Rutgers University
The State University of New Jersey
Department of Economics
Department of Mathematical Sciences

Fall 2025

Course Title: Introduction to Data Science

Course Number: 50:220:122/50:960:185 (cross-listed)

Instructors: Dr. I-Ming Chiu

Email: ichiu@camden.rutgers.edu

Classroom: BSB 336, 11:10 AM-12:30 PM, Tuesday & Thursday

Office: Armitage Hall #435

Office Hours: 1:00-2:00 PM (Tuesday/Thursday or by appointment)

Course Material: visit the Canvas site
Academic Calendar: https://camden.rutgers.edu/registrar/catalogs-calendars/2025-2026

Prerequisites:

High school algebra, computer literacy and familiarity with Microsoft Office products (Word, and Excel, etc.)

Required Readings:

Michael Freeman and Joel Ross (2019), Programming Skills for Data Science/Start Writing Code to Wrangle, Analyze, and Visualize Data with R (PSDS), Addison-Wesley. (ISBN: 978-0135133101).
John D. Kelleher and Brendan Tierney (2018), Data Science (DS), MIT Press (ISBN: 978-0262535434).

Recommended Readings:

1) Data Storage & Wrangling
   Bradley C. Boehmke (2016), Data Wrangling with R, Springer.
2) Data Exploration and Visualization
   John W. Tukey (1977), Exploratory Data Analysis, Addison Wesley.
   Hadley Wickham (2016), ggplot2: Elegant Graphics for Data Analysis 2nd ed., Springer.
3) Probability, Statistical Inference, and Computational Statistics
   Babak Shahbaba (2012), Biostatistics with R/An Introduction to Statistics through Biological Data, Springer.
4) Basic Programming (functional programming)
   Alain F. Zuur, Elena N. Ieno, and Erik H.W.G. Meesters (2009), A Beginner's Guide to R, Springer.

Course Description:
Data Science is a multi-disciplinary field that involves data exploration, computer programming and statistical modeling. Data scientists help decision makers extract reliable information and uncover knowledge from structured and unstructured data. Introduction to Data Science class prepares students to become sufficiently fluent in both inferential thinking and computational skill. The class consists of the following five main components: data wrangling (management), data visualization, data mining/machine learning, inferential thinking, and statistical computations.

Computation Using R & RStudio (an IDE, Integrated Development Environment, for R)
https://www.youtube.com/watch?v=Icawuhf0Yqo (for Mac)
https://www.youtube.com/watch?v=hxj0UG4boGU (for PC)
*Install R before install RStudio. **Please notice that the most recent R and RStudio versions are 4.5.1 and 2025.05.1+513, respectively.

We will also use a general-purpose programming language Python occasionally when needed

Other Tools:
There is a brief coverage to learn how to retrieve data from Data Bases. MySQL will be used to build a local database on your machine.

Objectives of the Course:
1) The ability to collect and organize data for further analysis.
2) The ability to explore and visualize data.
3) Understanding the basics of probability theory and inferential statistics.
4) The ability to write functions (functional programming) and understand the basics of machine learning algorithms.
5) The ability to conduct research projects and write reports using real-world data.

Evaluation:
10% - DataCamp training (an online learning platform, please visit https://www.datacamp.com)
25% - Homework Assignments (3~5)
40% - Two Midterm Exams
25% - Final Exam/Project
5% - Participation/Quiz in class meetings and others (extra credit points)
** Term grades will be based on the final distribution of the above grading weights.

Academic Conduct:
Make up exams will be given only upon prior notice. I request prior knowledge of any expected absence from an exam. If this is not feasible, you can document a valid reason for missing the exam. Unexcused absence on any exam will result in a grade of zero. Dishonesty in seeking an excused absence or in the examination process will result in a grade of zero on the exam involved and in university discipline. To review the academic integrity policy, go to
https://deanofstudents.camden.rutgers.edu/academic-integrity

Disability Services:

Rutgers University welcomes students with disabilities into all of the University's educational programs. In order to receive consideration for reasonable accommodations, a student with a disability must contact the appropriate disability services office at the campus where you are officially enrolled, participate in an intake interview, and provide documentation: https://ods.rutgers.edu/students/documentation-guidelines.

If the documentation supports your request for reasonable accommodations, your campus's disability services office will provide you with a Letter of Accommodations. Please share this letter with your instructors and discuss the accommodations with them as early in your courses as possible. To begin this process, please complete the Registration form at https://webapps.rutgers.edu/student-ods/forms/registration

Here is the link to the Office of Disability Service: https://success.camden.rutgers.edu/disability-services

Learning Center- Learning Specialists and Tutoring

I am committed to making course content accessible to all students. The Learning Center provides Learning Specialists who can help you build a learning plan based on your strengths and needs. Tutors, study groups and more services are available you for free. Many services are available in virtual formats and after normal business hours. In addition, if English is not your first language and this causes you concern about the course, the Learning Center can help. You can learn more about these services by calling 856-225-6442, emailing rclc@camden.rutgers.edu or learningcenter@camden.rutgers.edu, or visiting the website https://learn.camden.rutgers.edu. You can schedule an appointment with Learning Specialist to create a plan of action using the website.

Complaints

Rutgers University-Camden is committed to providing quality services, a great education and an engaged and caring experience for our students. Sometimes problems arise, and students may find that they would like to file a complaint about their experience or a particular situation. To file a complaint, students can complete the form at this link and someone will connect with you to discuss your complaint, explain options and to address the issue that was raised. Students do have the option of filing a complaint anonymously, but then there will be no way for the office handling the complaint to be able to let the student know how it was addressed. Filling out a form will allow students to understand all options and the different ways an issue can be addressed. The form is located here: https://deanofstudents.camden.rutgers.edu/reporting

More student resources information can be found from the following link:

https://studentaffairs.camden.rutgers.edu/student-resource-list

Course Outline:

Topic 1	Data Types/Structures and Introduction to R
Topic 2	Population, Samples, and Statistics
Topic 3	Data Wrangling Basics and the Use of dplyr Package
Topic 4	Write your own Functions
Exam I	**Date: TBA in the class**
Topic 5	Exploratory Data Analysis (EDA): Visualization Fundamentals
Topic 6	Exploratory Data Analysis (EDA): Visualization w/ ggplot2 Package
Topic 7	Introduction to Probability Theory
Topic 8	Learning Inferential Statistics using Simulation and Bootstrap Methods
Exam II	**Date: TBA in the class**
Topic 9[1]	Supervised Machine Learning: Linear Regression Model
Topic 10	Supervised Machine Learning: K-Nearest Neighbors (KNN)
Topic 11	Unsupervised Machine Learning: K-Means Clustering
Topic 12	Reproducible Research using R Markdown
Additional Topic	Introduction to Relational Database using MySQL
Final Exam/Project (school schedule)	11:30 AM-2:20 PM, Tuesday, December 16

[1] Three machine learning (ML) algorithms are introduced in Topic 9, 10 and 11.