



## Teaching Guide

Identifying Data					2022/23
Subject (*)	Big Data for the Social Sciences			Code	615545008d
Study programme	Máster Universitario en Métodos Avanzados de Investigación e Innovación na Análise Social (a distancia)				
Descriptors					
Cycle	Period	Year	Type	Credits	
Official Master's Degree	2nd four-month period	First	Obligatory	6	
Language	Spanish				
Teaching method	Face-to-face				
Prerequisites					
Department	Socioloxía e Ciencias da Comunicación				
Coordinador	León Medina, Francisco J.	E-mail	f.leon@udc.es		
Lecturers	León Medina, Francisco J.	E-mail	f.leon@udc.es		
Web					
General description	<p>As tecnoloxías dixitais revolucionaron a nosa forma de vivir e a forma en que recollemos e analizamos datos sobre o comportamento social. Por unha banda, a crecente capacidade computacional permitiu desenvolver e probar modelos de procesos sociais ata agora matematicamente imposibles. Por outra banda, o acceso aos rexistros da pegada dixital que deixa o noso comportamento online supuxo a posibilidade de traballar con bases de datos masivas, comportamentais, non reactivas, relacionais e dinámicas. Ambas posibilidades converxeron na chamada ciencia social computacional, que á súa vez enlaza cunha longa tradición empírica, analítica, computacional e xenerativista que foi a columna vertebral desa ciencia social máis orientada á explicación científica e rigorosa dos fenómenos sociais.</p> <p>Neste curso, o alumnado introducirase nos principais debates epistemolóxicos, teóricos e éticos da ciencia social computacional e, especialmente, os que se producen arredor do uso do Big Data para a investigación social. Ademais, ofrecerá as ferramentas e os coñecementos básicos para deseñar e executar unha investigación baseada en datos masivos: introducirase o alumno nas técnicas e procedementos destinados a obter ou producir, xestionar e analizar bases de datos masivas, así como as máis innovadoras para a presentación e visualización de resultados.</p>				

## Study programme competences

Code	Study programme competences
A1	C1. Know in detail the main theoretical and practical advances in qualitative and quantitative research and their application - critical, reflective and contextual - to understand, define and explain highly complex and uncertain phenomena associated with current social, economic and political challenges .
A6	C6. Deal with the theoretical and scientific foundations necessary to carry out quantitative and qualitative social research using advanced technological tools (R).
A7	C7. Recognize the different options for presenting and displaying research results and their methodological foundations to clearly convey the results of the research to different audiences.
A8	C8. Know the modes and potentialities of using mass data and data mining in the context of social sciences.
A11	H1. Evaluate and select the appropriate methodological paradigm (quantitative, qualitative, mixed, prospective, participatory) to formulate judgments based on incomplete or limited information on phenomena associated with social, economic and political challenges.
A14	H4 Transmit the results of basic and applied research, as well as the most relevant foundations on which they are based, clearly and unambiguously, taking into account the diversity of contexts, audiences and needs, using different presentation and display options.
A15	H5 Carry out qualitative and quantitative research in R.
A16	H6 Develop the ability to plan research, correctly choosing the phases and procedures that guarantee maximum rigor and that are suitable for different contexts within the public or private sphere.
A17	H7 Acquire the ability to continue delving into research and analysis methods in a way that will be largely self-directed and autonomous.
A19	H9 Develop an advanced capacity to evaluate the quality and results of research processes.
A20	CP1. Contrast the importance of the existing methodological diversity for the analysis of the social, political and economic reality, addressing different complex problems of current societies through creative and innovative solutions.
A22	CP3. Develop and implement research projects in new or little-known environments.



A24	CP5. Being able to work with data sources, scientific research methodologies and techniques, and advanced computer tools typical of the social sciences.
A27	CP8. Design and execute a research project based on the use of Big Data for scientific purposes, making original contributions to the field of study.
A28	CP9. Obtain, manage and analyze massive databases using the necessary computer tools and programming languages, especially the advanced language R.

## Learning outcomes

Learning outcomes	Study programme competences		
Os estudantes serán capaces de deseñar e executar un proxecto de investigación baseado no uso do Big Data con fins científicos, facendo achegas orixinais e sendo capaz de resolver problemas teóricos e prácticos utilizando estratexias innovadoras.	AC1		
	AC6		
	AC7		
	AC8		
	AC11		
	AC15		
	AC16		
	AC17		
	AC20		
	AC22		
	AC24		
	AC27		
	AC28		
O alumnado será capaz de obter, xestionar e analizar bases de datos masivas utilizando as ferramentas informáticas e linguaxes de programación necesarios, especialmente a linguaxe R avanzada, así como as técnicas estatísticas adecuadas para a súa análise.	AC8		
	AC24		
	AC28		
O alumnado será capaz de avaliar criticamente a calidade do coñecemento científico producido mediante a análise de datos masivos e reflexionar sobre a súa importancia e impacto socioeconómico.	AC19		
	AC20		
O alumnado poderá utilizar as ferramentas máis avanzadas para a visualización de datos e para a presentación dos resultados da investigación.	AC7		
	AC14		
	AC24		

## Contents

Topic	Sub-topic
TEMA 1: Fundamentos da ciencia social computacional	Discusións teóricas, metodolóxicas e éticas arredor do Big Data
TEMA 2: Obtención, xestión e almacenamento de datos masivos	Tipos e arquitecturas de big data. Obtención de datos masivos: web, APIs, etc.
TEMA 3: Tratamento e análise de datos masivos	Análise de sentimentos, análise de redes sociais, análise de redes sociais, aprendizaxe automática, etc.
TEMA 4: Ferramentas avanzadas de visualización masiva de datos	R (ggplot2), Gephi, etc.

## Planning

Methodologies / tests	Competencies	Ordinary class hours	Student?s personal work hours	Total hours
Student portfolio	A6 A7 A11 A14 A15 A16 A17 A19 A20 A22 A24 A27 A28	0	53	53



Workshop	A1 A6 A7 A8 A11 A14 A15 A16 A17 A19 A20 A22	21	21	42
Oral presentation	A7 A14	8	24	32
Guest lecture / keynote speech	A1 A6 A8 A11 A16 A17 A19 A20	21	0	21
Personalized attention		2	0	2

(\*)The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.

Methodologies	
Methodologies	Description
Student portfolio	Os diferentes exercicios que se desenvolverán nas sesións de ?Obradoiro?, así como os exercicios que se propoñan realizar de forma autónoma, incorporaranse a unha ?carpeta do alumno? que se entregará ao remate do curso e estará suxeita a avaliación..
Workshop	Ao longo do curso realizaranse diferentes sesións prácticas nas que se realizarán exercicios nos que o alumnado aprenderá os procedementos básicos para obter, estruturar, almacenar, analizar e visualizar datos masivos. As sesións prácticas realizaranse de forma presencial. Para aqueles que poidan asistir de forma virtualmente sincrónica, as sesións transmitiránse a través de Teams. Para os que non poidan asistir de forma sincronizada, e para o resto do alumnado, as sesións gravaranse en vídeo e a súa visualización estará dispoñible a través do Campus Virtual.
Oral presentation	Ao remate do curso, dedicarase unha sesión á exposición dalgúns dos exercicios que forman parte do portfolio do alumno, especialmente aqueles dirixidos á visualización e presentación de resultados. Para o alumnado en modalidade virtual, farase via TEAMS.
Guest lecture / keynote speech	Haberá clases maxistras nas que o peso da actividade recae principalmente no profesor. Este método pedagóxico complementase coa introdución doutras actividades e métodos que permitan un papel máis activo ao alumnado. Dado o contido eminentemente orientado á práctica investigadora desta materia, as sesións maxistras terán frecuentemente un contido técnico, orientado á aprendizaxe dos procedementos de investigación baseados no Big Data máis habituais. Esta orientación técnica complementarase cunha reflexión sobre os fundamentos epistemolóxicos, éticos e teóricos deste tipo de investigacións. As clases maxistras realizaranse de forma presencial. Para aqueles que poidan asistir de forma virtualmente sincrónica, as sesións transmitiránse a través de Teams. Para os que non poidan asistir de forma sincronizada, e para o resto do alumnado, as sesións gravaranse en vídeo e a súa visualización estará dispoñible a través do Campus Virtual.

Personalized attention	
Methodologies	Description
Guest lecture / keynote speech Student portfolio Workshop Oral presentation	A atención personalizada desta materia desenvolverase de forma virtual. Para a comprensión das lecturas e a correcta elaboración dos exercicios prácticos, os alumnos contarán coa presenza virtual do profesor durante a retransmisión das sesións de clase e, ademais, poderán programar as titorías virtuais que consideren necesarias no horario que se comunicará ao comezo de curso. O profesor tamén estará dispoñible a través do correo electrónico (e tamén a través do Campus Virtual).

Assessment			
Methodologies	Competencies	Description	Qualification
Student portfolio	A6 A7 A11 A14 A15 A16 A17 A19 A20 A22 A24 A27 A28	Ao longo do curso realizaranse diferentes sesións prácticas nas que se realizarán exercicios nos que o alumnado aprenderá os procedementos básicos para obter, estruturar, almacenar, analizar e visualizar datos masivos. Os diferentes exercicios que se desenvolverán nestas sesións, así como os que se propoñan para realizar de forma autónoma, incorporaranse a unha ?carpeta do alumno? que se entregará ao final do curso e será obxecto de avaliación.	90



Oral presentation	A7 A14	Ao remate do curso, dedicarase unha sesión á exposición dalgúns dos exercicios que forman parte do portfolio do alumno, especialmente aqueles dirixidos á visualización e presentación de resultados. Para o alumnado en modalidade virtual, farase via TEAMS.	10
-------------------	--------	--	----

### Assessment comments

### Sources of information

<b>Basic</b>	<ul style="list-style-type: none"> <li>- Flache, A., Mäs, M., &amp; Keijzer, M. A. (2022). Computational approaches in rigorous sociology: agent-based computational modeling and computational social science. In Handbook of Sociological Science (pp. 57-72).. Edward Elgar Publishing</li> <li>- Spaiser, V. (2021). Digital data and methods. In Research Handbook on Analytical Sociology (pp. 352-363).. Edward Elgar Publishing</li> <li>- Salganik, M. J. (2019). Bit by bit: Social research in the digital age. Princeton University Press</li> <li>- Marr, B. (2016). Big data in practice: how 45 successful companies used big data analytics to deliver extraordinary results. John Wiley &amp; Sons</li> <li>- Mayer-Schönberger, V., &amp; Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt</li> <li>- Lazer D, et al. (2009). Computational social science. Science, 323(5915):721?723..</li> <li>- Ruths, D., &amp; Pfeffer, J. (2014). Social media for large studies of behavior. Science, 346(6213), 1063-1064.</li> </ul>
<b>Complementary</b>	

### Recommendations

#### Subjects that it is recommended to have taken before

Descriptive and Inferential Statistical Analysis/615545001  
 Foundations of Social Research using R/615545003  
 Advanced Statistical Analysis/615545007

#### Subjects that are recommended to be taken simultaneously

#### Subjects that continue the syllabus

#### Other comments

(\*The teaching guide is the document in which the URV publishes the information about all its courses. It is a public document and cannot be modified. Only in exceptional cases can it be revised by the competent agent or duly revised so that it is in line with current legislation.