Syllabus

Introduction to Real World Data Management

2 credits/ 4 ECTS

Dr. Alexandru Cernat

Video lecture by
Dr. Alexandru Cernat

September 29 – November 17, 2025

Short Course Description

Data is omnipresent in the contemporary world, coming in different shapes and sizes: from survey data to found data. In order to make use of such data through analysis it is necessary first to import and clean it. This is often one of the most time consuming and difficult parts of data analysis. In this course you will learn both the conceptual steps needed in preparing data for analysis as well as the practical skills to do this. The course will cover all the essential skills needed to prepare data be it survey data, administrative data or found data.

Course Objectives

By the end of the course, students will…

- Understand the stages involved in preparing data for analysis
- Understand the concept of tidy data
- Understand the basics of using R
- Know how to write their own functions and loop over them
- Know how to import and export data
- Know how to clean data in R
- Know how to merge data
- Know how to manipulate textual data
- Know how to manipulate date/time data
- Know how to use tables and graphs to explore data

Prerequisites

Basic knowledge of R. Prior experience with working with quantitative data.

Class Structure and Course Concept

This is an online course using a flipped classroom design. It covers the same material and content as an on-site course but runs differently. In this course, you are responsible for watching video-recorded lectures and reading the required literature for each unit prior to participating in mandatory weekly one-hour online meetings where students have the chance to discuss the materials from a unit with the instructor. Although this is an online course where students have more freedom in when they engage with the course materials, students are expected to spend the same amount of time overall on all activities in the course – including preparatory activities (readings, studying), inclassactivities (watching prerecorded videos, attending the live online meetings), and followup activities (working on assignments and exams) – as in an on-site course. As a rule of thumb you can expect to spend approximately 3h/week on in-class-activities and 9 hours per week on out-of-class activities (preparing for class, readings, assignments, projects, studying for quizzes and exams). Therefore, the workload in all courses will be approximately 12h/week. Please note that the actual workload will depend on your personal knowledge.

Mandatory Weekly Online Meetings

Sec. 1: Mondays, September 29 -November 17, 2025, 12 pm ET/5 pm CET (4 pm UK time)

Meetings will be held online through Zoom. Follow the link to the meeting sessions on the course website. If video participation via Internet is not possible, arrangements can be made for students to dial in and join the meetings via telephone.

In preparation for the weekly online meetings, students are expected to watch the lecture videos and read the assigned literature before the start of the meeting. In addition, students are encouraged to post questions about the materials covered in the videos and readings of the week in the forum before the meetings (deadline for posting questions is Sunday 12 pm ET/5 pm CET (4 pm UK time)

Students have the opportunity to use the BigBlueButton feature in Canvas to connect with peers outside the scheduled weekly online meetings (e.g., for study groups). Students are not required to use BigBlueButton and can of course use other online meeting platforms such as Google Hangouts, Skype or Microsoft Teams.

Daylight saving time ends in the USA on November 2, 2025. Therefore, look carefully at the times of meetings and deadlines!

Grading

Grading will be based on:

- 4 fortnightly homework assignments (worth 60% total)
- Participation in discussion during the weekly online meetings and submission of questions to the forum (deadline: Sunday, 3:00 PM EDT/9:00 PM CEST) (demonstrating understanding of the required readings and video lectures (10% of grade)
- A final project (30% of grade)

A+ 100 - 97
A 96 - 93
A- 92 - 90
B+ 89 - 87
B 86 - 83
B- 82 - 80
Etc.

The grading scale is a base scale. Variations for grading on a scale are at the discretion of the instructor.

The final grade will be communicated under the assignment "Final Grade" in the Canvas course. Please note that the letter grade written in parentheses in Canvas is the correct final grade. The point-grade displayed alongside the letter grade is irrelevant and can be ignored. Dates of when assignment will be due are indicated in the syllabus. Extensions will be granted sparingly and are at the instructor's discretion.

Technical Equipment Needs

The learning experience in this course will mainly rely on the online interaction between the students and the instructors during the weekly online meetings. Therefore, we encourage all students in this course to use a web camera and a headset. Decent quality headsets and web cams are available for less than $20 each. We ask students to refrain from using built-in web cams and speakers on their desktops or laptops. We know from our experience in previous online courses that this will reduce the quality of video and audio transmission and therefore will decrease the overall learning experience for all students in the course. In addition, we suggest that students use a wire connection (LAN), if available, when connecting to the online meetings. Wireless connections (WLAN) are usually less stable and might be dropped.

Long Course Description

Data is ubiquitous in the contemporary world. It comes in a variety of shapes and sizes: surveys, administrative data or found data. Often we want to use this data to bring insights by applying different types of statistical analyses. Unfortunately, most often the data we are interested in do not come prepared for the analysis we want to carry out. This can be due to its format, due to missing cases or just because it

captures information in a way that we cannot use in our analysis. In this course you will learn both the conceptual and practical aspects of importing and manipulating data in order to be used both for exploratory and more advanced statistical analyses.

The course will first cover the main concepts needed to prepare real world data using R. We will start by understanding the steps we need to follow in order to prepare data for analysis. Then we will develop the core skills in R such as working with the different types of objects, such as data frames. We will then cover how to use techniques to make our work with data more efficient, for example by using loops or by applying functions over variables or data frames. After covering the main concepts and skills we will concentrate on data management. Here we will discuss how to manipulate data such as selecting cases/variables, recoding variables or reshaping datasets. We will then go on to learn how to explore the data using tables and graphics. Finally, we will cover the topic of cleaning and exploring text data as well as time data.

By the end of the course the students will be able to work with multiple types of data and be able to manipulate them in order to prepare them for analysis. They will know the main steps needed to achieve this in an efficient way.

The course will be divided in four topics. Each one will be covered in two weeks. The first week will cover the online course and the reading materials. In the second week students will have to prepare a project based on what they learned in the first week.

Readings

Mandatory Readings

Wickham, H., & Grolemund, G. (2017). R for Data Science. O'Reilly UK Ltd.

(: https://r4ds.hadley.nz/) Grolemund, G. (2017). Hands-On Programming with R. Write Your Own

Functions and Simulations. O'Reilly UK Ltd. (https://rstudio-education.github.io/hopr/)

Complementary Readings

Wickham, H. (2015). Advanced R. Boca Raton, FL: CRC Press. (https://adv-r.hadley.nz/)

Academic Conduct

Clear definitions of the forms of academic misconduct, including cheating and plagiarism, as well as information about disciplinary sanctions for academic misconduct may be found at

https://www.president.umd.edu/sites/president.umd.edu/files/documents/policies/III100A.pdf (University of Maryland)

Knowledge of these rules is the responsibility of the student and ignorance of them does not excuse misconduct. The student is expected to be familiar with these guidelines before submitting any written work or taking any exams in this course. Lack of familiarity with these rules in no way constitutes an excuse for acts of misconduct. Charges of plagiarism and other forms of academic misconduct will be dealt with very seriously and may result in oral or written reprimands, a lower or failing grade on the assignment, a lower or failing grade for the course, suspension, and/or, in some cases, expulsion from the university.

Accommodations for Students with Disabilities

In order to receive services, students at the University of Maryland must contact the Accessibility & Disability Service (ADS) office to register in person for services. Please call the office to set up an appointment to register with an ADS counselor. Contact the ADS office at 301.314.7682; https://www.counseling.umd.edu/ads/.

Please inform the instructor of any accommodations that may be required on or before the first meeting.

Course Evaluation

In an effort to improve the learning experience for students in our online courses, students will be invited to participate in an online course evaluation at the end of the course. Participation is entirely voluntary and highly appreciated.

Sessions

Mondays, September 29 -November 17, 2025, 12 pm ET/5 pm CET (4 pm UK time)

Week 1:

Video lecture: available Monday, September 22, 2025

Online meeting: Monday, September 29, 2025, 12 pm ET/5 pm CET

Online question week 1: due Sunday, September 28, 2025, 12 pm ET/5 pm CET

Required Readings (chapter numbers from online books): Chapters 1-7, 9 and 11 from “Hands-On Programming with R” (https://rstudioeducation.github.io/hopr/) Chapters 26-27 from "R for Data Science" (https://r4ds.hadley.nz/iteration) Wickham, H. (2014). Tidy Data. Journal of Statistical Software, 59(10). https://doi.org/10.18637/jss.v059.i10

Recommended Readings (chapter numbers form online book):
Chapters 3, 5, 7 from "R for Data Science" (https://r4ds.hadley.nz/)

Week 2:

Video lecture: available Monday, September 29, 2025

Online meeting: Monday, October 6, 2025, 12 pm ET/5 pm CET
Online question week 2: due Sunday, October 5, 2025, 12 pm ET/5 pm CET

Graded project 1: due Wednesday, October 8, 2025, 12 pm ET/5 pm CET

Week 3:

Video lecture: available Monday, October 6, 2025

Online meeting: Monday, October 13, 2025, 12 pm ET/5 pm CET

Online question week 3: due Sunday, October 19, 2025, 12 pm ET/5 pm CET

Required Readings (chapter numbers form online book):
Chapters 5, 10-13, 15 from “R for Data Science.” (http://r4ds.had.co.nz/)

Week 4:

Video lecture: available Monday, October 13, 2025

Online meeting: Monday, October 20, 2025, 12 pm ET/5 pm CET

Online question week 3: due Sunday, October 19, 2025, 12 pm ET/5 pm CET

Graded project 2: due Wednesday, October 22, 2025, 12 pm ET/5 pm CET

Week 5:

Video lecture: available Monday, October 20, 2025

Online meeting: Monday, October 27, 2025, 11 am ET/5 pm CET

Online question week 4: due Sunday, October 26, 2025, 11 am ET/5 pm CET

Required Readings (chapter numbers form online book):
Chapters 2, 10-12 from "R for Data Science"
(https://r4ds.hadley.nz/) **Recommended Readings**
(chapter numbers form online book): Chapters 27 and
29 from “R for Data Science.” (http://r4ds.had.co.nz/)
Please keep in mind that winter time begins in the UK on October 26, 2025 and clocks are turned back 1 hour. Winter time begins in the USA on November 2, 2025. Therefore, look carefully at the times of meetings and deadlines!

Week 6:

Video lecture: available Monday, October 27, 2025

Online meeting: Monday, November 3, 2025, 12 pm ET/5 pm CET

Online question week 6: due Sunday, November 2, 2025, 12 pm ET/5 pm CET

Graded project 3: due Wednesday, November 5, 2025, 12 pm ET/5 pm CET

Week 7:

Video lecture: available Monday, November 3, 2025

Online meeting: Monday, November 10, 2025, 12 pm ET/5 pm CET

Online question week 7: due Sunday, November 9, 2025, 12 pm ET/5 pm CET

CET Graded project 4: due Wednesday, November 12, 2025, 12 pm ET/5 pm CET

Required Readings (chapter numbers form online book): Chapters 2, 10-12 from "R for Data Science" (https://r4ds.hadley.nz/)

Week 8:

Online meeting: Monday, November 17, 2025, 12 pm ET/5 pm CET

Online question week 7: due Sunday, November 16, 2025, 12 pm ET/5 pm CET

Final graded project:

Due Monday, November 24, 2025, 12 pm ET/5 pm CET