

## **-CURSO DE EDUCACIÓN PERMANENTE 2022-**

# **R aplicado al análisis cualitativo**

**Docente:** **Elina Gómez(UMAD-FCS)**

- Desde 22 de noviembre al 7 de diciembre.
- Martes, Miércoles y Jueves de 16:30 a 19 hrs, Facultad de Ciencias Sociales.
- Lugar: Sala de informática C3, tercer piso.
- Carga horaria: 20 horas.
- Matrícula: \$3.486.
- Dirigido a egresada/os universitarios, estudiantes de posgrado, estudiantes de grado avanzadas/os<sup>1</sup>.
- [Formulario de inscripción](#).

### **Objetivos:**

Lograr una familiarización el lenguaje R y los conceptos más fundamentales que esto implica (como las clases de objetos, funciones, operadores, etc.).

Lograr que el estudiante pueda obtener datos de texto de diferentes fuentes directamente desde R (web scraping, prensa, APIs de redes sociales)

Lograr que cada estudiante se familiarice con las funciones más comunes en el trabajo con R aplicado al análisis cualitativo, tanto en los que refiere a codificación de textos (RQDA), pre - procesamiento, como análisis textual.

Contribuir a que cada estudiante se familiarice con los conceptos de corpus, matriz de términos, así como maneje técnicas vinculadas a la minería de texto: asociación y agrupación de palabras, uso de diccionarios (manuales y automáticos), análisis de sentimiento, aprendizaje automático aplicado a datos de texto.

Mostrar herramientas para la visualización estática e interactiva de resultados.

### **Contenidos:**

1. NIVELACIÓN EN LENGUAJE "R": R como software libre y gratuito y su interfaz gráfica RStudio. Uso de la ayuda y foros; trabajo en proyecto (Rproj); Paquetes y funciones. Operadores relacionales y lógicos; Clases de objetos: Vectores ; Matrices; Data.Frames y Listas.
2. FUENTES DE DATOS: Contemplar fuentes de datos textuales: Textos brutos (OCR); Web scraping; APIs de redes sociales; obtención mediante conexión desde R/RStudio.

---

<sup>1</sup> De acuerdo a lo establecido en el [Marco para las actividades de Formación y Actualización Permanente de la FCS](#), se admitirán estudiantes de grado con el 75% de la currícula aprobada y/o líneas de investigación vinculadas al tema del curso.

3. PRE - PROCESAMIENTO DE LOS DATOS: Procesamiento de strings, expresiones regulares y abordaje de los requerimientos previos (limpieza y homogenización) para el análisis de textos. Exploración y manipulación básica de datos, contemplando las diferentes fuentes y codificaciones posibles. Codificación manual de textos y creación de redes multinivel (categorías, códigos y citas) mediante la plataforma RQDA().

4. MINERÍA DE TEXTOS: Noción de Corpus, matriz de términos y sus posibilidades analíticas, desde lo más descriptivo a la aplicación de técnicas más complejas: asociación y agrupación de palabras, uso de diccionarios (manuales y automáticos), análisis de sentimiento.

5. VISUALIZACIÓN: Exploración de las diferentes posibilidades gráficas de visualización de los resultados del análisis textual (nubes de palabras, frecuencias, dendogramas, etc.).

6. APRENDIZAJE AUTOMÁTICO: Introducción a técnicas de aprendizaje automático (machine learning) aplicado a datos textuales, modelos supervisados y no supervisados, modelos de clasificación.

**\*\*SE REQUIERE CONOCIMIENTO INTERMEDIO DE R PREVIO AL CURSO\*\***

### **Método didáctico:**

Breves exposiciones seguidas de trabajos prácticos, individuales y grupales.

### **Bibliografía:**

Abedin, Jaynal & Kishor Kumar Das 2015. Data Manipulation with R (2nd Edition).

Benoit, Kenneth & Michael Laver. 2003. "Estimating Irish party policy positions using computer wordscoring." *Irish Political Studies* 18(1):97–107.

Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S and Matsuo A 2018. "quanteda: An R package for the quantitative analysis of textual data." *Journal of Open Source Software*, 3(30), pp. 774. doi: 10.21105/joss.00774.

Hart, Roderick P. & Jay P. Childers. 2005. "The Evolution of Candidate Bush." *American Behavioral Scientist* 49(2):180–197.

Hu, M. & B. Liu. 2004. Mining and summarizing customer reviews. In proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, page 168-177, Huang, Ronggui. 2018. RQDA: R-based Qualitative Data Analysis. R package version 0.3-1.

Klemmensen, Robert, Sara BinzerHobolt & Martin Ejnar Hansen. 2007. "Estimating policy positions using political texts: An evaluation of the Wordscores approach." *Electoral Studies* 26(4):746–755.

Klingemann, Hans-Dieter, Andrea Volkens, Judith Bara, Ian Budge & Michael McDonald. 2006. *Mapping Policy Preferences II: Estimates for Parties, Electors, and Governments in Eastern Europe, European Union and OECD 1990-2003*. Oxford: Oxford University Press.

Klüver, Heike. 2009. "Measuring interest group influence using quantitative text analysis." *European Union Politics* 10(4):535–549.

Krippendorff, Klaus. 2004. *Content Analysis: An Introduction to Its Methodology*. 2 ed. Thousand Oaks: Sage.

Langer, Ana Ines & Iñaki Sagarzazu. 2016 Are all policy decisions equal? Explaining the variation in media coverage of the UK budget. *Policy Studies Journal*

Laver, Michael & John Garry. 2000. "Estimating Policy Positions from Political Texts." *American Journal of Political Science* 44(3):619–634.

Laver, Michael, Kenneth Benoit & John Garry. 2003. "Extracting policy positions from political texts using word as data." *American Political Science Review* 97(2):311–331.

Thomas M., B Pang & L Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *EMNLP*, page 327335

Verzani, John 2011. *Getting Started with RStudio: An Integrated Development Environment for R* (1st Edition).

Wickham, Hadley & Garrett Golemund 2017. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*