# COMPUTER PROGRAMING I

**IE University**
Professor: **JAVIER CEREZO LAFUENTE**
E-mail: jcerezo@faculty.ie.edu

Academic year: 23-24
Degree course: FIRST
Semester: 2º
Category: COMPULSORY
Number of credits: 3.0
Language: English

## PREREQUISITES

Students should have taken Programming 0. There are no other prerequisites.

## SUBJECT DESCRIPTION

In this course, students will continue learning about data analysis in R. The goal of the course is to build knowledge across four dimensions:

(1) the R programming language

(2) general programming best practices

(3) data analysis techniques

(4) financial and economic concepts.

We will explore several more advanced applications, drawing examples from Finance and Economics whenever possible.

## OBJECTIVES AND SKILLS

R is a programming language that is particularly strong in statistics and what is now called “data science”, i.e. analysis of datasets in a broad sense. The rise of “Big Data” has popularized the use of R, previously confined to statistics, and expanded its capabilities into machine learning and other fields that were previously beyond its scope.
In a more tangible manner, R will be useful as a tool to manipulate, merge, and analyze datasets, and will substitute Excel and other visual interfaces both when the size of the data is large and when automatization is required.
The R community has developed libraries that enhance the basic R capabilities, and that we will cover throughout in the course, particularly tidyverse and ggplot2.

## METHODOLOGY

This course does not require sophisticated mathematical knowledge nor extensive programming experience. However, the nature of the material is somewhat technical, so it is required from the students to follow the activities and readings in order to benefit from the course.
Students are welcome to schedule a meeting with the professor or communicate with him via email / message in case they require help.
Since the course involves a lot of programming, students should bring their computers to every class. The recommended IDE (integrated development environment) for the class is RStudio, although students are welcome to use a classic editor (vim, emacs, etc.) and a terminal if they wish. RStudio is available for free for Windows, OSX and Linux (RStudio).

| Teaching methodology | Weighting | Estimated time a student should dedicate to prepare for and participate in |
|---|---|---|
| Lectures | 13.33 % | 10 hours |
| Discussions | 6.67 % | 5 hours |
| Exercises | 53.33 % | 40 hours |
| Group work | 13.33 % | 10 hours |
| Other individual studying | 13.33 % | 10 hours |
| TOTAL | 100.0 % | 75 hours |

## PROGRAM

## SESSION 1 (LIVE IN-PERSON)

Introduction to R programming environment, basic commands and concepts.
In this class we will learn which environments work best for R programming, with a focus on RStudio. We will also learn the basic R commands and the design philosophy of the language. We will also get familiar with some databases that will serve as example throughout the course, such as those in ourworldindata.org, finance.yahoo.com and in fred.stlouisfed.org.

## SESSION 2 (LIVE IN-PERSON)

Loops and vectors
Loops are one of the basic structures of all programming languages, and vectors are one of the basic data types of all programming languages. In R, they are both particularly important, and we will devote time to them early on in the course because they will serve as a basis for the following lessons.

## SESSION 3 (LIVE IN-PERSON)

Matrices and lists
Matrices are data structures that mimic the properties of algebraic matrices. They allow R to tackle problems in numerical analysis and more broadly to organize numerical data efficiently.

Lists are another fundamental data structure of most languages. They are also important in R, so we will introduce them early, with particular attention to indexing.

## SESSION 4 (LIVE IN-PERSON)

Simulated data and dataframes

Simulated data plays an important role in scientific computing; we will learn how to generate and use synthetic data that we later can feed into our models.

Dataframes are one of the most used data structures in modern data science. They are used in R, Python and other languages and allow for quick data transformation and manipulation.

## SESSION 5 (LIVE IN-PERSON)

Data import; tidy data

Data input/output is a nontrivial part of data analysis, and R provides several tools to minimize friction during this task. In order to practice those tools we will use the Titanic dataset from kaggle.com, which features real data from the passengers of the ship and is an old favourite of machine learning competitions. The data has numerical and categorical variables, which allows for several types of format manipulation.

## SESSION 6 (LIVE IN-PERSON)

Strings

Strings are a fundamental type of variable of every programming language. In R strings are called ‘characters’ and, although the number of things one can do with a string seems limited, the package string from tidyverse will prove otherwise.

## SESSION 7 (LIVE IN-PERSON)

Functions and flow control

Functions are one of the most important features of R and of programming languages in general. They are so important that a whole style of programming (functional pro- gramming) is named after them. R, together with more exotic languages such as Haskell or Lisp.

Flow control is a key element of all programming languages and consists of checking logical conditions and steering the execution of a program based on that.

## SESSION 8 (LIVE IN-PERSON)

Midterm

## SESSION 9 (LIVE IN-PERSON)

Data transformation: dyplr, logical operators, filter, arrange, select, mutate…

Dyplr is one of the most important libraries in the tidyverse: it enables handling and manipulating large datasets in an elegant way and it facilitates the comprehension of other languages such as python / pandas and SQL. In this class we will learn how it works by applying filters, selecting columns, arranging rows, and other operations into the dataset of historical gold prices, which we will download from Yahoo Finance.

## SESSION 10 (LIVE IN-PERSON)

Data wrangling

In this class we will explore the pipe operator, which enables concatenation of other operations in a clean and efficient way. For that we will replicate the examples in Hadley Wickham's book using the default diamonds and mtcars datasets. We will then move on to calculate some common statistcs by writing functions, which are an essential feature of most programming languages.

## SESSION 11 (LIVE IN-PERSON)

Exploratory Data Analysis with ggplot2

ggplot2 is the most important graphical library in R. As opposed to matplotlib in Python, gnuplot, or matlab, ggplot2 has a syntax that is particularly suited to operations of aggregation; this makes it initially less explicit and intuitive but in the long run facilitates exploratory data analysis without need for intermediate operations. We will practice ggplot2's most common visualizations with labor market data that we will download from fred.stlouisfed.org

## SESSION 12 (LIVE IN-PERSON)

Communicating insights with ggplot2

As mentioned, ggplot2 is the most important graphic library in R. In this lecture we will expand into its usage and comment on good and bad ways to present data. Nowadays we are flooded with visualizations and infographics that are not always displayed in a sensible or clear manner, so we will analyze some examples from fivethirtyeight and other websites to understand good design principles behind visualizations.

## SESSION 13 (LIVE IN-PERSON)

Relational data (join command)

Relational databases are at the center of data system design nowadays. R is well equipped to combine, filter and reorganize datasets, including a "join" command in SQL style. In order to build up some practice we will combine several datasets from fred.stlouisfed.org including commodity prices in different countries and timeframes, and we will then calculate correlation matrices to understand how those prices move together during different macro environments.

## SESSION 14 (LIVE IN-PERSON)

Equations and data tables

Many problems in statistics and math involve solving the roots of a polynomial or, more generally, the zeros of a function. R is well-equipped to solve these type of problems.

Data tables are a powerful tool to deal with datasets. They present an alternative approach to that of the tidyverse and have a better performance at the expense of a more complicated syntax. We will cover them only superficially.

## SESSION 15 (LIVE IN-PERSON)

Final exam

## BIBLIOGRAPHY

### Recommended

- Hadley Wickham. *R for Data Science.* O'Reilly. ISBN 1491910399 (Printed)

null

- Gareth James. *An introduction to Statistical Learning.* Springer. ISBN 1461471370 (Printed)

null

- Joseph Adler. *R in a Nutshell.* O'Reilly. ISBN 1449312084 (Printed)

null

## EVALUATION CRITERIA

You will need to submit the individual assignments on your own, but you are welcome to work in groups and help each other. It is important that you understand the code you upload into Blackboard even if you have used a few lines from a classmate or from online resources. For the final project, the recommendation is to form groups of 2 to 4 people.
The assignments are due by the specified due dates; no late work will be accepted. The answers will be available to view on Blackboard after the due date passes.
For the final group project, you will apply an approach that you learned in the class to study a real-world question that interests you, preferably in the field of finance and economics. The project that you choose should be challenging. You will present these projects in the last weeks of the course.

| Criteria | Percentage | Comments |
|---|---|---|
| Class Participation | 15 % | |
| Individual Work | 50 % | |
| Group Presentation | 35 % | |
| Examen Final | 0 % | |

## PROFESSOR BIO

Professor: **JAVIER CEREZO LAFUENTE**

E-mail: jcerezo@faculty.ie.edu

### JAVIER CEREZO LAFUENTE

Javier Cerezo is an adjunct professor at IE University. He holds an MSc in aerospace engineering from UPM and an MBA from The University of Chicago. He has developed his career in Airbus, where he worked in aeroelasticity and structural dynamics, McKinsey, where he led strategy and operations projects in Latin America, and Amazon, where he was product manager. He now works as Associate Director of Data Science in Novartis.

OTHER INFORMATION
Email: jcerezo@faculty.ie.edu
Office hours: by appointment
Phone number: +34 653 40 28 48

## OTHER INFORMATION

GENERAL OBSERVATIONS

Each student has four attempts over two consecutive academic years to pass this course. Dates and location of the final exam will be posted in advance and will not be changed. Students must attend at least 70% of the sessions. Students who do not comply with the 70% attendance rule will receive a 0.0 on their first and second attempts and go directly to the third one (they will need to enroll in this course again the following academic year). Students who are in the third or fourth attempt should contact the professor during the first two weeks of the course.

ATTENDANCE

Attendance is mandatory at IE University, as it is an essential factor of IE ?s learning methodology. While we do closely monitor attendance in each course, we also consider our students responsible for their own agenda and commitments, as adult university students. With that in mind, each student may miss up to 30% of the sessions within a given course and still maintain the possibility of passing that given course. This 30% "buffer" is to be used for any absences, such as: illnesses, personal emergencies, commitments, official/governmental matters, business and/or medical appointments, family situations, etc. Students should manage their various needs, and situations that may arise, within that 30% buffer. If a student is absent to more than the allowed 30% of the sessions (regardless of the reason), s/he will obtain a 0.0 grade for that course in both the ordinary and extraordinary calls of the current academic year, and s/he will have to retake the course during the following academic year. Having established the rule, we strongly discourage to use this buffer as granted, we highly recommend to attend 100% of the classes as it will improve your learning outcomes, it will increase the class performance and it might improve your participation grade. Extreme cases involving emergencies such as: extended hospitalizations, accidents, serious illnesses and other contexts involving force majeure, are to be consulted with the Program Management team for assessment of the situation and corresponding documentation, so that Program Management can support and guide each student optimally.

CODE OF CONDUCT IN CLASS

Be on time. Students arriving more than 5 minutes late will be marked as "Absent". Only students that notify in advance in writing that they will be late for a specific session may be granted an exception (at the discretion of the professor). Students attending online must always have their cameras on during the session or risk being marked absent.
If applicable, bring your name card and strictly follow the seating chart. It helps faculty members and fellow students learn your names.
Do not leave the room during the lecture: Students are not allowed to leave the room during lectures. If a student leaves the room during lectures, he/she will not be allowed to re-enter and, therefore, will be marked as "Absent". Only students that notify that they have a special reason to leave the session early will be granted an exception (at the discretion of the professor).
Do not engage in side conversation. As a sign of respect toward the person presenting the lecture (the teacher as well as fellow students), side conversations are not allowed. If you have a question, raise your hand and ask it. If you do not want to ask it during the lecture, feel free to approach your teacher after class. If a student is disrupting the flow of the lecture, he/she will be asked to leave the classroom and, consequently, will be marked as "Absent".
Use your laptop for course-related purposes only. The use of laptops during lectures must be authorized by the professor. The use of Social Media or accessing any type of content not related to the lecture is penalized. The student will be asked to leave the room and, consequently, will be marked as "Absent".
No cellular phones: IE University implements a "Phone-free Classroom" policy and, therefore, the use of phones, tablets, etc. is forbidden inside the classroom. Failing to abide by this rule entails expulsion from the room and will be counted as one absence.
Escalation policy: 1/3/5. Items 4, 5, and 6 above entail expulsion from the classroom and the consequent marking of the student as "Absent." IE University implements an "escalation policy": The first time a student is asked to leave the room for disciplinary reasons (as per items 4, 5, and 6 above), the student will incur one absence, the second time it will count as three absences, and from the third time onward, any expulsion from the classroom due to disciplinary issues will entail 5 absences.

RETAKE POLICY

Any student whose weighted final grade is below 5 will be required to sit for the retake exam to pass the course (except those not complying with the attendance rules, who are banned from this possibility). Grading for retakes will be subject to the following rules:

The retakes will consist of a comprehensive exam. The grade will depend only on the performance on this exam; continuous evaluation over the semester will not be taken into account.
The exam will be designed bearing in mind that the passing grade is 5 and the maximum grade that can be attained is 8 out of 10.
Dates and location of the retakes will be posted in advance and will not be changed.

PLAGIARISM / ACADEMIC HONESTY

Plagiarism is the dishonest act of presenting another person’s ideas, texts or words as your own.

This includes in order of seriousness of the offense:

providing faulty sources;
copy-pasting material from your own past assignments (self-plagiarism) without the instructor’s permission;
copy-pasting material from external sources even while citing them;
using verbatim translations from sources in other languages without citing them;
copy-pasting material from external sources without citing them;
and buying or commissioning essays from other parties.

IEU students must contact the professor if they don’t know whether the use of a document constitutes plagiarism. The professor will advise the student on how to present said material. All written assignments have to be submitted through Turnitin, which produces a similarity report and detects cases of plagiarism. Professors are required to check each student's academic work in order to guarantee its originality. If the originality of the academic work is not clear, the professor will contact the student in order to clarify any doubts. In the event that the meeting with the student fails to clarify the originality of the academic work, the professor will inform the Director of the Bachelor Program about the case, who will then decide whether to bring the case forward to the Academic Ethics Committee. Very high similarity scores will be automatically flagged and forwarded to the Academic Ethics Committee. Plagiarism constitutes a very serious offense and may carry penalties ranging from getting a zero for the assignment to expulsion from the university depending on the severity of the case and the number of times the student has committed plagiarism in the past.