COURSE DATA

Data Subject	
Code	36425
Name	Data visualisation
Cycle	Grade
ECTS Credits	6.0
Academic year	2020 - 2021

Study (s)

Degree	Center	Acad. year	Period
1406 - Degree in Data Science	School of Engineering	2	Second term
1407 - Degree in Multimedia Engineering	School of Engineering	4	Second term

Subject-matter

Degree	Subject-matter	Character
1406 - Degree in Data Science	8 - Information Management	Obligatory
1407 - Degree in Multimedia Engineering	19 - Optatividad	Optional

Coordination

Name	Department
GOMEZ SANCHIS, JUAN	242 - Electronic Engineering
MARTINEZ SOBER, MARCELINO	242 - Electronic Engineering
PORTALES RICART, CRISTINA	240 - Computer Science

SUMMARY

Data scientists often have to use graphics for exploratory purposes or for disseminating results. Students therefore need to know the theory of perception and color, which are the elements of a visualization system and the main tools for creating suitable visualizations. They also need to know the most suitable types of graphics for each type of data, especially those whose special nature requires specific tools and methods such as spatial data. Interactive graphics play an important role in the user's knowledge extraction process. Students will therefore learn how to handle this type of graphics. They will also learn how to 'fit the pieces' into the puzzle in order to draft a report with visual information or create a complete control panel.

This course will enable students to analyze the advantages and disadvantages of various technologies in order to select the most suitable ones for each situation.

The theoretical classes will be taught in Spanish. The language for the practical and laboratory classes will be specified in the course guidelines available on the website for this degree.

PREVIOUS KNOWLEDGE

Relationship to other subjects of the same degree

There are no specified enrollment restrictions with other subjects of the curriculum.

Other requirements

Students are recommended to have passed Programming Foundations (FP), Data Structures and Algorithms (EDA) and Data Processing (TD), which are taught in the first and second terms of the first year of the degree.

OUTCOMES

1406 - Degree in Data Science

- Students must have acquired knowledge and understanding in a specific field of study, on the basis of general secondary education and at a level that includes mainly knowledge drawn from advanced textbooks, but also some cutting-edge knowledge in their field of study.
- Students must be able to communicate information, ideas, problems and solutions to both expert and lay audiences.
- (CG02) Ability to solve problems with initiative and creativity and to communicate and transmit knowledge, abilities and skills, which should include the ethical and professional responsibility of the activity of a data scientist.
- (CG03) Capability to elaborate models, calculations, reports, to plan tasks and other works analogous to the specific field of data science.
- (CT03) Ability to defend your own work with rigor and arguments and to expose it in an adequate and accurate way with the use of the necessary means.
- (CT04) To be responsible for ones own professional development and specialisation, applying the acquired knowledge in the identification of career opportunities and sources of employment.
- (CE02) To methodologically know and apply the programming techniques and the algorithms necessary for the efficient processing of information and the computer resolution of problems that use large volumes of data.

- (CE06) Ability to represent and visualise data sets for the extraction of knowledge.
- (CE13) To know how to design, apply and evaluate data science algorithms for the resolution of complex problems.

1407 - Degree in Multimedia Engineering

- G2 - Have the learning skills needed to undertake further studies or to gain further training with a certain degree of autonomy. (RD1393/2007)
- MM7 - Be able to apply the principles of audiovisual graphic design and communication to multimedia products.

LEARNING OUTCOMES

● Know what characteristics lead to good visualization. (CB01, CB04, CG02, CT03).

● Use (representative and diagnostic) statistical graphs to characterize data. (CB04, CG03,CT04,CE13).

● Implement multidimensional data visualization methods. (CB04,CE02,CE06).

● Apply data visualization techniques with temporal and spatial dependence (CB04, CG03, CE06).

● Implement interactive visualization procedures. (CE06,CT03,CE6,CE13).

Students on this course will:

- learn how to choose the most suitable graph for a certain type of data and implement it in R or Python.
- learn how to choose the most suitable type of color palette for transmitting the most relevant information contained in the data.
- acquire basic concepts of what a map is in order to represent spatial variables on it. Cartographic projections.
- learn the characteristics of the correct representation of spatial data. Continuous vs. discrete variables. Interpolation.
- learn the main types of spatial data (vector and raster) and how to manage them using a Geographic Information System (GIS).
- learn how to perform spatial operations through a GIS and understand how to extract information from it.
- create evolutionary maps by taking into account the temporal variable.
- represent spatial data using Python and R. Connecting to a GIS.
- know how to implement interactive graphics with R and Python.

- be able to implement a dashboard with the main tools in the field of data science collected in the current state of the art.

DESCRIPTION OF CONTENTS

1. Introduction to data visualization (2h)

1.1. Visualization of data in the data processing problem.
1.2. Why visualize data?
1.3. Color and perception.
1.4. Representation and symbology.
1.5. Open data.

2. Types of graphics

2.1. Graphical representation of data.
2.2. Univariate, bivariate and multivariate graphs.
2.3. Graphical representation of time and space series.
2.4. Case study.

3. Implementing advance graphics (6h)

3.1. Python Matplotlib Library.
3.2. Seaborn Python Library.
3.2. ggplot2 library of R.

4. Spatial data visualization (14h)

4.1. Maps and cartographic projections
4.2. Geographic Information Systems. QGIS, gvSIG
4.3. Types of spatial data. Interpolation and multi-layer visualisation
4.4. Operations on spatial variables. Logic operations, filters
4.5. Generation of thematic and temporal maps
4.6. Spatial data in Python and R. Connection to QGIS
4.7. Case study

5. Interactive data visualization (14h)

5.1. Interactive graphics. Plotly.
5.2. Elaboration of control panels in Python. Dash
5.3. Elaboration of control panels in R. Shiny
5.4. Case study

6. Graph visualization (2h)

6.1. Types of graphs and variants
6.2. Ontologies and their representation by graphs
6.3. Drawing of graphs in Python and in R
6.4. Case study

7. Data visualization laboratory

In this block a series of practical assumptions will be presented in the form of laboratory practices conducted in a computer room.

Practice 0. Python Pandas (2h)
Practice 1. Implementation of graphics in Python. Matplotlib and Seaborn (3h).
Practice 2. Implementation of graphics in R. Ggplot2 (3h).
Practice 3. Visualization of spatial data through a GIS (3h).
Practice 4. Visualization of spatial data in Python and in R (3h).
Practice 5. Dashboards in Python. Dash (3h).
Practice 6. Dashboards in R. Shiny (3h).

WORKLOAD

ACTIVITY	Hours	% To be attended
Theory classes	34,00	100
Laboratory practices	20,00	100
Classroom practices	6,00	100
Development of group work	10,00	0
Development of individual work	10,00	0
Study and independent work	10,00	0
Readings supplementary material	5,00	0
Preparation of evaluation activities	20,00	0
Preparing lectures	15,00	0
Preparation of practical classes and problem	10,00	0
Resolution of online questionnaires	10,00	0
TOTAL	**150,00**	

TEACHING METHODOLOGY

Classes will combine theoretical and practical content

MD1 - Theoretical activities. Expository development of the subject. Students will be encouraged to help solve specific questions by completing individual evaluation questionnaires.

Students will be encouraged to participate in these face-to-face theoretical activities, which will develop the key and most complex aspects of the course (CB01, CB04, CT04, CE02, CE06).

MD2 - Practical activities. Students will learn by solving problems, completing exercises and analyzing case studies to acquire competences in the various aspects of the subject (CB04, CG02, CE02, CE06, CE13).

The main objective of these practical activities, which perfectly complement the theory classes, is that students apply the theoretical concepts and expand them with the knowledge and experience they will have acquired by completing their assignments.

MD4 - Work in the laboratory and/or computer classroom. Students will learn by conducting practical assignments individually or in small groups in the laboratory and/or computer room (CB04, CG02, CG03, CT03, CE02, CE06, CE13).

As well as conducting face-to-face activities, students will be expected to complete assignments outside the classroom on various issues and problems and to study for their classes and exams. These tasks will mainly be done individually in order to promote autonomous work. Some assignments, especially the preparation and completion of laboratory practices, will require work to be completed in small groups of 2 or 3 students in order to improve their ability to work in a team. Students may also be required to participate in a small-group activity as a ‘challenge’ sponsored by a company in the Data Science sector.

The University of Valencia’s e-learning platform (*Aula Virtual*) will be used to communicate with students. Students will also be able to access the teaching materials used in class and the problems and exercises they need to solve via this platform.

EVALUATION

Evaluation for this course will comprise the following components:

SE1 - Objective test, consisting of deliverables and tests consisting of both theoretical-practical questions and problems (assessment of competences CB01, CT03, CT04, CE02, CE06, CE13) (50%) (Note: All percentages refer to the final mark).

SE1-1 (40%) Theory exam

SE1-2 (10%) Laboratory test and deliverables

SE2 - Evaluation of practical activities based on the preparation of papers/memories and/or oral presentations and prelaboratori tests (competence evaluation CB04, CG02, CG03, CT03, CT04, CE02, CE06, CE13) (30%)

SE2-1 (25%) Realisation of one or two mini-projects about data visualisation or the elaboration of a scorecard with real data.

SE2-2 (5%) Prelaboratory tests

SE3 - Continuous assessment of each student. (20%)

SE3-1 (1%) Regular attendance at planned telematic or face to face activities (competency assessment CB04, CG01). (Activity NOT RECOVERABLE)

SE3-2 (19%) Resolution of proposed issues and problems (competency assessment CB01, CB04, CG03, CE06). (NON-RECUPERABLE Activity)

The final grade of the course will be calculated as the weighted average of each of the previous sections, according to the following criteria: SE-1 (50%), SE-2 (30%), SE-3 (20%).

Particular considerations on the evaluation:

- A minimum score of 4 (out of 10) is required for the evaluation sections SE1-1, SE1-2 and SE2-1.
- Students who do not pass the SE1 or SE2 assessment activities in the first trial must take a theoretical-practical exam to evaluate the SE1 activity and an individual project to evaluate the SE2 activity in second trial.
- Activities SE3-1 and SE3-2 are not recoverable.

In all cases the evaluation system will be governed by the University of Valencia's regulations on grading and assessment for bachelor's degrees and master's degrees, which is available at:

https://webges.uv.es/uvTaeWeb/MuestraInformacionEdictoPublicoFrontAction.do?accion=initio&idEdictoSeleccionado=5639

REFERENCES

Basic

- C. O. Wilke (2019) Fundamentals of Data Visualization. OReilly.
- QGIS Project (2019), QGIS User Guide, https://docs.qgis.org/3.4/pdf/en/QGIS-3.4-UserGuide-en.pdf
- Mas, J-F., (2018). Análisis espacial con R: Usa R como un Sistema de Información Geográfica, European Scientific Institute, 114 p. http://eujournal.org/files/journals/1/books/JeanFrancoisMas.pdf
- QGIS Project, (2019), PyQGIS developer cookbook Release 3.4, https://docs.qgis.org/3.4/pdf/en/QGIS-3.4-PyQGISDeveloperCookbook-en.pdf
- C. Adams, (2014) Learning Python Data Visualization. OReilly.

Additional

- Stephen Wise (2014), GIS fundamentals. CRC Press, Taylor & Francis Group, 305 p.
- C. Beeley, S.R.Sukhdeve. 2018. Web Application Development with R Using Shiny: Build stunning graphics and interactive data visualizations to deliver cutting-edge analytics, 3rd Edition. Packt.
- Dash User guide and documentation. https://github.com/plotly/dash-docs/blob/master/pdf-docs/Dash_User_Guide_and_Documentation.pdf
- Menno-Jan Kraak, Ferjan Ormeling (2013), Cartography: Visualization of Geospatial Data, Routledge, Taylor & Francis Group, 202 p.
- Kang-tsung Chang, (2015). Introduction to Geographic Information Systems, McGraw-Hill Education, 448 p.
- Colette Cauvin, Francisco Escobar, Aziz Serradj, (2010), Thematic Cartography and Transformations, Wiley, 465 p.

ADDENDUM COVID-19

This addendum will only be activated if the health situation requires so and with the prior agreement of the Governing Council

The teaching methodology of the course will follow the Teaching Model approved by the Data Science Academic Committee (https://go.uv.es/cienciadatos/ModelDocentGCD2Q). In the event that the facilities are closed for health reasons that affect all or part of the course sessions, these will be replaced by non-presential sessions following the established timetable. If the closure affects a presential assessment test for the subject, it will be replaced by a test of a similar nature that will be carried out in virtual mode through the computer tools supported by the University of Valencia. The percentages of each assessment test will remain unchanged, as established by this guide.