DATA INSIGHTS AND VISUALIZATION

Grado en Comportamiento y Ciencias Sociales BBSS
SEP-2023 DIV-N-BBSS.1.M.A
Area Others
Number of sessions: 20
Academic year: 23-24
Degree course: FIRST
Number of credits: 3.0
Semester: 2º
Category: COMPULSORY
Language: English

Professor: **RAFAEL BALLESTER RIPOLL**

E-mail: rballester@faculty.ie.edu

RAFAEL BALLESTER RIPOLL

Rafael Ballester-Ripoll (rafael.ballester@ie.edu) joined IE University in 2019 as an Assistant Professor in Data Science. Prior to that, he was a Postdoctoral Associate at the University of Zurich (UZH) and at the Swiss Federal Institute of Technology in Zurich (ETH). He holds a PhD in Computer Science from UZH (2017) as well as BSc and MSc degrees in Mathematics and Computer Science from the Technical University of Catalonia-UPC (2012).

Office Hours

Office hours will be on request. Please contact at:

rafael.ballester@ie.edu

SUBJECT DESCRIPTION

Have you ever wondered about the total amount of Terabytes of data that are generated per hour? Did you know that Americans use an average of 3M GB of internet data per minute and that 50k of photos are posted in Instagram per minute? Nowadays, companies are starting to adopt data-driven decisions to define future strategies. Most of these businesses are creating Advanced Analytics departments with data units to manage the huge amount of data generated every minute of every day.

The first question that comes to people's mind is what to do with this huge amount of data? The human brain cannot assimilate all this information and make sense of it, especially when it is in raw format. Data visualizations make big and small data easier for the human brain to understand, to detect patterns and trends, and to spot unusual observations (i.e. outliers) in datasets. Good data visualizations help extract insights from complicated datasets so that their message is clear and concise.

In this context, this course will cover the basics concepts and tools that will allow us to organize data into charts, derive insights and extract conclusions that will help in data-driven decisions. We will learn how to analyze data and create various kinds of visualizations in R, one of the most demanded programming languages as well as Tableau, the most widely used business analytical software (Drive your business with Tableau).

LEARNING OBJECTIVES

The overall objective of this course is to learn the basics of data insights and visualization: from obtaining data, to cleaning, to creating goal-driven charts and visualization in R programming language as well as Tableau.

Secondary goals include:

- Introduction to problem-solving using data analysis;
- Creating and analyzing various charts and graphs;
- Understanding programming languages and their importance in solving real situations based on data;
- Exploring data wrangling and visualization in R;
- Build advanced visualizations using Tableau.

TEACHING METHODOLOGY

IE University teaching method is defined by its collaborative, active, and applied nature. Students actively participate in the whole process to build their knowledge and sharpen their skills. Professor's main role is to lead and guide students to achieve the learning objectives of the course. This is done by engaging in a diverse range of teaching techniques and different types of learning activities such as the following:

Learning Activity	Weighting	Estimated time a student should dedicate to prepare for and participate in
Lectures	26.67 %	20.0 hours
Discussions	6.67 %	5.0 hours
Exercises in class, Asynchronous sessions, Field Work	33.33 %	25.0 hours
Group work	26.67 %	20.0 hours
Individual studying	6.67 %	5.0 hours
TOTAL	100.0 %	75.0 hours

PROGRAM

SESSION 1 (LIVE IN-PERSON)

INTRODUCTION/WELCOME

Course logistics, evaluation and organization; computer programming environment (IDEs, R, RStudio, etc…). Desirable properties of a visualization diagram. Visual variables and attributes: what they are and why they matter. Choosing the right visual channel for the right task. Scatterplots in R

SESSION 2 (LIVE IN-PERSON)

BEYOND SCATTERPLOTS

More on data wrangling; sequential vs. diverging vs. qualitative colormaps; barplots; line plots; filled line plots; area plots.

SESSION 3 (LIVE IN-PERSON)

GGPLOT2 LAB (1)

Together, we will solve a set of problems on a real-world data set. These will be guided: you will be asked to create specific plots with specific properties.

SESSION 4 (LIVE IN-PERSON)

DENSITY DATA

Smooth density lines; dot plots and advanced customizations; box plots; violin plots; grid density plots; long vs. wide data; some miscellaneous tricks (overlaying multiple plots, flipping coordinates, etc.).

SESSION 5 (LIVE IN-PERSON)

GGPLOT2 LAB (2)

You will have to solve a set of guided practical exercises, slightly more challenging than session 3, covering the visualization techniques you have seen until now in R.

SESSION 6 (LIVE IN-PERSON)

DIMENSIONALITY REDUCTION

The "curse of dimensionality": why dimensionality reduction is often needed. Plot ensembles; feature selection; principal component analysis (PCA); t-stochastic neighbor embedding (t-SNE).

SESSION 7 (LIVE IN-PERSON)

GGPLOT2 GRADED EXERCISES

10% of your grade

You will put your ggplot2 knowledge in action and visualize interesting information from a few example datasets.

SESSION 8 (LIVE IN-PERSON)

GGPLOT2 MINIPROJECT

10% of your grade

You will be given a real-world data set and, using ggplot2, you will have to prepare a PowerPoint presentation that delivers interesting insights. This will be a more creative project than the previous ones: you will have to decide by yourself what diagrams fit best the message you want to communicate.

SESSION 9 (LIVE IN-PERSON)

MIDTERM EXAM

15% of your grade

SESSION 10 (LIVE IN-PERSON)

INTRODUCTION TO TABLEAU

Tableau uses an entirely different approach to ggplot2: it is based on point-and-click and drag-and-drop interaction with the mouse. We will learn the fundamentals of Tableau: reading data files; bar plots; line plots; area plots; scatter plots; miscellaneous trcks (overlaying plots, swapping axes, etc.).

SESSION 11 (LIVE IN-PERSON)

TABLEAU LAB (1)

Together, we will solve a set of problems on a real-world data set. These will be guided: you will be asked to create specific plots with specific properties.

SESSION 12 (LIVE IN-PERSON)

DATES, PATHS, AND MORE

We will delve deeper into Tableau's more powerful features: handling date formats; filtering variables; stacked bars; pie charts; density plots; paths; multiple overlays.

SESSION 13 (LIVE IN-PERSON)

TABLEAU LAB (2)

You will have to solve a set of guided practical exercises, slightly more challenging than session 11, covering the visualization techniques you have seen until now in Tableau.

SESSION 14 (LIVE IN-PERSON)

CALCULATED FIELDS, ANIMATIONS, AND MORE

Calculated fields in Tableau; pages and animations; geovisualization; merging data sources; packed bubbles; treemaps; boxplots; donut charts.

SESSION 15 (LIVE IN-PERSON)

VISUALIZATION COMPETITION (1st ROUND)

By groups; 12.5% of your grade

You will form groups of ~4 people each. Each group will pick and study a real-world data set, explore it, and find hidden interesting patterns to visualize. You will put together and present your conclusions in the form of an inspiring and eye-catching slideshow, which you will have to present during this session. The professor will give feedback to all teams, and the best teams will be classified for the 2nd round.

30th October 2023

SESSION 16 (LIVE IN-PERSON)

STORYTELLING WORKSHOP (1)

By groups

Classified teams will work on the 2nd round presentation. The rest of the teams work on a presentation on another dataset. The professor will guide and mentor all teams.

SESSION 17 (LIVE IN-PERSON)

STORYTELLING WORKSHOP (2)

By groups; 12.5% of your grade

Each teams submits their presentation; the professor will give feedback to all.

SESSION 18 (LIVE IN-PERSON)

VISUALIZATION COMPETITION (2nd ROUND)

By groups

Classified teams have to compete in the IE Auditorium against teams from other cohorts. Everyone else is still required to attend (missing it counts as an absence).

SESSION 19 (LIVE IN-PERSON)

RECAP AND Q&A

This session will be 100% dedicated to troubleshooting, addressing common issues and doubts, and questions about ggplot2 and Tableau in general.

SESSION 20 (LIVE IN-PERSON)

FINAL EXAM

25% of your grade

EVALUATION CRITERIA

A variety of teaching and learning strategies will be used in this course. You will be assigned a grade based on your demonstrated knowledge on in-class quizzes, a group visualization competition, a midterm and final exam, and your participation in various class activities and discussions. The scores for these different activities are as follows:

Final Exam (25%)

The final exam will include all the material explained in the course. The exam format will include multiple questions to be solved numerically and from a graphical point of view using R and Tableau. The final exam will be held during the last session.

Midterm Exam (15%)

The midterm exam will include material from the PowerPoint slides. This exam will be solved in groups and it will include several questions to be solved using R for a dataset that will be provided that day.

Group Visualization Competition (25%)

The ability to effectively work together with others is critical for success in all areas of life. Thus, the purpose of the group competition is to provide you with an opportunity to develop and demonstrate your ability to effectively work with others on all aspects of a project; including generating an idea, select a proper dataset, delegating tasks, effective time management, critically analyzing an issue, and effectively presenting the information via a top-notch presentation.

For the group work one dataset from some of these resources will be selected per group:

- https://www.springboard.com/blog/free-public-data-sets-data-science-project/
- https://www.kaggle.com/
- https://datos.madrid.es/portal/site/egob/

Exercises and Miniproject (20%)

We will dedicate two sessions to mandatory exercises and a miniproject in ggplot2 where you will hone your storytelling skills. These will be graded and will have a weight of 10% each in the global score.

Class Participation (15%)

Active participation in class activities, discussions, and labs is an especially important aspect in this course because our focus will be on understanding how the concepts discussed in class can be applied in real-world contexts.

Marks:

A minimum passing grade in the final exam (3.5) is required to pass the subject. If a student scores lower than this minimum, he will have to go to June retake, irrespective of their overall course grade. The overall passing course grade is 5.0.

criteria	percentage	Learning Objectives	Comments
Final exam	25 %		
Midterm exam	15 %		
Group visualization Competition	25 %		
Exercises and miniproject	20 %		
Class Participation	15 %		

RE-SIT / RE-TAKE POLICY

BIBLIOGRAPHY

Recommended

- Tamara Munzner. (2014). *Visualization Analysis & Design.* CRC Press. ISBN 9781466508934 (Digital)

BEHAVIOR RULES

Please, check the University's Code of Conduct here. The Program Director may provide further indications.

ATTENDANCE POLICY

Please, check the University's Attendance Policy here. The Program Director may provide further indications.

ETHICAL POLICY

Please, check the University's Ethics Code here. The Program Director may provide further indications.