

Degree programme	Type	Course
Interactive Communication	OB	4

Contact lecturer

Name : Mireia Camacho Corrales

Email : mireia.camacho@uab.cat

Prerequisites

Group languages

You can consult this information at the [end](#) of the document.

It is recommended to have completed the course "Introduction to Big Data" in order to better understand and carry out the course.

Objectives

The aim of the course is to study and practice the handling of large databases, complemented by real use cases. Various techniques and functions will be explored to enable students to clean and analyze databases regardless of their size. The course also embraces the open data paradigm as a working system with multiple possibilities for practical application.

Learning outcomes

1. Distinguish the salient features in all types of documents within the subject.
2. Cross-check information to establish its veracity, using evaluation criteria.
3. Submit course assignments on time, showing the individual and/or group planning involved.
4. Identify situations in which a change or improvement is needed.
5. Propose new methods or well-founded alternative solutions.
6. Identify the social, economic and environmental implications of academic and professional activities within one's own area of knowledge.
7. Propose viable projects and actions to boost social, economic and environmental benefits.
8. Propose projects and actions that are in accordance with the principles of ethical responsibility and respect for fundamental rights and obligations, diversity and democratic values.
9. Propose projects and actions that incorporate the gender perspective.
10. Communicate using language that is not sexist or discriminatory.
11. Critically analyse the principles, values and procedures that govern the exercise of the profession.
12. Explain the explicit or implicit deontological code in your area of knowledge.
13. Evaluate the impact of problems, prejudices and discrimination that could be included in actions and projects in the short or medium term in relation to certain people or groups.
14. Analyse a situation and identify its points for improvement.
15. Weigh up the risks and opportunities of both one's own and other people's proposals for improvement.
16. Plan and execute academic projects in the field of big data.
17. Solve basic problems in big data.
18. Share experiences with the group as a path to learning, in order to work subsequently in

multidisciplinary groups.

19. Describe the infrastructure needed to store big data.
20. Explain the characteristics of the infrastructure needed to recover big data.
21. Explain the infrastructure needed to process big data.
22. Differentiate between the various types of existing architectures for working with big data.
23. Extract large volumes of data from social networks and the new digital media in particular.

Contents

1. Introduction to R and RStudio
2. Fundamentals of descriptive statistics
3. Types of databases
4. Open data and web scraping
5. Data cleaning and transformation
6. Text processing with regular expressions
7. Data visualization
8. Web application development with RShiny

The detailed programme will be announced on the first day of class with the presentation of the course timetable.

Learning activities and methodology

Title	Hours	ECTS	Learning outcomes
THEORETICAL SESSIONS	15	0.6	1, 10, 11, 13, 19, 20, 21
TUTORING	8.5	0.34	3, 4, 5, 6, 7, 8, 9, 10, 13, 14, 15, 16, 17, 18
SEMINARS	15	0.6	1, 2, 4, 13, 14, 19, 21, 22, 23
OTHER ACTIVITIES (study time; practice preparation; seminar preparation, readings, etc.)	50	2	1, 2, 3, 4, 6, 8, 9, 10, 11, 12, 13, 14, 16, 17, 19, 20, 21, 22, 23
ASSESSMENT	8.5	0.34	1, 2, 3, 8, 9, 10, 13, 16, 17, 18, 19, 23
Lab practices	18	0.72	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23

The methodology is based on the following activities:

Theoretical sessions: Introduction theoretical sessions to concepts

Laboratory practices: individual or team works in which practical activities are carried out with one task with time limit. Students must apply knowledge, distribute time and prepare the submission within the classroom and in the hours spent in practice under the professor's guidance.

Seminars: individual or teamwork in which more extensive practical activities are carried out and with tasks open to student creativity. There are no limited time in the classroom, but deadlines for submission. Students must apply knowledge, distribute time and prepare submissions by starting their work within the classroom, but continue it in the form of activities supervised by the professor's team.

Course final work: practical group assessment exercise in which students must solve, during course development, a practical application problem linked to the subject's objectives. Students must raise the problem and perform the four processes to provide a solution based on large amounts of data: search,

extraction, analysis and publication of data report including a proposal for a decision based on the information collected and analysed.

Annotation: within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

Assessment

Continuous assessment activities

Title	Weight	Hours	ECTS	Learning outcomes
PRACTICAL SESSIONS	50%	15	0.6	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23
STUDENT'S PARTICIPATION	10%	8	0.32	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23
SUBMISSION AND PRESENTATION OF THE COURSE PROJECT	40%	12	0.48	3, 4, 5, 7, 8, 9, 10, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23

Activity A. Course project and oral presentation (group) . 40% of the final grade.

Activity B: Laboratory Practice. 50% of the final grade.

Activity C. Student participation. 10% of the final grade.

To pass the module, a minimum passing grade (5.0) must be achieved in activities A and B.

SINGLE ASSESSMENT:

This course does not provide for a single-assessment system.

RE-EVALUATION:

During the last two weeks of term, students who have not passed the module may take part in a re-assessment process consisting of a theoretical exam and a practical exercise. The student must have completed at least two-thirds of the total practical work for the module (corresponding to Assessment B) and have achieved a grade of 3.5 or above (and under 5) in the weighted average of the assessment activities.

PLAGIARISM:

Should a student commit any irregularity that could lead to a significant variation in the mark for an assessment, that assessment will be marked as 0, regardless of any disciplinary proceedings that may be initiated. In the event that several irregularities occur in the module's assessments, the final grade for this module will be 0.

The use of AI

Students are permitted to use AI, even though the material uploaded to the course CV already contains the knowledge necessary to carry out the requested tasks without having to consult external sources. However, if phrases such as 'Here is the text you requested' or other expressions are detected, indicating that the text has been copied and pasted directly from AI, and therefore that the work has not been reviewed before submission, the exercise or assignment will be marked 0.

Bibliography

Basic bibliography

Fernández-Avilés, Gema; Montero, José-María; et al. (2024) Fundamentos de ciencia de datos con R. Editorial McGraw-Hill. Disponible a: <https://cdr-book.github.io/index.html>

Casas Roma, Jordi (2019) Big data: análisis de datos en entornos masivos. Barcelona: Editorial UOC.

Duran, Xavier (2019). El imperio de los datos: el big data, la privacidad y la sociedad del futuro. PUV Publicacions, Universitat de València: Càtedra de Divulgació de la Ciència, UCC+i, Unitat de Cultura Científica i de la Innovació, Universitat de València.

Dur Lahoz-Beltrá, Rafael (2019). En las entrañas del big data: una aproximación a la estadística. Emse Edapp, S.L.

Fuller A. (2012). The White Book of Big Data. The definitive guide to the revolution in business analytics. Fujitsu. <https://www.fujitsu.com/rs/Images/WhiteBookofBigData.pdf>

Further reading

Tascón, Mario. "Introducción: Big data. Pasado, presente y futuro" *Telos: Cuadernos de comunicación e innovación* 95 (2013): 47-50. <https://telos.fundaciontelefonica.com/archivo/numero095/#contenido>

Mohamed, Azlinah, et al. "The state of the art and taxonomy of big data analytics: view from new big data framework" *Artificial Intelligence Review* 53.2 (2020): 989-1037.

Gandomi, Amir, and Murtaza Haider. "Beyond the hype: Big data concepts, methods, and analytics" *International journal of information management* 35.2 (2015): 137-144.

Aldana Montes, José Francisco (2018). *Introducción al trabajo con datos*. Madrid: García-Maroto Editores

Lucivero, Federica. "Big data, big waste? A reflection on the environmental sustainability of big data initiatives." *Science and engineering ethics* 26.2 (2020): 1009-1030.

Software

The R programming language will be used throughout the course.

As the subject will carry out practical sessions during all its activities, it is recommended that students (if possible) always bring their laptop to the sessions.

Course groups and languages

The information provided is provisional until November 30. After this date, you will be able to consult the language of each group through this [link](#). To access the information, you will need to enter the course CODE

Type of teaching	Group	Language	Semester	Shift
(TE) Theory	6	Spanish	first semester	afternoon
(PLAB) Practical laboratories	61	Catalan	first semester	afternoon
(PLAB) Practical laboratories	62	Spanish	first semester	afternoon