This is a course for beginners! Students joining this course should not have any previous coding experience as the course is designed to teach them how to code for the first time.
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).
OBJECTIVES AND SKILLS
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;
- Learning the basic of programming with R;
- Writing fully functional console or R programs;
- Exploring data wrangling and visualization in R;

Build advanced visualizations using Tableau.
METHODOLOGY
The course is highly practical but has some solid theory as well. During the sessions, we will mix theoretical lecturing with practical assignments. The course is designed for you to always have a laptop so you can directly code the exercises during class. Approximately, 50% session’s time is dedicated to class discussions and in-class exercises.
Periodically, individual coding assignments will be given to help you reflect on what you have learned in class.
In this course, you will develop a group project to familiarize yourself with the concept of group programming. Since most of the real-life software projects are implemented by a group of programmers, it is good for you to be familiar with organizing work and interacting with other coders.
Towards the end of this course, groups will compete in building a top-notch visualization. Results will be presented in front of a Jury and the winning visualization(s) will earn a prize and will be displayed on the HST webpage. More information will be provided during the semester.
<table>
<thead>
<tr>
<th>Teaching methodology</th>
<th>Weighting</th>
<th>Estimated time a student should dedicate to prepare for and participate in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lectures</td>
<td>26.67 %</td>
<td>20 hours</td>
</tr>
<tr>
<td>Discussions</td>
<td>6.67 %</td>
<td>5 hours</td>
</tr>
<tr>
<td>Exercises</td>
<td>33.33 %</td>
<td>25 hours</td>
</tr>
<tr>
<td>Group work</td>
<td>26.67 %</td>
<td>20 hours</td>
</tr>
<tr>
<td>Other individual studying</td>
<td>6.67 %</td>
<td>5 hours</td>
</tr>
<tr>
<td>TOTAL</td>
<td>100.0 %</td>
<td>75 hours</td>
</tr>
</tbody>
</table>
PROGRAM

SESSION 1

Course introduction and overview of the computer programming environment
We will review the course logistics and organization together with the computer programming environment (IDEs, R, RStudio, etc…)

SESSIONS 2 - 3

Data types & variables
How to store information in R using variables. Different data-types for variables in R.

SESSIONS 4 - 5

Flow control statements, logic, scope and constants
We will expand the functionality we can code with iterations, logic and flow control with if/else structures

SESSION 6

Functions, arguments and return types
We will start encapsulating our code in methods that enable us to reuse the code preventing duplication and errors in large projects

SESSIONS 7 - 8

Lists and Data Frames.
We will learn how to store multiple items in lists and data Frames in R

SESSIONS 9 - 10

Reading and Writing Files
We will learn how to use files to store information on the long-term memory

SESSIONS 11 - 13

EDA: Exploratory Data Analysis
We will learn the basics for a descriptive analysis in order to extract conclusions and to take decisions based on data

SESSION 14

Midterm
We will solve some challenges in groups using R

SESSION 15

Merging datasets
We will learn how to combine two datasets using identifier variables

SESSIONS 16 - 17

dplyr library and pipes
We will learn how to filter, select and mutate datasets using pipes

SESSION 18

Introduction to data visualization
We will learn the importance of data visualization and how to select the proper chart for each specific objective

SESSIONS 19 - 20

Data Visualization using R
We will learn how to develop data visualizations with some visualization packages such as Ggplot2 finding hidden patterns in our data

SESSIONS 21 - 22

Insights and data visualization using Tableau
We will learn how to solve problems with numerical analysis and graphs using Tableau

SESSIONS 23 - 24

Visualization competition
Each group will show the top-notch visualization developed in front of a Jury

SESSION 25

Final exam
BIBLIOGRAPHY

Mandatory textbooks:
There is no mandatory reading for the course because we will review step by step the mechanics and concepts of programming in class but it would be helpful to have a R language reference book to ease you in your programming assignments.

Recommended textbooks:
There are many introductory books on R programming but we recommend:

Title: R for Data Science
Author: Hadley Wickham & Garrett Grolemund
Publisher / Edition / Year: O'Reilly/ 2017
ISBN / ISSN:
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:

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Percentage</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class Participation</td>
<td>15 %</td>
<td></td>
</tr>
<tr>
<td>Midterm exam</td>
<td>20 %</td>
<td></td>
</tr>
<tr>
<td>Group visualization</td>
<td>25 %</td>
<td></td>
</tr>
<tr>
<td>Competition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Exam</td>
<td>40 %</td>
<td></td>
</tr>
</tbody>
</table>

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.

Midterm Exam (20%)

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.kaggle.com/
- https://datos.madrid.es/portal/site/egob/

Final Exam (40%)

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.

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.
- All the presentations/videos/exams will be submitted via Turnitin on Campus Online. No other option will be accepted.

PROFESSOR BIO
Professor: PABLO MONFORT VINUESA  
E-mail: pmonfort@faculty.ie.edu

Pablo is a Senior Data Scientist, who has worked for McKinsey for 3 years after obtaining his PhD in Mathematics. In 2017 he founded his own company focused on Advance Analytics. He is a specialist in mathematical modeling, clustering, machine learning and data mining. He has been supporting several major client engagements by developing predictive models and data analysis. Pablo worked at the National Institute of Statistics (INE) as well.

Examples of his work include:
- Implementation of 4G telephony for a company leading Asian telecommunications.
- Clustering client base card of a major European bank.
- Prediction model for churned clients in Latin America telephone company.
- Sales analysis of a prominent Spanish textile company by segmenting its costume base and launch a product recommendation engine.

OTHER INFORMATION
WHEN QUESTIONS ARISE OUT OF CLASS:
email: if you have a question(s) that was not answered in class, you are welcome to ask your question(s) via email. I can be reached at pmonfort@faculty.ie.edu. Although I will make every effort to respond to your question(s) as quickly and thoroughly as possible, please recognize that I may not be available when you send an email. Thus, please allow me up to 48 hours to respond before sending a follow-up email.
Office Hours: if your question cannot be properly answered via email and/or you would prefer to meet in person, please send an email and we will meet during my scheduled office hours.

OTHER INFORMATION:
As per University Policy:
Each student has 4 chances to pass any given course distributed in two consecutive academic years (regular period and July period).

It is mandatory to attend 100% of the classes. Students who do not comply with at least 70% attendance will lose their 1st and 2nd chance, and go directly to the 3rd one (they will need to enroll again in this course the next academic year).

Grading for retakes will be subject to the following rules:
1. Those students who failed the subject in the first regular period will have to do a retake in July (except those not complying with attendance rules who are banned from this possibility).
2. Dates and location of the July retakes will be posted in advance and will not be changed. Please take this into consideration when planning your summer.
3. The maximum grade that a student may obtain in the 2nd exam session is 8 out of 10. Those students in the 3rd call will be required to attend 50% of the classes. If due to schedule overlap, a different option will be discussed with the professor in order to pass the subject.

Attendance:
Attendance at all scheduled classes is mandatory and essential for success in the course. If you miss class for any reason, you are responsible for getting notes from classmates. If you have questions about any assignment please send me an email. Under most circumstances, students who miss a class in which a presentation, mid-term, or final exam is held will not be granted an exception or given an opportunity to do a make-up assignment or exam. However, if illness or other circumstances prevent you from adhering to the assignment/presentation due dates stated in this syllabus, an exception may be granted at the discretion of the professor. In all cases, the student must provide official documentation (e.g., from a medical doctor, counsellor) to the professor within 24 hours of the missed due date.

Special need Students:
To request academic accommodations due to a disability, please contact Rafif Srour via email at: Rafif.Srour@ie.edu.

Student Privacy Statement:
At times, students may disclose personal information through class discussions. It is expected that all members of the class will respect the privacy of their classmates. This means that the information disclosed in the class will not be repeated or discussed with other students outside of the course.

Decisions about Grades:
Decisions about grades are made very carefully, and are final at the end of the course. If you have questions regarding a certain grade or you would like to receive personal feedback, you must request a meeting with me to discuss grades on specific assignments before the last class of the course. Any disputes regarding grades must be resolved before the final exam. “Extra credit” or makeup assignments will only be allowed under extenuating circumstances at the sole discretion of the course professor.

ACADEMIC INTEGRITY
Unless you are specifically instructed to work with other students in a group, all of your assignments, papers, projects, presentations, and any work I assign must reflect your own work and thinking.

What is academic integrity? When you do the right thing even though no one is watching. The core values of integrity, both academic and otherwise include: honesty, fairness, respect, responsibility, and trust. Academic Integrity requires that all students within Instituto de Empresa (IE) act in accordance with these values in the conduct of their academic work, and that they follow the rules and regulations concerning the accepted conduct, practices and procedures of academic research and writing. Academic Integrity violations are defined as Cheating, Plagiarism or other violations of academic ethics.

Cheating and plagiarism are very serious offenses governed by the IE student code of conduct. Any student found cheating or plagiarizing on any assignment or component of this course will at a minimum receive a “0” on the affected assignment. Moreover, the student will also be referred to the University Judicial System for further action. Additional penalties could include a note on your transcript, failing the class, or expulsion from the university.

It is important to note that, while the list below is comprehensive, it should not be considered exhaustive.

Cheating includes:

a. An act or attempt to give, receive, share, or utilize unauthorized information or unauthorized assistance at any time for assignments, papers, projects, presentations, tests or examinations. Students are permitted to mentor and/or assist other students with assignments by providing insight and/or advice. However, students must not allow other students to copy their work, nor will students be permitted to copy the work of other students. Students must acknowledge when they have received assistance from others.

b. Failure to follow rules on assignments, papers, projects, presentations, tests or examinations as provided by the course professor and/or as stipulated by IE.
c. Unauthorized co-operation or collaboration.
d. Tampering with official documents, including electronic records.
e. The impersonation of a student on presentations, exercises, tests or an examination. This includes logging onto any electronic course management tool or program (e.g. Black Board, etc.) using someone else’s login and password.

**Plagiarism includes:**

a. Using the work of others and attempting to present it as your own. For example, using phrases or passages from books, articles, newspapers, or the internet and not referencing them properly in your document. This includes using information from others without citing it, misrepresentation of cited work, and misuse of quotation marks.
b. Submitting an assignment or paper that is highly similar to what someone else has written (i.e., minimal changes in wording, or where the sentences are similar, but in a different order).
c. You don’t have to commit “word for word” copying to plagiarize – you can also plagiarize if you turn in something that is “thought for thought” the same as someone else.

**Other violations of academic ethics include:**

a. Not acknowledging that your work or any part thereof has been submitted for credit elsewhere.
b. Misleading or false statements regarding work completed.
c. Knowingly aiding or abetting anyone in committing any form of an Academic Integrity violation.
CODE OF CONDUCT IN CLASS

1. **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).

2. **If applicable, bring your name card and strictly follow the seating chart**. It helps faculty members and fellow students learn your names.

3. **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).

4. **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".

5. **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".

6. **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.

7. **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.