Data Science 6101-10, Spring 2026
Introduction to Data Science
6:10 - 8:40pm Wednesday, 1957 E Street, Room 314

Instructor: Dr. Darcy Steeg Morris
Office Hours: after class or by appointment
E-mail: dsteeg@gwu.edu

Course Description: More than a booming industry, Data Science emerges at the forefront in shaping every aspect of our daily life. This course introduces the basic concepts and underlying principles, and provides guidance in practical applications using team projects with real life data walking through a typical data science life cycle – from developing SMART questions and EDA, building models, to assembling a compelling team presentation. The class covers estimation, hypothesis testing, and various modeling techniques from basic to advanced. R programming and Git workflow are used throughout the course.

Prerequisite(s): You are expected to have a basic understanding of introductory statistics and R coding, either through online resources or previous experiences.

Learning Outcomes: As a result of completing this course, students will be able to:

- construct research questions based on SMART principles on real life data driven problems;
- choose appropriate statistical tests in data analyses;
- describe and choose appropriate models for data analyses;
- apply the fundamental principles of data science to solve real life data driven problems;
- evaluate the validity and value of data models;
- use R programming to perform data analyses, and present results in oral and written reports from such analyses;
- collaborate with teammates using Git system to perform data analyses in R, and maximize the collaborative development cycle efficiency; and
- utilize new or existing libraries in R programming to perform data analysis.

Textbooks:

No one book fully captures the variety of skills and knowledge required in this course. Consequently, we have pre-selected chapters from a variety of sources (links provided for those available free online).

- [AoRP] Matloff (2011), *The Art of R Programming.**
- [AoDS] Peng and Matsui (2016), *Art of Data Science.**
- [OIS] Diez et al. (2015), *OpenIntro Statistics.*

  https://leanpub.com/os (set price to $0)
- [IS] Illowsky et al. (2018), *Introductory Statistics.*

  https://openstax.org/details/books/introductory-statistics

- [ISLR] James et al. (2013), *An Introduction to Statistical Learning*.

  https://www.statlearning.com/s/ISLR-Seventh-Printing.pdf

- [DDS] Schutt and O'Neil (2014), *Doing Data Science*.

  Search O'Reilly Books through https://libguides.gwu.edu/ebooks/disciplines

Course Resources:

This course uses R for computing, Git for file sharing, and some online tutorials. Complete these tasks as early as possible to maximize your learning experience in this class.

- Install R on your laptop: https://www.r-project.org/.
- Install RStudio on your laptop: https://posit.co/products/open-source/rstudio/.
- Sign up for an account on DataCamp: https://www.datacamp.com/ with your GWU email.
- Set up a free GitHub account: https://github.com/.
- Install GitHub desktop on your laptop: https://desktop.github.com/.

Data Science Helpdesk:

The Data Science Helpdesk can help you resolve any technical issues relating to software installation, updates, and crashes. Students can also create a service ticket to initiate an online support session by sending an email to dshelpdesk@gwu.edu. The Helpdesk will be available for in-person resolution at Desk 1, Room 304, Samson Hall Monday-Thursday. See https://helpdesk.datasci.land for more information and in-person times.

Grading:

Homework Assignments	20%
Blackboard Discussion Boards	10%
Quizzes	30% (10% each)
Project 1	20%
Project 2	20%

- *Homework Assignments*: Students will complete homework assignments about every week to demonstrate applied understanding of concepts. These homework assignments are based on readings and lectures. Homework is submitted on Blackboard with R assignments written in RMarkdown. Homework may be discussed with classmates, but submissions should represent individual work and written up independently – homeworks in violation of this policy will receive a zero.

- *Discussion Board*: Throughout the course, students will interact with classmates to discuss course topics on Blackboard. Posts and responses will be based on concepts learned in lectures, readings, or online resources. Students are expected to post original ideas and respond substantively to peers.

- *Quizzes*: Students will complete three, non-cumulative quizzes in the course to demonstrate proficiency with course concepts. Keeping up with readings and lectures will support successfully completing quizzes. Quizzes are taken on Blackboard during class time in the classroom on the student's labtop/tablet, are closed-book/closed-note, and timed (75-90 minutes).

- *Class Projects:* Students will complete two projects during the course working as a team of 2-4 students to practice applying data science principles and methods to various problems. This opportunity will help develop teamwork, communication, and technical skills.

- Students are expected to spend a minimum of 7.5 hours a week on out-of-class/independent learning, assuming the knowledge of basic statistics that is a prerequisite to take this class.. Students are requried to attend all lectures during class time (2.5 hours a week) – if attendance becomes an issue, the instructor reserves the right to record attendance and factor it into the final course grade.

Use of Electronic Course Materials and Class Recordings:

Students are encouraged to use electronic course materials, including recorded class sessions, for private personal use in connection with their academic program of study. Electronic course materials and recorded class sessions should not be shared or used for non-course related purposes unless express permission has been granted by the instructor. Students who impermissibly share any electronic course materials are subject to discipline under the Student Code of Conduct. Contact the instructor if you have questions regarding what constitutes permissible or impermissible use of electronic course materials and/or recorded class sessions. Contact Disability Support Services at www.disabilitysupport.gwu.edu if you have questions or need assistance in accessing electronic course materials.

Academic Integrity Code:

Academic integrity is an essential part of the educational process, and all members of the GW community take these matters very seriously. As the instructor of record for this course, my role is to provide clear expectations and uphold them in all assessments. Violations of academic integrity occur when students fail to cite research sources properly, engage in unauthorized collaboration, falsify data, and otherwise violate the Code of Academic Integrity. If you have any questions about whether particular academic practices or resources are permitted, you should ask me for clarification. If you are reported for an academic integrity violation, you should contact Student Rights and Responsibilities (SRR) to learn more about your rights and options in the process. Consequences can range from failure of assignment to expulsion from the University and may include a transcript notation. For more information, refer to the SRR website at www.studentconduct.gwu.edu/academic-integrity, email rights@gwu.edu, or call 202-994-6757.

Use of Generative AI:

Work submitted for evaluation is represented as the student's own intellectual product. Students may not submit content (e.g., ideas, text, code, images) for evaluation that was generated, in whole or in part, by Generative Artificial Intelligence (GAI) tools (such as ChatGPT and other large language models). Doing so without instructor's explicit permission constitutes cheating under the Code of Academic Integrity and is therefore prohibited. For example, you cannot use GAI for writing your project papers or Blackboard discussion posts. *Exceptions: You can use GAI to research R functions (e.g. syntax, minimum working examples) to implement a procedure for HW and projects, but the final code must be your own; and you can use GAI to brainstorm project topics and search for project datasets.*

Students are permitted to use GAI tools to generate content that is not submitted to an instructor for evaluation. For example, using GAI tools to study for examinations, tests, and quizzes is permitted. Likewise, GAI tools may be used for learning, studying, and brainstorming.

The use of GAI tools during any assessment (e.g., examination, test, quiz) whether taken in the classroom or elsewhere, constitutes cheating under the Code of Academic Integrity and is therefore prohibited.

GW Campus Emergency Information:

GW Emergency Services: 202-994-6111

For situation-specific instructions, refer to GW's Emergency Procedures guide: www.safety.gwu.edu/emergency-response-handbook.

GW Alert

GW Alert is an emergency notification system that sends alerts to the GW community. GW requests students, faculty, and staff maintain current contact information by logging on to alert.gwu.edu. Alerts are sent via email, text, social media, and other means, including the Guardian app. The Guardian app is a safety app that allows you to communicate quickly with GW Emergency Services, 911, and other resources. Learn more at safety.gwu.edu.

Protective Actions

GW prescribes four protective actions that can be issued by university officials depending on the type of emergency. All GW community members are expected to follow directions according to the specified protective action. The protective actions are Shelter, Evacuate, Secure, and Lockdown (details below). Learn more at safety.gwu.edu/gw-standard-emergency-statuses.

University Policy on Observance of Religious Holidays:

Students must notify faculty during the first week of the semester in which they are enrolled in the course, or as early as possible, but no later than three weeks prior to the absence, of their intention to be absent from class on their day(s) of religious observance. If the holiday falls within the first three weeks of class, the student must inform faculty in the first week of the semester. For details and policy, see www.provost.gwu.edu/policies-procedures-and-guidelines.

Support for Students Outside the Classroom:

Writing Center

GW Writing Center cultivates confident writers in the University community by facilitating collaborative, critical, and inclusive conversations at all stages of the writing process. Working alongside peer mentors, writers develop strategies to write independently in academic and public settings. Appointments can be booked online. See https://gwu.mywconline.com.

Academic Commons

Academic Commons is the central location for academic support resources for GW students. To schedule a peer tutoring session for a variety of courses visit go.gwu.edu/tutoring. Visit https://academiccommons.gwu.edu for study skills tips, finding help with research, and connecting with other campus resources. For questions email academiccommons@gwu.edu.

Disability Support Services (DSS): 202-994-8250

Any student who may need an accommodation based on the potential impact of a disability should contact the Disability Support Services to establish eligibility and to coordinate reasonable accommodations. For additional information see: www.disabilitysupport.gwu.edu.

Student Health Center: 202-994-5300, 24/7

The Student Health Center (SHC) offers medical, counseling/psychological, and psychiatric services to GW students. More information about the SHC is available at www.healthcenter.gwu.edu. Students experiencing a medical or mental health emergency on campus should contact GW Emergency Services at 202-994-6111.

Course Schedule:

The weekly coverage and number of homework discussions/assignments and discussion/homework/quiz due dates may change as they depend on the progress of the class. The instructor reserves the right to make changes to this course schedule as necessary.

Week	Module	Content
Week 1 (1/14)	1: DS & Stats Basics	• *Lecture*: Git, R/Rstudio; What is Data Science? • *Suggested Reading*: *DDS* Preface & Chapter 1.
Week 2 (1/21)		• *Lecture*: R & Git practice; SMART Questions / Data Science lifecycle. • *Suggested Reading*: *AoRP* Intro & Getting Started; *AoDS* Chapters 1 – 3. • **HW #1 Due: DataCamp R Completion** and **HW #2 Due: AI Artwork**
Week 3 (1/28)		• *Lecture*: EDA; Basic Statistics; Normal Distribution; Project Overview. • *Suggested Reading*: *AoDS* Chapter 4; *OIS* Chapter 1-3, 4.1 • **HW #3 Due: R Basics** and **BB Discussion Due**
Week 4 (2/4)	2: Inference & Testing	• *Lecture*: CLT; z- and t- score; Confidence Intervals. • *Suggested Reading*: *OIS* Chapters 5.1–5.2; *IS* Chapters 6–8. • **HW #4 Due: EDA in R** and **Project 1 Proposal Due**
Week 5 (2/11)		• *Lecture*: Hypothesis Testing; p-values. • *Suggested Reading*: *OIS* Chapters 5.3; *IS* Chapters 9–11. • **Quiz 1**
Week 6 (2/18)		• *Lecture*: z-test, t-test, anova, chi-squared (χ^2) test. • *Suggested Reading*: *OIS* Chapters 5.3; *IS* Chapters 9–11. • **HW #5 Due: Confidence Intervals in R** and **BB Discussion Due**
Week 7 (2/25)		• *Lecture*: Linear Model Part I. • *Suggested Reading*: *ISLR* Chapter 3. • **HW #6 Due: Hypothesis Testing in R**
Week 8 (3/4)		• **Project 1 Presentations in class**; **Project 1 Paper due 3/8**
Week 9 (3/11)	3: Regression & Other Modeling	• *Spring Break (No Class)*
Week 10 (3/18)		• *Lecture*: Linear Model Part II • *Suggested Reading*: *OIS* Chapter 8 & 9. • **Quiz 2**
Week 11 (3/25)		• *Lecture*: Generalized Linear Models and Logistic Regression • *Suggested Reading*: *ISLR* Chapters 4.1–4.3. • **HW #7 Due: LM in R** and **Project 2 Proposal Due**
Week 12 (4/1)		• *Lecture*: Logistic Regression Part II • *Suggested Reading*: *ISLR* Chapters 4.1–4.3. • **HW #8 Due: More LM in R** and **BB Discussion Due**
Week 13 (4/8)		• *Lecture*: Classification and Regression Trees • *Suggested Reading*:*ISLR* Chapter 8.1, 8.3. • **HW #9 Due: LR in R**
Week 14 (4/15)		• *Lecture*: Student-Voted Topic • **Quiz 3**
Week 15 (4/22)		• **Project 2 Presentations in class**; **Project 2 Paper due 4/26**