

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

Fall 2020

Course Title: Introduction to Data Science (Index: 01312)

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

Instructors: Prof. I-Ming Chiu

Classroom: Remote Teaching (11:10 AM-12:30 PM via Zoom, Tuesday & Thursday)

Office: Camden Towers 214B

Office Hours: 1:00-2:00 PM (Thursday or by appointment via Zoom)

Academic Calendar: <https://registrar.camden.rutgers.edu/academic-calendar-2020-2021>

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 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
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, inferential thinking, and statistical computations.

Objectives of the Course:

- 1) The ability to import data and organize data for 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 (procedural programming) and understand the basics of mining algorithms.
- 5) The ability to write research project using real-world data.

Evaluation:

- 10% - DataCamp training (an online learning platform, please visit <https://www.datacamp.com/>)
 - 30% - Homework Assignments (3~5)
 - 30% - Midterm Exam
 - 30% - Final Exam/Project
 - 5% - Participation in Zoom meetings (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. More detailed information can be found at the following site: <https://academicintegrity.rutgers.edu>

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>

Course Outline:

Topic 1	Data Types/Structures and Introduction to R
Topic 2	Population, Sample and Statistics
Topic 3	Data Wrangling Basics and the Use of dplyr Package
Topic 4	Write your own Functions
Topic 5	Exploratory Data Analysis (EDA): Visualization Fundamentals
Midterm Exam	Date: TBA in the class
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
Topic 9	Data Mining Algorithms: Case Studies (I)
Topic 10	Data Mining Algorithms: Case Studies (II)
Additional Topic	Reproducible Research Project using R Markdown
Additional Topic	Version Control using Git and GitHub
Final Exam (school schedule)	11:30 AM-2:20 PM, Tuesday, December 15