly scaffolded project in a timely manner. Thanks, Julia - excellent work!"

Official course evaluations on HMB200 Summer 2022, HMB300 Fall 2022 & 2023

At the end of each semester, the University of Toronto's Human Biology Department administers standardized, department-wide course evaluations that include a component specifically designed to gather feedback on teaching assistants. However, as of Fall 2023, the department has paused the administration of TA-specific evaluations due to updated faculty-level policies. Below is a summary of the weighted mean scores across each evaluated course.

Scale: 1 - Poor 2 – Fair 3 - Good 4 - Very Good 5 - Excellent

Question **For your TA, did they …**	**HMB200** **Summer 2022**	**HMB300** **Fall 2022**	**HMB300** **Fall 2023**
Prepare and know the subject thoroughly?	4.0	4.8	4.8
Present material clearly, in a logical, organized way?	4.0	4.8	5.0
Facilitate your learning by aiming at an appropriate level?	3.5	4.6	4.8
Use creative teaching methods effectively?	4.5	4.2	4.6
Encourage questions?	4.5	4.6	4.8
Answer students' questions effectively?	4.0	4.8	5.0
Encourage group discussion?	4.0	4.6	4.4
Mark fairly?	4.0	4.0	5.0
Create a positive learning environment/atmosphere?	4.0	4.8	5.0
What is your overall rating of your TA's contribution to your learning experience?	5.0	4.8	5.0

Note. These evaluations reflect a clear and consistent trajectory of growth in my teaching effectiveness over time. Initial feedback from Summer 2022 suggested a solid foundation, with room to improve in areas such as facilitating learning at an appropriate level. By Fall 2022 and continuing into Fall 2023, scores improved markedly across nearly all categories, with several perfect ratings by the most recent term. Notably, my ability to present material clearly, respond to questions effectively, and create a positive learning environment reached peak scores in Fall 2023, indicating increasing pedagogical confidence and student engagement. This upward trend underscores my commitment to reflective teaching and continuous improvement.

<u>Qualitative comments</u>

The following are anonymized, unsolicited student comments collected from the official course evaluations in HMB300 Fall 2022 and Fall 2023.

HMB300 Fall 2022:

Student 1

> “I'm so lucky I had Julia this semester. She was an amazing TA and was always transparent with what her expectations were. She welcomed questions and always answered them to the best of her abilities. She cared about what we wanted to learn as well.”

Student 2

> “Julia is a great TA, she always answers questions and encourages for questions to be asked. Additionally, she reminds us that extensions are available for projects, which is very helpful.”

Student 3

> “Julia is the best! I've had her for multiple courses, and she always does a great job explaining the content and answering any questions.”

HMB300 Fall 2023:

Student 1

> “Julia was an excellent TA. She always encouraged us to come to tutorial with questions and was very helpful in assisting with my learning.”

Student 2

> “Julia came prepared to each tutorial and answered questions throughout every single session. Super friendly, knowledgeable on the course material, and provides a lot of insight and help for both assignments! I also liked how she provided recordings of lectures and created an environment that encouraged questions.”

Feedback on LMP2004, Fall 2024

At the conclusion of the course, I distributed a brief anonymous Google Form on the final day of class. Students were invited to share any final thoughts, reflections, or suggestions. The purpose of this informal anonymous feedback was to identify strengths of the course as well as areas for continued improvement, in order to better align with student expectations in future iterations.

Scale: 1 - Poor 2 – Fair 3 - Good 4 - Very Good 5 - Great

Question	Mean	Median
General Feedback: How would you rate the tutorial overall?	4.7	5.0
Delivery and Engagement: How engaging was your TA?	4.3	4.0

Qualitative comments

The following are anonymized, unsolicited student comments collected via an anonymous Google Form distributed on the final day of the course.

Student 1

> "I liked how Julia organized her tutorials, as they made logical sense to me and helped me understand the lecture content much better. It was helpful to hear her explain concepts and theory first, then move into equations and R script. She also often paused between topics to check in with the class to ensure we understood the content."

Student 2

> "Julia really clarified basically everything about the lectures. I really appreciate Julia's rapid adjusting and modifying of the tutorials based on what was not covered clearly in the lectures."

Student 3

> "Julia, you were amazing and if you choose to be a professor one day, your students would be lucky to have you. I wish you best of luck with your PhD and your defense!"

These comments reflect my clear and organized teaching style, thoughtful integration of theory and practice, and a responsive approach to student needs. Students appreciated the logical progression from concepts to code and noted my flexibility in adjusting tutorials to address gaps in lecture content. Several also expressed strong appreciation for my teaching and encouragement toward a future in academia, highlighting both pedagogical impact and positive rapport.

Feedback on HMB300, Summer 2025

At the end of the semester, I shared a short anonymous Google Form during the final tutorial session, inviting students to provide any closing thoughts, reflections, or suggestions. The goal of this informal feedback was to highlight both the strengths of the tutorials and opportunities for improvement, helping to better align future versions with student expectations.

Scale: 1 - Not at all 2 – Somewhat 3 - Moderately 4 - Mostly 5 - A great deal

Question	**Mean**	**Median**
General Feedback: I found the tutorials intellectually stimulating.	4.3	5.0
General Feedback: The tutorials provided me with a deeper understanding of coding with R.	4.5	5.0
Delivery and Engagement: **Julia** created an atmosphere that was conducive to my learning.	4.9	5.0
Delivery and Engagement: **Julia** generated enthusiasm for learning in the course and was enthusiastic about the tutorial material.	4.7	5.0

Scale: 1 - Poor 2 – Fair 3 - Good 4 - Very Good 5 - Excellent

Question	**Mean**	**Median**
General Feedback: Overall, the quality of my learning experience in Tutorials was	4.6	5.0

Quantitative feedback was consistently positive. Students found the tutorials intellectually stimulating, gained confidence in R coding, and described the environment as supportive and engaging. As Head TA, I designed the tutorial materials and led live coding sessions. The feedback affirms the effectiveness of my hands-on, supportive teaching approach while also identifying areas for ongoing refinement.

Evaluation of Teaching

Feedback on PSY496, Winter 2025

Below are a few representative emails I received from students following the conclusion of the Winter 2024 term. These unsolicited messages reflect the impact of the course on students' engagement, understanding, and academic interests. They are reproduced in full with minor formatting adjustments (e.g., spacing, anonymization) to preserve their original tone and content.

Email 1

> "Hi Professor Gallucci,
>
> I hope you're doing well and that you're enjoying the warm weather!
> I wanted to say how much I truly enjoyed PSY496. It was genuinely one of my favourite courses during undergrad. The content was so engaging that it felt more like learning for interest than fulfilling a requirement. Your teaching style made complex topics much more accessible, and I really appreciate how you took the time to check in with us, clarify concepts, and make sure no one was left behind.
>
> [...]
> Thank you again for a meaningful term. I learned so much from this course, and its actually helped me solidify my passion for clinical psychology and research! :)"

This student emphasizes the course's ability to foster motivation and sustained engagement. They also highlight the accessibility of complex material and the individualized attention provided, both central to my teaching approach. Notably, the course appeared to have had a formative impact on the student's academic direction.

Email 2

> "Good afternoon professor,
>
> I hope you're doing well. As the semester comes to an end, I wanted to take a moment to express my sincere gratitude for the incredible learning experience you provided in PSY496.
> Your teaching style and the way you designed the course made a significant impact on me. The clarity of your lectures and the depth of the content created a stimulating environment that truly enhanced my understanding of the subject.
>
> I also wanted to let you know how much this course meant to me personally, as I have a strong interest in this subject. Your passion for the topic was evident and contagious, and it made all the difference in keeping me motivated and excited to learn. This experience has solidified my interest in continuing to pursue this field, and I feel very fortunate to have had you as my professor.

Thank you once again for your dedication, support, and for fostering such a great learning environment. I look forward to carrying the knowledge gained in this course with me as I continue my studies.
Wishing you a restful summer!“

This message reinforces the importance of structure, clarity, and enthusiasm in maintaining a stimulating and motivating academic environment. The student specifically attributes their continued interest in the field to the learning experience in PSY496, indicating the long-term impact of effective, passion-driven instruction.

Email 3

“Hi Julia,
I hope you are enjoying the warm weather! I first wanted to extend my thanks for an excellent semester. I really enjoyed the class environment you fostered as well as the content discussed in the lecture.
[...]
Thank you, and I hope to hear from you soon!”

Though brief, this note underscored the importance of classroom climate and content relevance. It reflects a sense of comfort and connection in the learning environment, consistent with my goal of fostering a respectful and engaging upper-year course experience.

Email 4

“Hi Julia,

I hope you are doing well and having a lovely summer!

I wanted to reach out and discuss volunteering at your lab to develop my research experience. Your class and your teaching made a major impact on my research aspirations and my desire to pursue opportunities in neuropsychiatry, and I wanted to thank you for such an incredible learning experience!”

This message highlights the course’s role in inspiring student motivation and research engagement. It reflects the accessibility of complex material and a supportive learning environment, aligning with my goal of fostering curiosity and confidence in students.

Official course evaluations on PSY496 Winter 2025

At the University of Toronto, the Faculty of Arts and Science conducts official course evaluations during the final two weeks of each semester. The evaluation includes standardized questions at three levels: institutional, divisional, and departmental. In addition, students are invited to provide open-ended comments to elaborate on their learning experience.

In *PSY496 – Cognitive Dysfunction in Neurological Disorders*, 20 of 49 enrolled students (41%) completed the course evaluation. According to the 2025 Course Evaluation Interpretation Guidelines issued by the Centre for Teaching Support and Innovation (CTSI), this response rate provides a general estimate of the overall class perception, with a typical margin of error of approximately ±0.6 to ±1.0.

Institution-wide items

Scale: 1 - Not At All 2 – Somewhat 3 - Moderately 4 - Mostly 5 - A Great Deal

Question	Mean	Median
I found the course intellectually stimulating.	4.7	5.0
The course provided me with a deeper understanding of the subject matter.	4.8	5.0
The instructor created an atmosphere that was conducive to my learning.	4.7	5.0
Course projects, assignments, tests, and/or exams improved my understanding of the course material.	4.5	5.0
Course projects, assignments, tests, and/or exams provided an opportunity for me to demonstrate an understanding of the course material.	4.4	5.0

Scale: 1 - Poor 2 – Fair 3 - Good 4 - Very Good 5 - Excellent

Question	Mean	Median
Overall, the quality of my learning experience in this course was:	4.5	5.0

Divisional items

Scale: 1 - Not At All 2 – Somewhat 3 - Moderately 4 - Mostly 5 - A Great Deal

Question	Mean	Median
The instructor generated enthusiasm for learning in the course.	4.7	5.0
I would recommend this course to other students	4.6	5.0

Scale: 1 - Very light   2 – Light   3 - Average   4 - Heavy   5 - Very heavy

Question	**Mean**	**Median**
Overall, the quality of my learning experience in this course was:	3.4	3.0

Departmental items

Scale: 1 - Not At All   2 – Somewhat   3 - Moderately   4 - Mostly   5 - A Great Deal

Question	**Mean**	**Median**
The course material inspired me to learn more about the subject matter.	4.5	5.0
The course instructor was enthusiastic about the course material	4.8	5.0
The course instructor (Julia Gallucci) expressed an interest in student understanding when explaining course concepts.	4.9	5.0

Qualitative comments

The following are some anonymized, unsolicited student comments collected from the official course evaluations in Winter 2025. These testimonials reflect consistent themes across student feedback: clarity and accessibility of instruction, depth and relevance of content, supportive learning environment, and overall teaching excellence.

Student 1

> "I have nothing but positive things to say about Julia and about this course. I genuinely looked forward to each week, and I have never felt this way towards any class at UofT (maybe one other, thanks to Elli Weisbaum!). This course was intellectually stimulating difficult, challenging, but enjoyable in every way. Julia made sure to distill complex topics and studies down and at each point asked us if we understood the material. This simple ask makes the world of a difference. Most instructors never do this, instead they treat the lecture as a keynote presentation and hammer the details to you without giving you a break or a chance to digest the material. I love how Julia brought recent, cutting edge research to always inform us how the field has evolved, and where we are with research and treatment options today. In this type of course and program, that is incredibly important. It's not just about learning the theories but also knowing where the field is today, and what the current research says. While this course was challenging, I absolutely loved the material and Julia's teaching approach. She summarized the ideas so eloquently, and relayed her teachings to us so well that I always stayed behind class to write more notes and ask questions. She made herself available to us whenever she could, and it's evidently clear she cares for her students beyond what is expected for instructors. This

made the utmost difference for her students, and we could really feel her impact! Thank you for a wonderful course to end my last semester as a student here.”

Student 2

“The overall quality in this course was done very well. Professor Gallucci is an excellent professor and I really appreciate how in depth she went in terms of the concepts surrounding the course. As this topic was very memorization heavy, she made sure to provide as much time to cover concepts where we tended to struggle in and ensured that she answered any questions the students had.
She would consistently go over concepts that were heavy or confusing and I truly appreciate the time she would take to make sure we understood the topics of the course correctly.”

Student 3

“Professor Gallucci was one of my favorite instructors this year. She explained the concepts really well and in a really digestible manner without long–winded explanations that made the information confusing. We always spent the most amount of time on the most relevant information instead of focusing on tangents, which I really liked because it allowed me to focus on what is most important. Because of this, it was easy to know how to organize my studying the best. I also liked that if certain concepts were confusing, we would spend extra time discussing them instead of skipping over the clarifications.”

Student 4

“The overall quality of the instruction in this course was excellent! The lectures were very well–strcutured – they were easy to follow along with, making it simple to grasp key concepts, and they effectively highlighted the most important points (I especially appreciated the summary/recap slides). Professor Gallucci helped create a seamless learning experience, her teaching approach was very organized and engaging! She was always available for support, whether that was during class or staying after lecture to answer everyone's questions. Her passion for the various neurological disorders we covered in class made me even more excited about the course, and I enjoyed each and every lecture. Overall, she fostered a very supportive learning environment, and the quality of instruction was fantastic.”

Student 5

“Amazing professor. One of the best in all my years at uoft. Is incredibly knowledgeable, well articulated, engaging, and incredibly helpful. Always ready to answer any questions or repeat explanations to support our learning. Pays attention to and truly cares about students' learning/understanding of the material. One of the only courses where I would leave classes actually feeling confident in what was taught rather than leaving knowing I'd have to 'figure it out later'.”

Official module evaluations DSI Linear Regression, Classification & Resampling, Winter and Summer 2025

At the University of Toronto, the Data Science Institute administers official module evaluations mid-certificate to gather structured feedback on learners' experiences. These evaluations included one standardized question targeting the performance of the course instructor (also known as the technical facilitator).

In Winter 2025, 111 of 125 enrolled participants (89%) completed the certificate evaluation. In Summer 2025, 114 of 125 enrolled participants (91%) completed the evaluation. The DSI combined both evaluations into a single report to assess trends across offerings and inform improvements to the certificate program.

Scale: Exceptional Good Average Poor

Module evaluation	Winter and Summer 2025
Rate the facilitation and teaching for LCR	40.3 % Exceptional 42.0 % Good 15.0 % Average 2.65 % Poor

Note. These ratings represent a marked improvement from a previous iteration of the module, in which only 13.3% of learners rated the facilitator as **Exceptional** and 18.7% rated them as **Poor**, suggesting meaningful gains in facilitation quality and overall learner experience.

Representative Teaching Materials

PSY496 – Cognitive Dysfunction in Neurological Disorders

Sample Syllabus
This syllabus is from the most recent iteration of *Cognitive Dysfunction in Neurological Disorders*, taught at the University of Toronto. It reflects my efforts to establish an engaging and supportive tone with students. Visual formatting (e.g., bold text, color highlights) is intentionally used to emphasize key information and set clear expectations.

Sample Exam Questions
This selection of exam questions demonstrates my approach to assessment, which emphasizes critical thinking and the application of course content beyond rote memorization. Multiple-choice questions are designed to test conceptual understanding and real-world relevance, while short-answer questions assess students' ability to synthesize and interpret empirical research.

Sample Check-In Quiz
To reinforce engagement with both lectures and assigned readings, I incorporated five low-stakes "check-ins" throughout the term. These short quizzes were designed to assess students' comprehension and encourage regular review of course material.

Term Paper Assignment
A major component of PSY496, the term paper allows students to explore a topic of personal interest within the course framework. They are encouraged to develop a focused research question and construct a clear, evidence-based argument. This assignment fosters independent inquiry and strengthens academic writing skills, which are key goals of 400-level psychology courses at the University of Toronto.

Sample course syllabus

PSY496H1 S

Cognitive Dysfunction in Neurological Disorders

Winter 2025 Syllabus

Course Meetings

PSY496H1 S

Section	Day & Time	Delivery Mode & Location
LEC0101	Wednesday, 2:00 PM - 5:00 PM	In Person: ES B142

Refer to ACORN for the most up-to-date information about the location of the course meetings.

Course Contacts

Course Website: https://q.utoronto.ca/courses/382584/

Instructor: Julia Gallucci
Email: julia.gallucci@mail.utoronto.ca
Office Hours and Location: Online, Monday 11:00 am -12:00 pm
https://utoronto.zoom.us/j/5457297556

Teaching Assistant: [redacted]
Email: [redacted]
Office Hours and Location: Online, Thursday 10:00 am - 11:00 am
[redacted]

Additional Notes: Please allow 24-48 hours for a response during regular business hours. When emailing, please include the course code, lecture number, and slide number in the subject line, for example, "PSY496 L1:Slide 12."

Course Overview

This lecture course provides an in-depth examination of the cognitive dysfunction found in neurological disorders, including Alzheimer's Disease, Parkinson's Disease, and Schizophrenia. The course focuses on how cognitive impairments relate to neuropsychological models of the specific disease (specifically, how these models provide insights into the biological mechanisms underlying the cognitive symptoms) and cognition, more broadly.

Course Learning Outcomes

Students will engage with primary research articles examining the connection between cognition and brain function across various neurological and psychological conditions. They will integrate this knowledge into a final paper, where they will propose a novel research experiment.

Prerequisites: PSY201H1/ ECO220Y1/ EEB225H1/ GGR270H1/ IRW220H1/ POL222H1/ SOC202H1/ STA220H1/ STA238H1/ STA248H1/ STA288H1/ ECO220Y5/ PSY201H5/ STA215H5/ STA220H5/ PSYB07H3/ STAB22H3/ STAB23H3/ STAB57H3, *and one of* PSY270H1/ PSY270H5/ PSYB57H3/ COG250Y1 *or* PSY290H1/ PSY290H5/ PSYB64H3/ HMB200H1/ PSL300H1
Corequisites: None
Exclusions: None
Recommended Preparation: None
Credit Value: 0.5

Marking Scheme

Assessment	Percent	Due Date
Term Test 1	25%	2025-02-12
Term Test 2	25%	2025-04-02
Paper Topic	10%	2025-02-26
Final Paper	35%	2025-03-19
Quiz	5%	Multiple dates

Late Assessment Submissions Policy

Assignments will be penalized 10% for each day after the deadline, to a maximum of 4 days (40%), after which papers will no longer be accepted. Extensions will only be granted in exceptional circumstances (i.e., illness or family emergency) with valid documentation in the Specific Medical Circumstances section.

Course outline/Schedule			
Lecture	**Date**	**Topic**	**Due**
1	Jan 8	Neuropsychology and methods overview	
2	Jan 15	Neuroanatomy and imaging methods overview	
3	Jan 22	Epilepsy	Quiz 1
4	Jan 29	Parkinson's Disease	Quiz 2
5	Feb 5	Frontotemporal Dementia	Quiz 3

6	Feb 12	**Term test 1**	
	Feb 19	**Reading week**	
7	Feb 26	Alzheimer's Disease 1	**Paper topic due**
8	March 5	Alzheimer's Disease 2	Quiz 4
9	March 12	Schizophrenia	Quiz 5
10	March 19	Mood Disorders	**Final Paper due**
11	March 26	Autism	
12	April 2	**Term test 2**	

Policies & Statements

Religious Accommodations

As a student at the University of Toronto, you are part of a diverse community that welcomes and includes students and faculty from a wide range of cultural and religious traditions. For my part, I will make every reasonable effort to avoid scheduling tests, examinations, or other compulsory activities on religious holy days not captured by statutory holidays. Further to University Policy, if you anticipate being absent from class or missing a major course activity (such as a test or in-class assignment) due to a religious observance, please let me know as early in the course as possible, and with sufficient notice (at least two to three weeks), so that we can work together to make alternate arrangements.

Students with Disabilities or Accommodation Requirements

Students with diverse learning styles and needs are welcome in this course. If you have an acute or ongoing disability issue or accommodation need, you should register with Accessibility Services (AS) at the beginning of the academic year by visiting https://studentlife.utoronto.ca/department/accessibility-services/. Without registration, you will not be able to verify your situation with your instructors, and instructors will not be advised about your accommodation needs. AS will assess your situation, develop an accommodation plan with you, and support you in requesting accommodation for your coursework. Remember that the process of accommodation is private: AS will not share details of your needs or condition with any instructor, and your instructors will not reveal that you are registered with AS.

Academic Integrity

All suspected cases of academic dishonesty will be investigated following procedures outlined in the Code of Behaviour on Academic Matters (https://governingcouncil.utoronto.ca/secretariat/policies/code-behaviour-academic-matters-july-1-2019). If you have questions or concerns about what constitutes appropriate academic behaviour or appropriate research and citation methods, please reach out to me. Note that you are expected to seek out additional information on academic integrity from me or from other institutional resources. For example, to learn more about how to cite and use source material appropriately and for other writing support, see the U of T writing support website at http://www.writing.utoronto.ca. Consult the Code of Behaviour on Academic Matters for a complete outline of the University's policy and expectations. For more information, please see A&S Student Academic Integrity (https://www.artsci.utoronto.ca/current/academic-advising-and-support/student-academic-integrity) and the University of Toronto Website on Academic Integrity (https://www.academicintegrity.utoronto.ca).

Mental Health and Well-Being

Your mental health is important. Throughout university life, there are many experiences that can impact your mental health and well-being. As a University of Toronto student, you can access free mental health and wellbeing services at Health & Wellness (https://studentlife.utoronto.ca/department/health-wellness/) such as same-day counselling, brief counselling, medical care, skill-building workshops, and drop-in peer support. You can also meet with a Wellness Navigation Advisor who can connect you with other campus and community services and support. Call the mental health clinic at 416-978-8030 ext. 5 to book an appointment or visit **https://uoft.me/mentalhealthcare** to learn about the services available to you.

You can also visit your College Registrar to learn about the resources and supports available: https://www.artsci.utoronto.ca/current/academic-advising-and-support/college-registrars-offices

If you're in distress, you can access immediate support: https://uoft.me/feelingdistressed

Re-marking Policy - Timeline and Protocol

If you believe an assignment has received a grade in error, you may submit an appeal. An appeal must be submitted within 14 days after the graded assignment is made available to students. Documents submitted for an appeal will be regraded in their entirety. As a result, your grade may increase, but it may also stay the same or even decrease.

Generative AI

On the use of Generative Artificial Intelligence (Gen AI) Tools: Students may not use generative artificial intelligence tools (e.g., ChatGPT) for the term tests in this course, but may use these tools for other assignments. If you choose to use generative artificial intelligence tools to assist you in the assignments in this course, this use must be documented in an appendix for each assignment. The documentation should include what tool(s) were used, how they were used (e.g., include your prompts), and how the results from the AI were incorporated into the submitted work. These tools can be most helpful in improving your writing and the clear expression of your ideas (rather than trying to generate complete content, which is unlikely to meet the standards of the assignments).

Make-Up Quizzes/Tests

Missed or unexcused tests will be treated as zeros unless valid documentation is provided, as outlined in the Specific Medical Circumstances section. Students have up to one week from the date of the missed term test to contact the instructor and provide the necessary documentation. Should term test 1 or term test 2 be missed with valid documentation, a make-up test will be arranged. If a student is unable to attend a make-up test due to documented circumstances, then the remaining assessments can be reweighted to account for the missed test.

Specific Medical Circumstances

If you become ill and it affects your ability to do your academic work, consult me right away. Normally, I will ask you for documentation in support of your specific medical circumstances. This documentation can be an Absence Declaration (via ACORN) or the University's Verification of Student Illness or Injury (VOI) form. The VOI indicates the impact and severity of the illness, while protecting your privacy about the details of the nature of the illness. If you cannot submit a VOI due to limits on terms of use, you can submit a different form (like a letter from a doctor), as long as it is an original document, and it contains the same information as the VOI (including dates, academic impact, practitioner's signature, phone and registration number). For more information on the VOI, please see http://www.illnessverification.utoronto.ca. For information on the Absence Declaration Tool for A&S students, please see https://www.artsci.utoronto.ca/absence. If you get a concussion, break your hand, or suffer some other acute injury, you should register with Accessibility Services as soon as possible.

Departmental Guidance for Undergraduate Students in Psychology

The Department of Psychology recognizes that, as a student, you may experience disruptions to your learning that are out of your control, and that there may be circumstances when you need extra support. Accordingly, the department has provided a helpful guide to clarify your and your instructor’s responsibilities when navigating these situations. This guide consolidates Arts & Science Policies for undergraduate students in one place for your convenience. As an instructor in the department, I will frequently consult with these recommendations when providing you with support, and I recommend that you also consult them to learn more about your rights and responsibilities before reaching out to me.

Submitted term test questions

Multiple choice

Q: Deep brain stimulation for epilepsy can be either open loop, in which __________ or it can be closed loop, in which ________.

a. **stimulation is delivered intermittently according to device programming; stimulation is delivered when the device detects a seizure beginning**
b. stimulation is delivered when the device detects a seizure beginning; stimulation is delivered intermittently according to device programming
c. Stimulation is high frequency; Stimulation is applied in theta-bursts
d. Stimulation is applied in theta-bursts; Stimulation is high frequency

Q: When performing a working memory task, patients with Schizophrenia showed:

a. The same level of brain variability as healthy controls
b. Lower brain variability compared to healthy controls
c. **Increased brain variability compared to healthy controls, primarily driven by a subset of patients with atypical activity patterns.**
d. Enhanced cognitive performance is associated with higher brain variability.

Short answer

Q: Deep brain stimulation (DBS) to treat Parkinson's Disease is beneficial to treat motor symptoms, but provides little benefit to cognitive symptoms. Describe how DBS has been modified to improve cognition in Parkinson's disease (Lee et al, 2021). Where and how is the stimulation done? What are the areas of cognition implicated/what tests were their main question about? How did experimenters run the experiment to test the effects of DBS on cognition? What were the results of the experiment?

A: In Lee et al (2021), DBS was altered in terms of frequency to improve cognition. (**1 mark**)

DBS was provided to the *dorsal* and *ventral* subthalamic nucleus (STN) (**1 mark**) at theta frequency and baseline clinical frequencies (also will accept 130Hz or gamma) (**1 mark**).

Verbal fluency was implicated (episodic and non-episodic) (**1 mark**).

The experimenters ran each of the different stimulation paradigms (dorsal vs ventral and theta vs high frequency) on each patient and measured their verbal fluency (episodic and non-episodic) for each change (**2 marks**) starting 5 minutes after the change.

They found that theta frequency (**1 mark**) to the dorsal STN (**1 mark**) improved [overall or episodic] verbal fluency (**1 mark**).

There was no effect for non-episodic verbal fluency / other stimulation sites/frequencies (**1 mark**).
OR
Theta stimulation did not have an effect on motor symptoms (**1 mark**).

Sample check-in quiz

1. According to the study by Audrain et al. (2018), how is matrix similarity between a patient's brain network and the normative template related to post-surgical language abilities?
    a. Higher similarity predicts greater declines in language abilities post-surgery.
    b. Lower similarity predicts better language resilience post-surgery.
    c. Higher similarity predicts better resilience in post-surgical language abilities.
    d. Similarity levels have no significant relationship with post-surgical language outcomes.

2. What does the greater activation of the right hemisphere during language tasks suggest about a patient undergoing a left anterior temporal lobectomy?
    a. The patient is more likely to experience language deficits after surgery.
    b. The patient is less likely to experience language deficits after surgery.
    c. The patient will have no change in language abilities post-surgery.
    d. The patient will likely lose all language functions after surgery.

3. Which of the following is true regarding cognitive abilities in patients with temporal lobe epilepsy?
    a. Left medial temporal lobe epilepsy (L-mTLE) leads to lower visual-spatial memory scores compared to right medial temporal lobe epilepsy (R-mTLE).
    b. Right medial temporal lobe epilepsy (R-mTLE) leads to lower verbal memory scores compared to left medial temporal lobe epilepsy (L-mTLE).
    c. Left medial temporal lobe epilepsy (L-mTLE) leads to lower verbal memory scores, while right medial temporal lobe epilepsy (R-mTLE) leads to lower visual-spatial memory scores.
    d. Both left and right medial temporal lobe epilepsy (L-mTLE and R-mTLE) equally affect verbal and visual-spatial memory.

4. Which of the following is true about sarcasm comprehension in patients with temporal lobe epilepsy (TLE) and anterior temporal lobectomy?
    a. Patients with TLE and anterior temporal lobectomy show no significant differences in sarcasm comprehension compared to controls.
    b. Sincere interaction comprehension is significantly impaired in patients with TLE and anterior temporal lobectomy compared to controls.
    c. Sarcasm comprehension is preserved in patients with TLE and anterior temporal lobectomy, but sincere comprehension is impaired.
    d. Sarcasm comprehension is significantly impaired in patients with TLE and anterior temporal lobectomy, especially in the left anterior lobectomy group.

Term paper assignment

Paper Topic Assignment

Learning Objectives for the <u>Final</u> Research Paper:

1. **Thorough Exploration of a Neurological Disorder in the Field**
    - Develop a deeper understanding of the theories and methodologies in the field.
    - Identify gaps in the current body of knowledge within the field.
2. **Cultivate Skills to Design Experiments in the Field**
    - Construct research methods aimed at addressing identified gaps in the literature.
    - Clearly explain how these methods target the gap.
    - Articulate the significance of potential findings.
    - Highlight how these findings could influence the broader scientific literature.
3. **Enhance Scientific Writing Proficiency**
    - Present ideas in a coherent, organized, and engaging manner, emulating the structure and style of a primary research article.

Paper Topic Description

Outline your plan for the final research paper in complete sentences. Specify the neurological (e.g., Parkinson's disease, Epilepsy, Alzheimer's disease) and cognitive domains (e.g., episodic memory, language, executive function) to be explored. Clearly define the gap in the literature that your study will address and briefly outline your study design. At this stage, there is no requirement to include results or conclusions.

Your teaching Assistant will provide feedback on the suitability of your topic and study design, helping you refine your final paper. The paper topic should be one page, single-spaced, using a 12-point font (Times New Roman or Arial). References are not included in the page count. Use APA formatting for citations. While there is no minimum number of citations required, any factual assertions about a disorder or cognitive function must be supported by appropriate references. If you're uncertain about citation usage, refer to a course reading to see examples of when citations are included. Generally, statements based on research require citations, while descriptions of hypotheses or novel methods do not.

Purpose of the Paper Topic Assignment

This assignment serves as an opportunity to present your ideas and receive constructive feedback in preparation for the final paper. It encourages early planning and provides a framework for refining your study before the final submission.

Research Paper Topic Rubric /10

Indicator	Not Demonstrated	Below Standard	Meets Standard	Exceeds Standard
Identify cognitive and neurological domains (3 marks)	No cognitive or neurological domain identified	Only one domain (cognitive or neurological) was identified	Both domains are identified	Both domains were identified and their significance explained
Identify a gap in the literature (3 marks)	No gap identified	Gap identified but already addressed	Gap identified	The gap was identified, and its exploration was motivated
Design a study to address the gap (3 marks)	No study design provided	Partial design provided	Comprehensive design provided	Comprehensive design and explanation of methods addressing the gap
Writing style (1 mark)	Numerous errors and poor organization	Few errors and reasonably organized	Few errors with clear reasoning	Error-free with strong, clear reasoning

General Advice

To identify gaps in the literature, start by reviewing recent papers on your topic. These often summarize key studies, highlight unresolved issues, and suggest future research directions. For example, when findings between studies conflict, consider factors like differing methodologies or participant characteristics (e.g., early-onset vs. late-onset Parkinson's disease).

An excellent paper topic submission will emphasize the significance and rationale for the proposed study. For example: "This research will investigate the impairment of **[specific cognitive domain]** in **[specific disorder]**. This is important because **[relevant statistics or impact of the disorder]**. Few studies have addressed this issue, and understanding the associated clinical factors may inform the development of targeted interventions to improve outcomes."

When proposing your study, you do not have to intentionally place limitations on yourself that may be typical in clinical research. For example, any given clinical site may not be able to collect a reasonable sample size for subtypes of disorders (in the real world). In this paper, if you have a question about subtypes, then you can say that you will collect data from the various subtypes. Assume you have this ability (it could be that your experiment proposes collecting data from a collection of collaborating clinical sites around the world!). However, you cannot ignore limitations that would exist in your proposed experiment due to properties of the physical world. For example, you cannot say that you will use an MRI machine that can collect functional imaging data at a millisecond resolution because such a machine does not exist.

DSI- Linear Regression, Classification, and Resampling

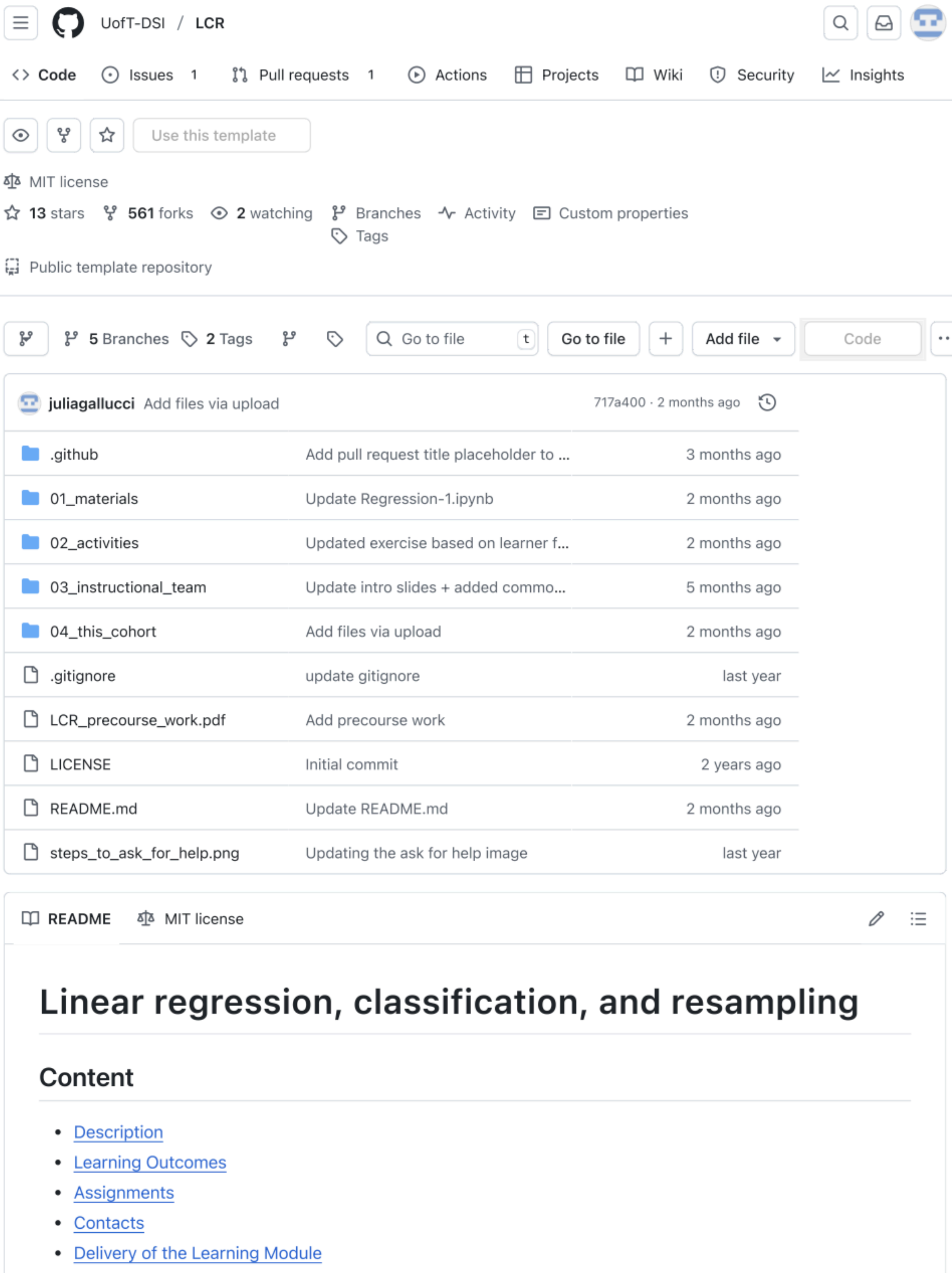
Sample Syllabus
This README serves as a syllabus from the most recent offering of Linear Regression, Classification, and Resampling, taught at the Data Science Institute, University of Toronto. It reflects my efforts to create an engaging and supportive learning environment. Visual formatting, such as bold text and hyperlinks, is intentionally used to highlight key information and communicate expectations clearly.

Sample Notebook
This notebook reflects my teaching approach, which emphasizes active student engagement. It is structured to support live, code-along demonstrations, enabling learners to practice new techniques in real time and develop fluency with core data science workflows.

Sample Assignment
To deepen hands-on engagement in an introductory statistical modeling course, I developed a Python-based assignment that leads students through the entire data analysis process, from data loading and cleaning to exploratory analysis and basic model fitting. The assignment is designed to reinforce conceptual understanding while building practical coding skills.

See **Teaching Innovation & Curriculum Development** section above for further details.

Sample module syllabus



- Schedule
- Requirements
- Resources
  - Documents
  - Simple Linear Regression
  - Multiple linear regression, interactions, qualitative predictors
  - Classification (logistic regression, generative models)
  - Resampling methods (CV, bootstrap) and Linear model selection and regularization
- Folder Structure

Description

This module introduces the skills required to design, implement, and test basic statistical learning methods, including regression, classification, and clustering, as well as validating models with resampling techniques. It compares the differences between modeling for prediction purposes and inference, exploring the trade-offs between prediction accuracy, model interpretability, and the bias-variance trade-off. Participants also gain exposure to key tools such as Pandas, NumPy, and scikit-learn.

Learning Outcomes

By the end of the module, participants will be able to:

- Implement and interpret the results from several supervised learning approaches for classification and regression.
- Use resampling methods such as cross-validation and bootstrapping to select and evaluate models.
- Understand the requirements for reproducible machine learning and ensure consistency across model implementations.
- Analyze the uncertainties and limitations associated with model results and understand the ethical implications of applying these models in real-world decision-making.
- Explain the trade-offs and considerations of various statistical methods to both technical and non-technical audiences.
- Apply `pandas`, `numpy`, and `scikit-learn` for data manipulation, model implementation, and evaluation.

Assignments

Participants should review the Assignment Submission Guide for instructions on how to complete assignments in this module.

Assignment 1

Assignment 2

Assignment 3

Assignment Due Dates

Assessment	Content	Due Date
Assignment 1	Classification (Sessions 1, 2)	May 18

Assessment	Content	Due Date
Assignment 2	Regression (Sessions 3, 4)	May 25
Assignment 3	Clustering & Resampling (Sessions 5, 6)	June 1

Contacts

Questions can be submitted to the *#cohort-6-help* channel on Slack

- Technical Facilitator: **Julia**. Questions can be sent via Slack
- Learning Support Staff: **Vishakh**. Questions can be sent via Slack
- Learning Support Staff: **Aditya**. Questions can be sent via Slack
- Learning Support Staff: **Gayathri**. Questions can be sent via Slack

Delivery of the Learning Module

This module will include live learning sessions and optional, asynchronous work periods. During live learning sessions, the Technical Facilitator will introduce and explain key concepts and demonstrate core skills. Learning is facilitated during this time. Before and after each live learning session, the instructional team will be available for questions related to the core concepts of the module. Optional work periods are to be used to seek help from peers, the Learning Support team, and to work through the homework and assignments in the learning module, with access to live help. Content is not facilitated, but rather, this time should be driven by participants. We encourage participants to come to these work periods with questions and problems to work through. Participants are encouraged to engage actively during the learning module. The key to developing the core skills in each learning module is through practice. The more participants engage in coding with the instructional team and apply the skills in each module, the more likely these skills will solidify.

The technical facilitator will introduce the concepts through a collaborative live coding session using the Python notebooks, which can be found under `/01_materials/notebooks/` . Slides can be found under `/01_materials/slides/` .

Schedule

- Week 1 will focus on intro and classification methods
- Week 2 will focus on regression methods
- Week 3 will focus on clustering and statistical inference topics

Requirements

- Participants are expected to have completed Shell, Git, and Python learning modules.
- Participants are encouraged to ask questions and collaborate with others to enhance learning.
- Participants must have a computer and an internet connection to participate in online activities.
- Participants must not use generative AI such as ChatGPT to generate code to complete assignments. It should be used as a supportive tool to seek out answers to questions you may have.
- We expect participants to have completed the steps in the onboarding repo.
- We encourage participants to default to having their camera on at all times and turning the camera off only as needed. This will greatly enhance the learning experience for all participants and provides real-time feedback for the instructional team.

Sample notebook

Install Packages and Import Dataset

In this notebook, we'll be working with the **Wisconsin Diagnostic Breast Cancer (WDBC)** dataset. This dataset is commonly used in machine learning tasks related to classification, and it contains several features derived from digitized images of breast tumor cells obtained via fine needle aspirates (FNAs).

Each row in the dataset represents one tumor, with a set of measurements calculated from the cell nuclei present in the image. The dataset has the following columns:

- **ID**: Unique identifier for each tumor sample.
- **Diagnosis**: The classification label for the tumor (Malignant or Benign).
- **Radius Mean, Texture Mean, Perimeter Mean, Area Mean**: Various statistical properties of the tumor.
- **Compactness, Concavity, Symmetry**: Other characteristics calculated from the shape and structure of the tumor cells.

The target column is **Diagnosis**, which we will try to predict based on the other features in the dataset.

This dataset was obtained from the UCI Machine Learning Repository, a well-known resource for datasets in the machine learning community.

```
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.colors as mcolors
from mpl_toolkits import mplot3d
```

```
cancer = pd.read_csv('dataset/wdbc.csv')
cancer
```

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_me
0	842302	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.300
1	842517	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.086
2	84300903	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.197
3	84348301	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.241
4	84358402	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.198
...	...	...	...	...	...	...	...	...	
564	926424	M	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.243
565	926682	M	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.144
566	926954	M	16.60	28.08	108.30	858.1	0.08455	0.10230	0.092
567	927241	M	20.60	29.33	140.10	1265.0	0.11780	0.27700	0.351
568	92751	B	7.76	24.54	47.92	181.0	0.05263	0.04362	0.000

569 rows × 32 columns

What is Pandas?

"But wait, what is that `pd.read_csv()` thing?" Well, it comes from **pandas**, a powerful Python library designed for data manipulation and analysis.

Pandas allows us to efficiently handle datasets, similar to how you would in Excel, but with far more capabilities. It can handle large datasets and perform complex transformations with just a few lines of code.

The `pd.read_csv()` function is used to load data from a CSV file into a **DataFrame**, which is like a table of data with rows and columns. Each column can hold different types of data (e.g., numbers, text), making it a versatile structure for analysis.

By using pandas in this notebook, we can:

- Load and explore the dataset easily.
- Identify missing or problematic data.
- Group, sort, and manipulate data efficiently.
- Prepare the data for machine learning or statistical analysis.

Inspect the data

The `.info()` method in pandas provides a concise summary of a DataFrame, including the dtype, column names, and non-null values. It's particularly useful for getting an overview of the dataset's structure and identifying any missing data.

```
cancer.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 569 entries, 0 to 568
Data columns (total 33 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   id                       569 non-null    int64
 1   diagnosis                569 non-null    object
 2   radius_mean              569 non-null    float64
 3   texture_mean             569 non-null    float64
 4   perimeter_mean           569 non-null    float64
 5   area_mean                569 non-null    float64
 6   smoothness_mean          569 non-null    float64
 7   compactness_mean         569 non-null    float64
 8   concavity_mean           569 non-null    float64
 9   concave points_mean      569 non-null    float64
 10  symmetry_mean            569 non-null    float64
 11  fractal_dimension_mean   569 non-null    float64
 12  radius_se                569 non-null    float64
 13  texture_se               569 non-null    float64
 14  perimeter_se             569 non-null    float64
 15  area_se                  569 non-null    float64
 16  smoothness_se            569 non-null    float64
 17  compactness_se           569 non-null    float64
 18  concavity_se             569 non-null    float64
 19  concave points_se        569 non-null    float64
 20  symmetry_se              569 non-null    float64
 21  fractal_dimension_se     569 non-null    float64
 22  radius_worst             569 non-null    float64
 23  texture_worst            569 non-null    float64
 24  perimeter_worst          569 non-null    float64
 25  area_worst               569 non-null    float64
 26  smoothness_worst         569 non-null    float64
 27  compactness_worst        569 non-null    float64
 28  concavity_worst          569 non-null    float64
 29  concave points_worst     569 non-null    float64
 30  symmetry_worst           569 non-null    float64
 31  fractal_dimension_worst  569 non-null    float64
 32  Unnamed: 32              0 non-null      float64
dtypes: float64(31), int64(1), object(1)
memory usage: 146.8+ KB
```

The `.unique()` method in pandas returns the unique values in a column. This method is useful for identifying all distinct values within a dataset, especially when you want to see the different categories or groups present in a column.

```
cancer["diagnosis"].unique()
```

```
array(['M', 'B'], dtype=object)
```

Clean data by renaming "M" to "Malignant" and "B" to "Benign" using the `.replace` method. The `.replace` method takes one argument: a dictionary that maps previous values to desired new values.

```
cancer["diagnosis"] = cancer["diagnosis"].replace({
    "M" : "Malignant",
    "B" : "Benign"
})
```

```
cancer["diagnosis"].unique()
```

```
array(['Malignant', 'Benign'], dtype=object)
```

To analyze the distribution of benign and malignant tumor observations in the dataset, we can use the `.groupby` and `.size` methods with pandas.

Group the Data by the Class Variable:

1. Use `.groupby` on the Class column to group the data by benign and malignant tumor observations. Count the Observations:
2. Apply the `.size` method to count the number of observations in each group.

Calculate Percentages:

3. Divide the counts by **the total number of observations** (found using `df.shape`[0]) and multiply by 100 to get the percentage.

```
100 * cancer.groupby("diagnosis").size() / cancer.shape[0]
```

```
diagnosis
Benign       62.741652
Malignant    37.258348
dtype: float64
```

The `.value_counts` method in pandas efficiently counts the occurrences of each unique value in a column.

```
cancer["diagnosis"].value_counts()
```

```
diagnosis
Benign       357
Malignant    212
Name: count, dtype: int64
```

By using the `normalize=True` argument, it can also return the fraction (or percentage) of each value, providing a quick and convenient way to analyze the distribution of data.

```
cancer["diagnosis"].value_counts(normalize=True)
```

```
diagnosis
Benign       0.627417
Malignant    0.372583
Name: proportion, dtype: float64
```

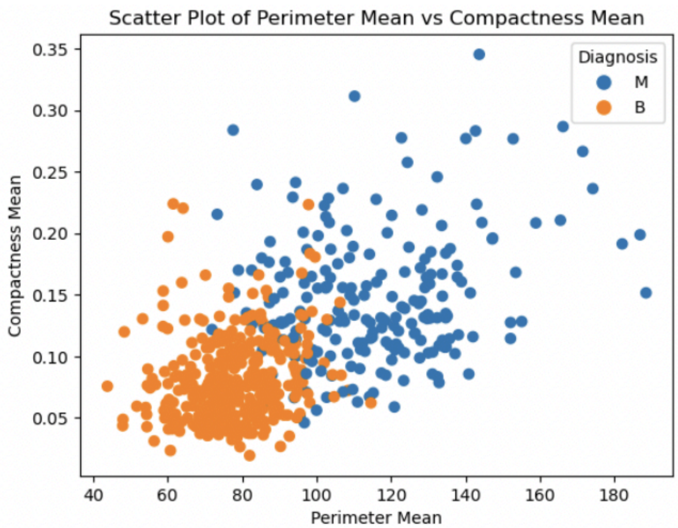
Visualizing the Data

Now, let's create a scatter plot to visualize the relationship between **perimeter mean** and **concavity mean**. This will help us see how these features relate to whether a tumor is benign or malignant.

The code below may look a bit intimidating, but don't worry! You don't need to understand every line right now. It uses matplotlib to create the plot and map different colors to the benign and malignant tumor groups.

Even if this feels complex, the key takeaway is seeing how we can visualize patterns in the data. Later, as you get more familiar with plotting, you can revisit this code to dissect it further.

```
# Create mapping between values and colors
labels = cancer["diagnosis"].unique().tolist()
colors = list(mcolors.TABLEAU_COLORS.keys())
color_map = {l: colors[i % len(colors)] for i, l in enumerate(labels)}

# Plot
plt.scatter(cancer["perimeter_mean"], cancer['concavity_mean'],
            color=cancer["diagnosis"].map(color_map))

# Create custom legend handles
handles = [plt.Line2D([0], [0], marker='o', color='w', label=label,
                      markersize=10, markerfacecolor=color_map[label])
           for label in labels]

# Add labels and legend
plt.xlabel('Perimeter Mean')
plt.ylabel('Concavity Mean')
plt.title('Scatter Plot of Perimeter Mean vs Concavity Mean')
plt.legend(handles=handles, title='Diagnosis')
plt.show()
```



It's okay if you don't grasp everything right away—focus on understanding the output and come back to the code when you're ready!

- **Malignant Observations**: These tend to be located in the upper right-hand corner of the plot (blue), indicating higher values for both concavity and perimeter.
- **Benign Observations**: These are generally found in the lower left-hand corner (orange), suggesting lower values for concavity and perimeter.

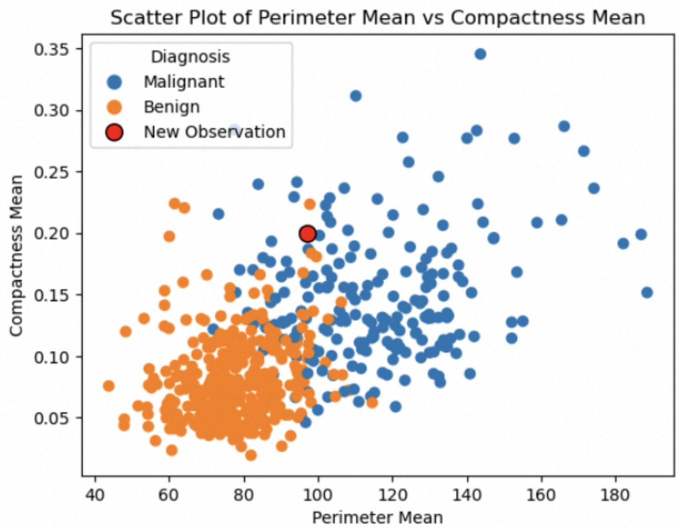
This visual pattern suggests that it may be possible to predict the diagnosis of new tumor images by using their concavity and perimeter mean values.

K-Nearest Neighbors Manually

In this section, we'll manually implement the K-nearest neighbors (KNN) algorithm to classify a new data point. The idea behind KNN is simple: given a new observation, the algorithm looks at the "K" closest points in the dataset and assigns the class (benign or malignant) based on the majority class of its neighbors.

We'll plot the existing data and add a new observation to the plot, then calculate the distances between the new point and the other points in the dataset to find its nearest neighbors.

```
# Plot existing data
plt.scatter(cancer["perimeter_mean"], cancer['concavity_mean'],
            color=cancer["diagnosis"].map(color_map))

# Create custom legend handles
handles = [plt.Line2D([0], [0], marker='o', color='w', label=label,
                      markersize=10, markerfacecolor=color_map[label])
           for label in labels]

# Add new observation
new_observation = {'perimeter_mean': 97, 'concavity_mean': 0.20}
plt.scatter(new_observation['perimeter_mean'], new_observation['concavity_mean'],
            color='red', edgecolor='black', s=100, label='New Observation')

# Add labels and legend
plt.xlabel('Perimeter Mean')
plt.ylabel('Concavity Mean')
plt.title('Scatter Plot of Perimeter Mean vs Concavity Mean')
plt.legend(handles=handles + [plt.Line2D([0], [0], marker='o', color='w',
                                          markerfacecolor='red', markeredgecolor='black',
                                          markersize=10, label='New Observation')],
           title='Diagnosis')
plt.show()
```



To find the K nearest neighbors to our new observation, we compute the distance from that new observation to each observation in our training data, and select the K observations corresponding to the smallest distance values. For example, suppose we want to use $K = 5$ neighbors to classify a new observation with mean perimeter 97 and concavity 0.20. We can calculate the distances between our new point and each of the observations in the training set to find the 5 neighbors that are nearest to our new point.

$$Distance = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}$$

In order to find the $K = 5$ nearest neighbors, we will use the `nsmallest` function from pandas.

```
new_obs_Perimeter = 97
new_obs_Concavity = 0.20
cancer["dist_from_new"] = (
       (cancer["perimeter_mean"] - new_obs_Perimeter) ** 2
     + (cancer["concavity_mean"] - new_obs_Concavity) ** 2
)**(1/2)

nearest_5 = cancer.nsmallest(5, "dist_from_new")[[
    "perimeter_mean",
    "concavity_mean",
    "diagnosis",
    "dist_from_new"
]]

nearest_5
```

	perimeter_mean	concavity_mean	diagnosis	dist_from_new
291	97.03	0.05940	Benign	0.143765
138	96.85	0.15390	Malignant	0.156924
15	96.73	0.16390	Malignant	0.272403
514	97.26	0.07486	Malignant	0.288548
54	97.26	0.05253	Malignant	0.298910

Tumor Data with Distance Calculations

Perimeter Mean	Concavity Mean	Distance	Diagnosis
97.03	0.05940	$\sqrt{(97.03 - 97)^2 + (0.05940 - 0.20)^2}$	Benign

96.85	0.15390	$\sqrt{+ (0.15390 - 0.20)^2}$	Malignant
96.73	0.16390	$\sqrt{(96.73 - 97)^2 + (0.16390 - 0.20)^2}$	Malignant
97.26	0.07486	$\sqrt{(97.26 - 97)^2 + (0.07486 - 0.20)^2}$	Malignant
97.26	0.05253	$\sqrt{(97.26 - 97)^2 + (0.05253 - 0.20)^2}$	Malignant

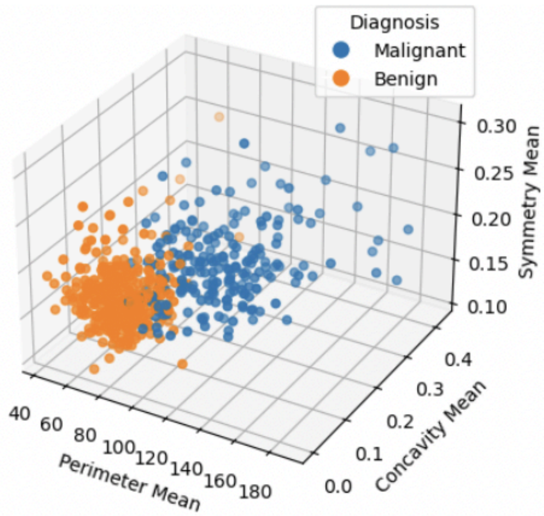
The result of this computation shows that 4 of the 5 nearest neighbors to our new observation are malignant

What if we have more than 2 variables?

So far, we've only looked at the relationship between two features: **perimeter mean** and **concavity mean**. But what if we want to include more features in our analysis? In that case, we can extend our scatter plot into three dimensions to visualize how a third feature (in this case, **symmetry mean**) interacts with the others.

This 3D scatter plot will help us see if adding another feature might give us more insight into classifying the data. By visualizing multiple features at once, we can start to understand how they work together to differentiate benign from malignant tumors.

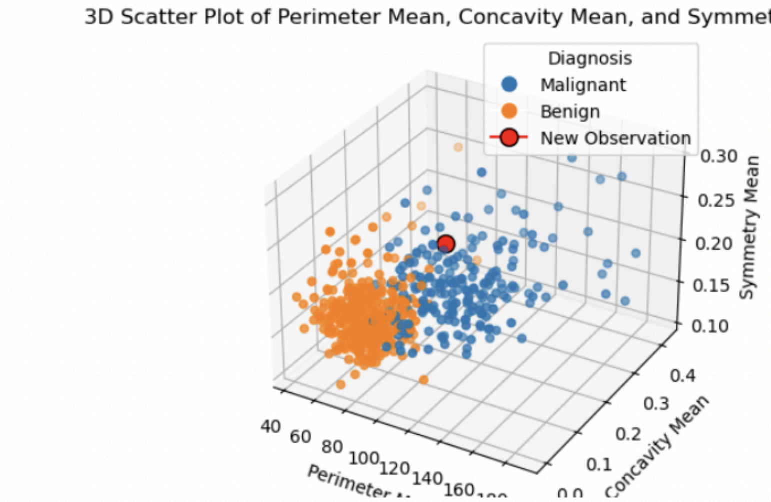
```
# Create mapping between values and colors
labels = cancer["diagnosis"].unique().tolist()
colors = list(mcolors.TABLEAU_COLORS.keys())
color_map = {l: colors[i % len(colors)] for i, l in enumerate(labels)}

# Create a 3D plot
ax = plt.axes(projection="3d")

# Plot data points with color corresponding to diagnosis
sc = ax.scatter3D(cancer['perimeter_mean'], cancer['concavity_mean'], cancer['symmetry_mean'],
                  c=cancer['diagnosis'].map(color_map), marker='o')

# Add axis labels
ax.set_xlabel('Perimeter Mean')
ax.set_ylabel('Concavity Mean')
ax.set_zlabel('Symmetry Mean')
ax.set_title('3D Scatter Plot of Perimeter Mean, Concavity Mean, and Symmetry Mean')

# Create custom legend handles
handles = [plt.Line2D([0], [0], marker='o', color='w', label=label,
                      markersize=10, markerfacecolor=color_map[label])
           for label in labels]

# Add legend
plt.legend(handles=handles, title='Diagnosis')

# Show plot
plt.show()
```



Suppose we want to use $K = 5$ neighbors to classify a new observation with a perimeter of 97, concavity of 0.20, and symmetry of 0.22.

$$Distance = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2}$$

```
# Create mapping between values and colors
labels = cancer["diagnosis"].unique().tolist()
colors = list(mcolors.TABLEAU_COLORS.keys())
color_map = {l: colors[i % len(colors)] for i, l in enumerate(labels)}

# Create a 3D plot
ax = plt.axes(projection="3d")

# Plot data points with color corresponding to diagnosis
sc = ax.scatter3D(cancer['perimeter_mean'], cancer['concavity_mean'], cancer['symmetry_mean'],
                  c=cancer['diagnosis'].map(color_map), marker='o')

# Define the new observation
new_observation = {'perimeter_mean': 97, 'concavity_mean': 0.20, 'symmetry_mean': 0.22}

# Plot the new observation
ax.scatter3D(new_observation['perimeter_mean'], new_observation['concavity_mean'],
             new_observation['symmetry_mean'], color='red', edgecolor='black',
             s=100, marker='o', label='New Observation')

# Add axis labels
ax.set_xlabel('Perimeter Mean')
ax.set_ylabel('Concavity Mean')
ax.set_zlabel('Symmetry Mean')
ax.set_title('3D Scatter Plot of Perimeter Mean, Concavity Mean, and Symmetry Mean')

# Create custom legend handles
handles = [plt.Line2D([0], [0], marker='o', color='w', label=label,
                      markersize=10, markerfacecolor=color_map[label])
           for label in labels]

# Add custom legend for new observation
handles.append(plt.Line2D([0], [0], marker='o', color='red', label='New Observation',
                          markersize=10, markeredgecolor='black'))

# Add legend
plt.legend(handles=handles, title='Diagnosis')
```




```
new_obs_Perimeter = 97
new_obs_Concavity = 0.20
new_obs_Symmetry = 0.22
cancer["dist_from_new"] = (
      (cancer["perimeter_mean"] - new_obs_Perimeter) ** 2
    + (cancer["concavity_mean"] - new_obs_Concavity) ** 2
    + (cancer["symmetry_mean"] - new_obs_Symmetry) ** 2
)**(1/2)
cancer.nsmallest(5, "dist_from_new")[[
    "perimeter_mean",
    "concavity_mean",
    "symmetry_mean",
    "diagnosis",
    "dist_from_new"
]]
```

	perimeter_mean	concavity_mean	symmetry_mean	diagnosis	dist_from_new
291	97.03	0.05940	0.1879	Benign	0.147305
138	96.85	0.15390	0.1957	Malignant	0.158795
15	96.73	0.16390	0.2303	Malignant	0.272597
514	97.26	0.07486	0.1561	Malignant	0.295539
54	97.26	0.05253	0.1616	Malignant	0.304562

Based on $K = 5$ nearest neighbors with these three predictors we would classify the new observation as malignant since 4 out of 5 of the nearest neighbors are malignant class.

K-Nearest Neighbours using `scikit-learn`

Coding the K-nearest neighbors algorithm from scratch in Python can become complex, particularly when dealing with multiple classes, more than two variables, or predicting the class for several new observations.

Fortunately, Python's `scikit-learn` package offers a built-in implementation of the KNN algorithm. This implementation simplifies the process, making our code more readable, accurate, and less prone to errors.

To ensure compatibility with pandas DataFrames when using scikit-learn, it's important to configure the package appropriately using the `set_config` function before starting with KNN.

```
from sklearn import set_config

# Output dataframes instead of arrays
set_config(transform_output="pandas")
```

Step 1. import the `KNeighborsClassifier` from the `sklearn.neighbors` module.

```
from sklearn.neighbors import KNeighborsClassifier
```

Similar to above, we will use perimeter mean and concavity mean as predictors and $K = 5$ neighbors to build our classifier.

```
cancer_train = cancer[["diagnosis", "perimeter_mean", "concavity_mean"]]
cancer_train
```

	diagnosis	perimeter_mean	concavity_mean
0	Malignant	122.80	0.30010
1	Malignant	132.90	0.08690
2	Malignant	130.00	0.19740
3	Malignant	77.58	0.24140
4	Malignant	135.10	0.19800
...	...	...	...
564	Malignant	142.00	0.24390
565	Malignant	131.20	0.14400
566	Malignant	108.30	0.09251
567	Malignant	140.10	0.35140
568	Benign	47.92	0.00000

569 rows × 3 columns

Step 2. Create a model object for K-nearest neighbors classification by creating a `KNeighborsClassifier` instance, specifying that we want to use $K = 5$ neighbors

```
knn = KNeighborsClassifier(n_neighbors=5)
knn
```

```
▾ KNeighborsClassifier i ?
KNeighborsClassifier()
```

Step 3. Fit the model on the breast cancer data. The `X` argument is used to specify the data for the predictor variables, while the `y` argument is used to specify the data for the response variable.

Here,

- `X=cancer_train[["perimeter_mean", "concavity_mean"]]` *to specify both Perimeter and Concavity means are to be used as the predictors.*
- `y=cancer_train["diagnosis"]` *to specify that diagnosis is the response variable (the one we want to predict)*

```
knn.fit(X=cancer_train[["perimeter_mean", "concavity_mean"]], y=cancer_train["diagnosis"])
```

```
▾ KNeighborsClassifier i ?
KNeighborsClassifier()
```

Step 4. Make a prediction on a new observation by calling `.predict` on the classifier object, passing the new observation itself.

```
new_obs = pd.DataFrame({"perimeter_mean": [97], "concavity_mean": [0.20]})
knn.predict(new_obs)
```

```
array(['Malignant'], dtype=object)
```

Prediction is the same as what we manually computed above!

Sample assignment

Assignment 1

You only need to write one line of code for each question. When answering questions that ask you to identify or interpret something, the length of your response doesn't matter. For example, if the answer is just 'yes,' 'no,' or a number, you can just give that answer without adding anything else.

We will go through comparable code and concepts in the live learning session. If you run into trouble, start by using the help `help()` function in Python, to get information about the datasets and function in question. The internet is also a great resource when coding (though note that **no outside searches are required by the assignment!**). If you do incorporate code from the internet, please cite the source within your code (providing a URL is sufficient).

Please bring questions that you cannot work out on your own to office hours, work periods or share with your peers on Slack. We will work with you through the issue.

Classification using KNN

Let's set up our workspace and use the **Wine dataset** from `scikit-learn`. This dataset contains 178 wine samples with 13 chemical features, used to classify wines into different classes based on their origin.

The **response variable** is `class`, which indicates the type of wine. We'll use all of the chemical features to predict this response variable.

```
# Import standard libraries
import pandas as pd
import numpy as np
import random
import matplotlib.pyplot as plt
import matplotlib.colors as mcolors
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import recall_score, precision_score
from sklearn.model_selection import cross_validate
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import accuracy_score


from sklearn.datasets import load_wine

# Load the Wine dataset
wine_data = load_wine()

# Convert to DataFrame
wine_df = pd.DataFrame(wine_data.data, columns=wine_data.feature_names)

# Bind the 'class' (wine target) to the DataFrame
wine_df['class'] = wine_data.target

# Display the DataFrame
wine_df
```

Question 1:

Data inspection

Before fitting any model, it is essential to understand our data. **Use Python code** to answer the following questions about the **Wine dataset**:

(i) How many observations (rows) does the dataset contain?

```
# Your answer here
```

(ii) How many variables (columns) does the dataset contain?

```
# Your answer here
```

(iii) What is the 'variable type' of the response variable `class` (e.g., 'integer', 'category', etc.)? What are the 'levels' (unique values) of the variable?

```
# Your answer here
```

(iv) How many predictor variables do we have (Hint: all variables other than `class`)?

```
# Your answer here
```

You can use `print()` and `describe()` to help answer these questions.

Question 2:

Standardization and data-splitting

Next, we must preform 'pre-processing' or 'data munging', to prepare our data for classification/prediction. For KNN, there are three essential steps. A first essential step is to 'standardize' the predictor variables. We can achieve this using the scaler method, provided as follows:

```
# Select predictors (excluding the last column)
predictors = wine_df.iloc[:, :-1]

# Standardize the predictors
scaler = StandardScaler()
predictors_standardized = pd.DataFrame(scaler.fit_transform(predictors), columns=predictors.columns)

# Display the head of the standardized predictors
print(predictors_standardized.head())
```

(i) Why is it important to standardize the predictor variables?

> Your answer here...

(ii) Why did we elect not to standard our response variable `Class`?

> Your answer here...

(iii) A second essential step is to set a random seed. Do so below (Hint: use the random.seed function). Why is setting a seed important? Is the particular seed value important? Why or why not?

> Your answer here...

(iv) A third essential step is to split our standardized data into separate training and testing sets. We will split into 75% training and 25% testing. The provided code randomly partitions our data, and creates linked training sets for the predictors and response variables.

Extend the code to create a non-overlapping test set for the predictors and response variables.

```
# set a seed for reproducibility
np.random.seed(123)

# split the data into a training and testing set. hint: use train_test_split !

# Your code here ...
```

Question 3:

Model initialization and cross-validation

We are finally set to fit the KNN model.

Perform a grid search to tune the `n_neighbors` hyperparameter using 10-fold cross-validation. Follow these steps:

1. Initialize the KNN classifier using `KNeighborsClassifier()`.
2. Define a parameter grid for `n_neighbors` ranging from 1 to 50.
3. Implement a grid search using `GridSearchCV` with 10-fold cross-validation to find the optimal number of neighbors.
4. After fitting the model on the training data, identify and return the best value for `n_neighbors` based on the grid search results.

```
# Your code here...
```

Question 4:

Model evaluation

Using the best value for `n_neighbors`, fit a KNN model on the training data and evaluate its performance on the test set using `accuracy_score`.

```
# Your code here...
```

Criteria

Criteria	Complete	Incomplete
Data Inspection	Data is inspected for number of variables, observations and data types.	Data inspection is missing or incomplete.
Data Scaling	Data scaling or normalization is applied where necessary (e.g., using `StandardScaler`).	Data scaling or normalization is missing or incorrectly ap
Model Initialization	The KNN model is correctly initialized and a random seed is set for reproducibility.	The KNN model is not initialized, is incorrect, or lacks a ra
Parameter Grid for `n_neighbors`	The parameter grid for `n_neighbors` is correctly defined.	The parameter grid is missing or incorrectly defined.
Cross-Validation Setup	Cross-validation is set up correctly with 10 folds.	Cross-validation is missing or incorrectly set up.
Best Hyperparameter (`n_neighbors`) Selection	The best value for `n_neighbors` is identified using the grid search results.	The best `n_neighbors` is not selected or incorrect.
Model Evaluation on Test Data	The model is evaluated on the test data using accuracy.	The model evaluation is missing or uses the wrong metri

Submission Information

🚨 **Please review our [Assignment Submission Guide](#)** 🚨 for detailed instructions on how to format, branch, and submit your work. Following these guidelines is crucial for your submissions to be evaluated correctly.

Note:

If you like, you may collaborate with others in the cohort. If you choose to do so, please indicate with whom you have worked with in your pull request by tagging their GitHub username. Separate submissions are required.

Submission Parameters:

- Submission Due Date: `11:59 PM - 05/18/2025`
- The branch name for your repo should be: `assignment-1`
- What to submit for this assignment:
  - This Jupyter Notebook (assignment_1.ipynb) should be populated and should be the only change in your pull request.
- What the pull request link should look like for this assignment: `https://github.com/<your_github_username>/LCR/pull/<pr_id>`
  - Open a private window in your browser. Copy and paste the link to your pull request into the address bar. Make sure you can see your pull request properly. This helps the technical facilitator and learning support staff review your submission easily.

Checklist:

- [ ] Created a branch with the correct naming convention.
- [ ] Ensured that the repository is public.
- [ ] Reviewed the PR description guidelines and adhered to them.
- [ ] Verify that the link is accessible in a private browser window.

If you encounter any difficulties or have questions, please don't hesitate to reach out to our team via our Slack at `#cohort-6-help`. Our Technical Facilitators and Learning Support staff are here to help you navigate any challenges.