

IMPORTANT: This syllabus form should be submitted to OAA (gsbs_academic_affairs@uth.tmc.edu) a week before the start of each semester.

NOTE to STUDENTS: If you need any accommodations related to attending/enrolling in this course, please contact the Graduate School's 504 Coordinator, Natalie Sirisaengtaksin, PhD. We ask that you notify GSBS in advance (preferably at least 3 days before the start of the semester) so we can make appropriate arrangements.

Term and Year: Spring 2026 Course Number and Course Title: GS01 1022: Statistical Communication, Consulting and Collaborative Data Science Credit Hours: 2 Prerequisites: Consent of Instructor Meeting Location: MD Anderson Cancer Center Building/Room#: 1MC12.3311	Program Required Course: No Approval Code: Yes (If yes, the Course Director or the Course Designee will provide the approval code.) Audit Permitted: No Classes Begin: January 15, 2026 Classes End: April 16, 2025 Final Exam Week: Not Applicable, no final exam				
Class Meeting Schedule <table border="1"> <thead> <tr> <th>Day</th> <th>Time</th> </tr> </thead> <tbody> <tr> <td>Thursday</td> <td>2:00-4:00 p.m.</td> </tr> </tbody> </table>		Day	Time	Thursday	2:00-4:00 p.m.
Day	Time				
Thursday	2:00-4:00 p.m.				
Course Director Name and Degree: Wenyi Wang, PhD Title: Professor Department: Bioinformatics and Computational Biology, and Biostatistics Institution: MDACC Email Address: wwang7@mdanderson.org Contact Number: 650-224-0952 Course Co-Director/s: (if any) Name and Degree: Ye Zheng, PhD Title: Assistant Professor Department: Bioinformatics and Computational Biology Institution: MDACC Email Address: yzheng8@mdanderson.org	Instructors 1. Wenyi Wang, PhD Institution: MDACC Email Address: wwang7@mdanderson.org 2. Ye Zheng, PhD Institution: MDACC Email Address: yzheng8@mdanderson.org 3. Liang Li, PhD Institution: MDACC Email Address: lli15@mdanderson.org				

NOTE: Office hours are available by request. Please email me to arrange a time to meet.

Teaching Assistant:

TBA

Course Description:

This course is designed to help students build essential statistical communication skills that are often underemphasized in traditional training. It focuses on preparing students to collaborate effectively with researchers from diverse backgrounds by teaching them how to:

- Effectively interview collaborators to understand their research questions and objectives,
- Articulate mutual goals and expectations specific to statistical consulting and interdisciplinary collaboration,
- Define statistical objectives and deliverables that can guide the research process, and
- Provide regular progress updates in a clear and actionable manner.

Through two core components—a consulting clinic and a project-based learning curriculum—students gain hands-on experience in applying statistical and data science knowledge while refining their communication and collaboration skills.

In the **consulting** clinic, students offer pro bono statistical consulting services to UTHealth/MDA community researchers, working under instructor supervision to apply these skills in real-world settings. The **project-based learning** component focuses on project scoping, collaborative practices, and reproducible workflows, equipping students with tools to translate research questions into actionable solutions and foster productive interdisciplinary collaborations.

This course will prepare students for professional collaborations in academic and industry settings.

Prerequisites: Students are expected to have knowledge in basic Statistics Inference, Probability, and Linear Regression. Prior programming experience in R or Python is required.

Note: Students who are interested in the course but not sure whether they meet the prerequisites can contact the course directors. If the registration number goes above 12, the course directors will make decisions on who to admit to the course.

Priorities will be given to QS program 1st and 2nd year PhD students and those who meet the prerequisites. The directors will provide guidance on preparing the prerequisites, through taking other basic statistical courses in class or through Coursera courses, for those who are interested in taking it in future

Textbook/Supplemental Reading Materials

- Irizarry, R. A. (2019). Introduction to data science: Data analysis and prediction algorithms with R. Chapman and Hall/CRC. <https://rafalab.dfci.harvard.edu/dsbook/>
- Wickham (2017) R for Data Science <http://r4ds.had.co.nz/>
- Yu, B. and Barter, R. L. (2024) Veridical Data Science. <https://vdsbook.com/>
- Venables, W. N., & Ripley, B. D. (2013). Modern applied statistics with S-PLUS. Springer Science & Business Media

Course Objectives:

Upon successful completion of this course, students will be able to effectively communicate with clients, translate data science challenges into structured workflows, and provide actionable solutions. They will have developed the skills to conduct effective interviews with collaborators, establish shared goals and expectations, and deliver clear, well-structured reports that help clients advance their research while ensuring they understand the statistical solutions and their relevance to the project.

Specific Learning Objectives:

1. Gain hands-on experience applying statistical and data science techniques to solve real-world research problems.
2. Learn important scientific communication skills to interview collaborators, define goals, and provide clear updates and deliverables to help students make progress with their research.
3. Work effectively with researchers from diverse fields to foster productive partnerships.
4. Work effectively with researchers from diverse fields to foster productive partnerships.
5. Translate research questions into structured statistical workflows and actionable solutions.

Student responsibilities and expectations:

Students enrolled in this course will be expected to perform the following activities each week.

1. Review and analyze materials related to statistical consulting and data science projects, including assigned readings and case studies.
2. Complete weekly assignments, including written reports on real-world data science projects, demonstrating understanding and application of statistical and machine learning techniques.
3. Actively participate in the Consulting Clinic, working with client researchers to address real-world data challenges and providing statistical guidance and actionable solutions.
4. Prepare and deliver oral presentations based on project/consulting clinic progress and findings, highlighting problem-solving strategies and outcomes.
5. Engage in class discussions, and review sessions, contributing insights and feedback on peers' consulting approaches.

Students are expected to complete all assigned readings and project work during and after class. While collaboration on project ideas and discussions is encouraged, all written reports and presentations must be original work. Plagiarism, failure to properly cite sources, or unethical behavior during consulting sessions or evaluations will not be tolerated and may result in dismissal from the course and GSBS disciplinary action.

Grading System: Letter Grade (A-F)	
Student Assessment and Grading Criteria: <i>(May include the following:)</i>	
Percentage	Description
Homework (35 %)	Four written reports on data science projects (Projects 1-4 worth 5%, 10%, 10%, 10% separately), evaluated on the quality of literature review, statistical application, and solution development.
Presentation (30 %)	Oral presentations on data science projects and consulting clinic (data science projects 20%, where 5% each project, consulting clinic 10%), graded on the clarity and effectiveness of communicating project outcomes and recommendations/solutions. <i>Failures of projects are allowed as long as work is conducted in a respectful manner.</i>
Workshop or Breakout-Session (25 %)	Graded on contributions during consulting clinic activities, and feedback from both client researchers and the instructor.
Participation and/or Attendance (10 %)	Assessed based on active engagement and regular attendance in class and consulting sessions.

CLASS SCHEDULE

Date	Duration (Hour(s) taught by lecturer)	Lecture Topic	Lecturer/s
01/15/2026	2 hours	Introduction and Class Organization; Consulting 101	Wenyi Wang/Ye Zheng
01/22/2026	2 hours	Project 1 release and discussion (Topic: Prepare a plain language presentation of a classic statistical methodology)	Wenyi Wang
01/29/2026	2 hours	Project 1 presentations; Consulting clinic presentation (The topics for the consulting clinic will be gathered through online forms accessible to the public within the MD Anderson and UTHealth)	Wenyi Wang
02/05/2026	2 hours	Project 2 release and discussion (Topic: Read and summarize the statistics or machine learning method(s) and discuss the choice of	Ye Zheng

		data used in a scientific paper); Consulting clinic presentation	
02/12/2026	2 hours	Project 2 discussion; Consulting clinic presentation	Ye Zheng
02/19/2026	2 hours	Project 2 discussion; Consulting clinic presentation	Ye Zheng
02/26/2026	2 hours	Project 3 release and discussion (Topic: Present a current statistical application to an intro stat class); Consulting clinic presentation	Liang Li
03/05/2026	2 hours	Project 3 discussion & presentation	Liang Li
03/12/2026	2 hours	No Class, Spring Break	
03/19/2026	2 hours	Project 4 release and discussion (Topic: Reproduce an applied statistics paper); Consulting clinic presentation	Wenyi Wang
03/26/2026	2 hours	Project 4 discussion; Consulting clinic presentation	Wenyi Wang
04/02/2026	2 hours	Project 4 discussion; Consulting clinic presentation	Wenyi Wang
04/09/2026	2 hours	Project 4 discussion; Consulting clinic presentation	Wenyi Wang
04/16/2026	2 hours	Wrapping up	Wenyi Wang/Ye Zheng