

Syllabus

Introduction to Real World Data Management

2 credits / 4 ECTS

Dr. Alexandru Cernat

Video lecture by

Dr. Alexandru Cernat

September 28 – November 23, 2026

Short Course Description

Data is omnipresent in the contemporary world, coming in different shapes and sizes: from survey data to found data. To use such data through analysis, you must first 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 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 before 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), in-class activities (watching prerecorded videos, attending the live online meetings), and follow-up 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 28 – November 23, 2026, 16:00–17:00 UK / 11:00–12:00 Eastern. On October 26 only: 16:00–17:00 UK / 12:00–13:00 EDT. No meeting on October 12 (Fall Break).

Meetings will be held online through Zoom. Follow the link to the meeting sessions on the course website. If video participation via the 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 at 16:00 UK / 11:00 Eastern (12:00 EDT on October 25)).

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 or Microsoft Teams.

Clock changes: UK clocks go back on October 25; US Eastern clocks go back on November 1. UK meeting and deadline times remain fixed. EDT and EST below identify Eastern daylight and standard time.

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 (Sunday, 16:00 UK / 11:00 Eastern; 12:00 EDT on October 25), 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 assignments will be due are indicated in the syllabus. Extensions will be granted sparingly and are at the instructor's discretion.

Technical Equipment Needs

This course will mainly rely on online interaction between students and instructors during weekly online meetings. Therefore, we encourage all students in this course to use a web camera and a headset. Decent quality headsets and webcams are available for less than \$20 each. We ask students to refrain from using built-in webcams and speakers on their desktops or laptops. From our experience in previous online courses, this reduces video and audio quality and therefore decreases the overall learning experience for all students in the course. In addition, we suggest that students use a wired 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 gain insights by applying different types of statistical analyses. Unfortunately, the data we are most interested in often do not come prepared for the analysis we want to carry out. This can be because of its format, missing cases, or because it captures information in a way we cannot use for our analysis. In this course, you will learn both the conceptual and practical aspects of importing and manipulating data 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 in R. We will start by understanding the steps we need to follow to prepare data for analysis. Then we will develop core R skills, such as working with different types of objects, including 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 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 manipulate them to prepare them for analysis. They will know the main steps needed to do this efficiently.

The course will be divided into 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., Çetinkaya-Rundel, M., & Grolemund, G. (2023). R for Data Science (2nd ed.). O'Reilly.

[Online book](#)

Grolemund, G. (2014). Hands-On Programming with R: Write Your Own Functions and Simulations. O'Reilly.

[Online book](#)

Complementary Readings

Wickham, H. (2019). Advanced R (2nd ed.). Chapman & Hall/CRC.

[Online book](#)

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

[University of Maryland Code of Academic Integrity](#)

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. They 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

Students at the University of Maryland requesting disability accommodations should submit the ADS registration application, schedule a meeting with a disability specialist, and provide a self-report and supporting documentation. Contact the Accessibility & Disability Service at 301.314.7682 or visit [the ADS website](#).

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 28 – November 23, 2026. Meetings are always 16:00–17:00 UK / 11:00–12:00 Eastern, except October 26: 12:00–13:00 EDT.

Fall Break: October 12–13, 2026 (Monday–Tuesday). No meeting on October 12; meetings resume October 19.

All questions and assignments are due at 16:00 UK time. This is 11:00 Eastern except during October 25–31, when it is 12:00 EDT. Each dated entry below gives both times.

Week 1

Video lecture: available Monday, September 21, 2026

Online meeting: Monday, September 28, 2026

16:00–17:00 UK / 11:00–12:00 EDT

Online question week 1: due Sunday, September 27, 2026

16:00 UK / 11:00 EDT

Required Readings

Chapter numbers and links refer to the online editions.

Hands-On Programming with R: [1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [7](#), [9](#), [11](#).

R for Data Science (2nd ed.): [Ch. 25 Functions](#); [Ch. 26 Iteration](#).

Wickham, H. (2014). Tidy Data. Journal of Statistical Software, 59(10).
<https://doi.org/10.18637/jss.v059.i10>

Week 2

Video lecture: available Monday, September 28, 2026

Online meeting: Monday, October 5, 2026

16:00–17:00 UK / 11:00–12:00 EDT

Online question week 2: due Sunday, October 4, 2026

16:00 UK / 11:00 EDT

Graded project 1: due Wednesday, October 7, 2026

16:00 UK / 11:00 EDT

Week 3

Video lecture: available Monday, October 12, 2026

Online meeting: Monday, October 19, 2026

16:00–17:00 UK / 11:00–12:00 EDT

Online question week 3: due Sunday, October 18, 2026

16:00 UK / 11:00 EDT

Required Readings

Chapter numbers and links refer to the online editions.

R for Data Science (2nd ed.): [Ch. 3 Data transformation](#); [Ch. 5 Data tidying](#); [Ch. 7 Data import](#); [Ch. 16 Factors](#); [Ch. 19 Joins](#).

Week 4

Video lecture: available Monday, October 19, 2026

Online meeting: Monday, October 26, 2026

16:00–17:00 UK / 12:00–13:00 EDT

Online question week 4: due Sunday, October 25, 2026

16:00 UK / 12:00 EDT

Graded project 2: due Wednesday, October 28, 2026

16:00 UK / 12:00 EDT

Week 5

Video lecture: available Monday, October 26, 2026

Online meeting: Monday, November 2, 2026

16:00–17:00 UK / 11:00–12:00 EST

Online question week 5: due Sunday, November 1, 2026

16:00 UK / 11:00 EST

Required Readings

Chapter numbers and links refer to the online editions.

R for Data Science (2nd ed.): [Ch. 1 Data visualization](#); [Ch. 9 Layers](#); [Ch. 10 Exploratory data analysis](#); [Ch. 11 Communication](#).

Recommended Readings

Chapter numbers and links refer to the online editions.

R for Data Science (2nd ed.): [Ch. 28 Quarto](#); [Ch. 29 Quarto formats](#).

Week 6

Video lecture: available Monday, November 2, 2026

Online meeting: Monday, November 9, 2026

16:00–17:00 UK / 11:00–12:00 EST

Online question week 6: due Sunday, November 8, 2026

16:00 UK / 11:00 EST

Graded project 3: due Wednesday, November 11, 2026

16:00 UK / 11:00 EST

Week 7

Video lecture: available Monday, November 9, 2026

Online meeting: Monday, November 16, 2026

16:00–17:00 UK / 11:00–12:00 EST

Online question week 7: due Sunday, November 15, 2026

16:00 UK / 11:00 EST

Required Readings

Chapter numbers and links refer to the online editions.

R for Data Science (2nd ed.): [Ch. 14 Strings](#); [Ch. 15 Regular expressions](#); [Ch. 17 Dates and times](#).

Week 8

Video lecture: available Monday, November 16, 2026

Online meeting: Monday, November 23, 2026

16:00–17:00 UK / 11:00–12:00 EST

Online question week 8: due Sunday, November 22, 2026

16:00 UK / 11:00 EST

Graded project 4: due Wednesday, November 25, 2026

16:00 UK / 11:00 EST

Final Graded Project

Final graded project: due Monday, December 7, 2026

16:00 UK / 11:00 EST