

Software libre en la formación sociológica: R para el análisis de datos sociales en la era digital

Free Software in Sociological Education: R for Social Data Analysis in the Digital Age

David Moreno-Alameda¹

Resumen

Cada vez con mayor frecuencia, observamos un aumento en la formación y demanda del lenguaje de programación R por parte del alumnado y personal investigador en Sociología y otras Ciencias Sociales. En este sentido, en este artículo abordaremos el estado actual de la enseñanza de software estadístico, destacando las ventajas, desafíos y recomendaciones para su integración en la docencia universitaria del área. Para ello, la revisión se complementa con los resultados del cuestionario aplicado como parte del proyecto de Innovación Docente «Aula de Innovación Tecnológica para el Uso de Aplicaciones Informáticas Accesibles en las Ciencias Sociales» dirigido al conjunto de la comunidad universitaria de la Facultad de Ciencias Políticas y Sociología de la Universidad Complutense de Madrid. Se discutirá por tanto la conveniencia de utilizar *software* libre como R, frente alternativas de *software* propietario, fundamentalmente por su ausencia de costo, el desarrollo de extensiones por la propia comunidad, y sus ventajas en el marco de la docencia.

Palabras clave

R, programación, estadística, enseñanza, competencias, *software* libre.

Abstract

With increasing frequency, we observe the training and demand for the R programming language among students and researchers in Sociology and other Social Sciences. In this context, this article addresses the current state of statistical software education, highlighting its advantages, challenges, and recommendations for its integration into university-level teaching in the field. To this end, the review is supplemented by the results of a questionnaire conducted as part of the «Aula de Innovación Tecnológica para el Uso de Aplicaciones Informáticas Accesibles en las Ciencias Sociales» Teaching Innovation Project, aimed at the entire university community of the Faculty of Political Science and Sociology at the Complutense University of Madrid. Thus, the discussion will focus on the suitability of using Free Software such as R, compared to proprietary software alternatives, primarily due to its cost-free nature, the development of extensions by the community itself, and its advantages within the teaching framework.

Keywords

R, programming, statistics, teaching, competences, free software.

Cómo citar/Citation

Moreno-Alameda, David (2025). *Software libre en la formación sociológica: R para el análisis de datos sociales en la era digital*. *Revista de Sociología de la Educación - RASE*, 18 (1), 234-249. <http://dx.doi.org/10.7203/RASE.18.1.29809>.

Recibido: 18-11-2024

Aceptado: 13-01-2025

¹ David Moreno-Alameda, Universidad Complutense de Madrid, davidmorenoalameda@gmail.com.

1. Introducción²

Los avances tecnológicos de las últimas décadas han reconfigurado las herramientas empleadas de la enseñanza y la investigación de las Ciencias Sociales, y, en particular, de la Sociología. Este cambio ha conducido a un mayor acceso a datos masivos (Robles *et al.*, 2022; Tinati *et al.*, 2014), la atención a comunidades virtuales a través de Internet (Fielding, Lee, y Blank, 2017), una reducción de costes de cuestionarios web frente a opciones más tradicionales (Díaz de Rada, Domínguez, y Pasadas, 2019), así como una demanda cada vez mayor de técnicas avanzadas y específicas para el análisis de datos. En consecuencia, nos encontramos ante una demanda creciente de las habilidades y conocimientos tecnológicos, tanto en la docencia universitaria, como en la investigación académica y de mercado.

Dentro de este contexto, en el marco de la Universidad Complutense de Madrid y como parte de las actividades realizadas en el proyecto de Innovación Docente «Aula de Innovación Tecnológica para el Uso de Aplicaciones Informáticas Accesibles en las Ciencias Sociales» 2023/2024, se realizó un curso formativo abierto y gratuito para la adquisición de competencias y el manejo de herramientas informáticas. Dicho curso estuvo abierto a todo el alumnado de Sociología y Ciencias Sociales afines (como Economía, Psicología Social, Ciencias Políticas, etc.), Personal Técnico de Administración y Servicios, así como Personal Docente e Investigador de la Facultad de Ciencias Políticas y Sociología. El énfasis en este último grupo fue considerado fundamental, ya que es este quien transmite el conocimiento a partir de las distintas asignaturas en titulaciones y asignaturas de Sociología, y, por consiguiente, incide en el alumnado matriculado en dichas materias. Profundizando en la perspectiva adoptada, se impartieron talleres orientados a una mayor comprensión computacional y operacional de los fenómenos sociales, generalmente de naturaleza cuantitativa y con operaciones estadísticas. Para ello, se utilizaron como apoyo datos de encuestas sociales comunes dentro del área de la Sociología y áreas afines, como las realizadas por el Centro de Investigaciones Sociológicas o el Instituto Nacional de Estadística de España.

En cuanto a herramientas en las que formar, los talleres se centraron mayoritariamente en la herramienta de análisis R (R Core Team, 2022). Sin embargo, también hubo sesiones temáticas dedicadas a otros lenguajes, como Python (Van Rossum y Drake, 2009) y PostgreSQL (PostgreSQL Global Development Group, 2024), todos ellos *software* libre. El término de *software* libre, tal y como lo define Stallman (2002), remite a aquel software que proporciona a la persona usuaria la libertad de ejecutarlo, redistribuirlo y mejorarlo, siendo imprescindible para ello el acceso a su código fuente, sin necesidad de notificar a la autoría³. De acuerdo con dicho autor, ello no impide que un *software* libre pueda tener un carácter comercial, sino más bien que no se tendría restringida su copia, modificación o venta una vez se ha adquirido. Es por ello por lo que el espíritu del curso fue implementar y formar en el uso de *software* libre de acceso gratuito, principalmente por las ventajas de su distribución y la colaboración entre personas usuarias. Como eje central a nivel didáctico, el curso se centró en favorecer la comprensión de las instrucciones y la socialización en los distintos entornos de programación. De esta manera, el aprendizaje de una herramienta informática facilitaría el aprendizaje de herramientas familiares (Kay 2007): si una persona presenta bases previas en el manejo de R, es esperable que aprenda más fácilmente lenguajes como Python.

2 Este artículo ha sido posible gracias a la experiencia en el curso formativo de R del proyecto de Innovación Docente «Aula de Innovación Tecnológica para el Uso de Aplicaciones Informáticas Accesibles en las Ciencias Sociales» de la Universidad Complutense de Madrid, queriendo por mi parte agradecer el gran trabajo realizado por los docentes con los que compartí dicha labor: Jorge Blanco Iglesias, Alberto Sanz Gimeno y Alejandro Díaz Guerra Romero.

3 Aunque el término Open Source (Open Source Initiative, 2006) se muestra habitualmente vinculado al concepto de *software* libre, al incidir en el acceso al código, el desarrollo colaborativo del *software* y su redistribución libre, los conceptos no son estrictamente iguales. A diferencia del software libre, Open Source posibilita la existencia de derivados del *software* original sin la obligación de publicar el código fuente modificado, así como puede obligar a indicar explícitamente los cambios realizados sobre el código fuente, en función de las licencias. Un ejemplo de licencia que permite ambas situaciones es Apache 2.0. (Apache Software Foundation 2004).

2. Contexto actual de la enseñanza de Sociología y programas estadísticos

Así pues, una de las competencias clave en el marco de los programas de las asignaturas de Sociología, así como en la formación del mismo grado, remite a la adquisición de competencias asociadas al análisis de datos, las cuales han mostrado una relación favorable con el aprendizaje de la disciplina sociológica (Dietz, 2006). En esta línea, con el abaratamiento del coste de los ordenadores y la mayor difusión de Internet en las últimas décadas, se ha extendido la disposición de laboratorios o aulas de informática. No obstante, si atendemos al contexto actual dentro de las distintas asignaturas del Grado Sociología, en el caso de la Universidad Complutense de Madrid y en su plan de estudios vigente durante el curso 2022/2023, se observa como la utilización de programas informáticos no siempre es una práctica obligatoria o transversal a las distintas asignaturas. De este modo, queda relegada a asignaturas de metodología o de competencias digitales, utilizándose de forma opcional en otras asignaturas según la voluntad del equipo docente.

Además, la revisión de los distintos programas de asignaturas indica que cuando se hacía referencia a algún *software*, en casi todos los casos presentaban eran de carácter propietario. En este sentido, los *software* propietarios requieren, generalmente, la compra de una licencia para hacer uso del programa, extender sus funcionalidades e imposibilitando su distribución por parte de usuarias/os. Ello hará que tanto el alumnado como los profesionales del área solo puedan hacer uso de ella si pertenecen a una institución, empresa o tiene personalmente suficientes recursos económicos para costeárselo. Por el lado contrario, encontraríamos el *Software Libre* utilizado en el curso, el cual facilitará la descarga y utilización de los programas sin necesidad de la compra de una licencia. Por consiguiente, este será una alternativa estratégica al posibilitar su uso gratuito independientemente del lugar o condición administrativa en la que se encuentre.

En segundo lugar, podemos indicar las dificultades asociadas a la interfaz de usuario (por sus siglas en inglés, IU). La IU se puede entender como el punto de contacto entre la persona usuaria y la máquina, de manera que es a través de esta donde ejecuta o visualiza las principales órdenes que se le demandan. Dentro de la IU, podemos distinguir la Interfaz Gráfica de Usuario (por sus siglas en inglés, GUI) y la interfaz de línea de comandos (por sus siglas en inglés, CLI). La primera facilita el uso del programa a través de pestañas o ventanas, haciéndolo más amigable para la población menos familiarizada con el código. Por su parte, la CLI se centra en la programación y los códigos de instrucciones hacia el programa, siendo más retador para la gente en formación, pero con claras ventajas a la hora de demostrar que se ha utilizado y replicar los resultados, siendo un eje fundamental a la hora de evaluar prácticas de clase.

En este marco, sobre la IU podemos comentar que, por lo general, sea *software* propietario o libre, se requiere una formación mediante manuales del propio programa, cursos, etc. que o bien se realizan de manera autodidacta y por voluntad propia, o bien a través del diseño de demandas obligatorias (por ejemplo, la necesidad en asignaturas de grado o posgrado que requieran adquirir dicha destreza como requisito para su superación). En el caso de los *software* propietario, las ofertas formativas muchas veces dependen de la autorización y credenciales educativas de las empresas distribuidoras del *software*. De esta manera, en muchas ocasiones suelen requerir el pago y la adquisición del servicio formativo como paso previo a su uso. Por tanto, ello puede limitar las posibles barreras de entrada a usar un *software* nuevo, a menos que la potencialidad esperada motive su uso frente a otras prácticas realizadas anteriormente. Por su parte, los *software* libres por lo general tienden a tener mayor oferta formativa, tanto gratuita como de

pago. Ello se deberá fundamentalmente a que su uso suele estar más generalizado, al no depender su uso de la adquisición de una licencia, así como a la ausencia de restricciones para producir material formativo. Si bien la misma barrera inicial de aprendizaje está presente, es esperable que la mayor accesibilidad a recursos formativos haga ligeramente más atractiva su aprendizaje.

Aunque la mayoría de los programas permiten un buen desarrollo tanto de la GUI como de la CLI, la principal baza en la que han incidido su mejora por parte de los programas propietarios ha sido el desarrollo de una GUI. Así pues, se ha tendido a crear un entorno visual amigable para un público no necesariamente formado en la programación, como suele ser el que encontramos en titulaciones de Sociología. No obstante, el evitar enseñar a programar al alumnado de esta disciplina ha acabado generando una cada vez mayor resistencia o falta de conocimiento sobre esta. De forma colateral, también se genera una mayor dependencia hacia programas concretos, como consecuencia de una GUI diferente para cada *software* propietario. Ello supondrá una barrera de aprendizaje más de entrada, solo superada si se considera los beneficios frente a las desventajas, como el tiempo que tardará en formarse, que aporta dicho cambio.

Además, como dificultad añadida, se muestra más difícil para el personal docente la evaluación de contenido realizado directamente mediante una GUI. Así pues, en el caso de un resultado erróneo obtenido a través de GUI, se presentará la tarea de identificar qué procedimiento o qué parámetros se ha ejecutado. De esta manera, se complejiza la devolución de retroalimentación en la tarea. Por su parte, el hacer uso del análisis estadístico mediante formatos como la CLI permite al profesorado reconocer los procedimientos concretos utilizados y las opciones pedidas con mayor precisión (Wallace *et al.*, 2012), de manera que pueden evaluar con mayor evidencia los conocimientos demostrados, y facilitar por consiguiente propuestas de mejora.

Otra diferencia que debemos señalar con respecto al *software* propietario reside en que este, por lo general, limita su desarrollo al equipo encargado de diseñar el programa, si bien en ocasiones puede estar abierto a la participación de la comunidad de clientes. Por su lado, el *software* libre permitirá, a cualquier persona usuaria, el acceso al código fuente siempre. Así mismo, posibilita el desarrollo de extensiones que permitan ampliar la cantidad y calidad de sus funcionalidades. Aunque el grupo encargado del *software* propietario por lo general estará también atento a su mejora, el *software* libre permite una adaptación más rápida y a medida realizada por la propia comunidad, de manera colaborativa, instantánea y sin necesidad de permisos. En el caso de R, la mayor parte de las extensiones de funcionalidades y sus actualizaciones se encuentra en el denominado *The Comprehensive R Archive Network* (<https://cran.r-project.org/>), donde tendremos la libertad de descargar tanto el programa base, sus extensiones (librerías), así como los ficheros para distribuirlos independientemente.

Este carácter abierto se considerará un elemento esencial, tanto para adaptar las funciones a los nuevos objetos de investigación de sociología, como a novedosos análisis y, con especial atención para el uso docente, al desarrollo de entornos adaptados al Diseño Universal de Aprendizaje (Centro de Tecnología Especial Aplicada, CAST). De esta manera, podremos encontrar extensiones orientadas a la visión por personas con distintos tipos de daltonismo, la personalización de las fuentes para la dislexia, o el aumento del tamaño de los elementos. Así mismo, como consecuencia de un mayor énfasis en la CLI por parte del *software* libre, basado en la escritura de código, se hace también más accesible para personas con ausencia de visión al combinarlo con conversores de texto a voz.

3. Demandas en la enseñanza de R

Dicho todo lo anterior, se hace un llamamiento a la adaptación de la Sociología a un escenario marcado por la convergencia entre el mundo físico y el mundo digital. Es en esta convergencia donde aparecerán datos sobre realidades ya existentes, así como se crearán nuevos objetos de estudio sociológicos. Por ejemplo, podríamos identificar la necesidad de atender a las comunidades virtuales, los patrones de navegación en Internet, el análisis de los elementos en las páginas web, o el efecto de los algoritmos en las relaciones sociales digitales. De igual forma, con el desarrollo de R y Python, cada vez se puede llevar a cabo un análisis de datos más sofisticado, de manera que se muestra imprescindible para la Sociología entrar en el marco de las Ciencias Sociales Computacionales. Aunque aún no existen datos para cuantificar el uso de R o Python entre los profesionales de la Sociología, encuestas como la de *Stack Overflow* (2024) ya vienen señalando como ambos lenguajes son ampliamente utilizados por las personas desarrolladoras de código.

Para intentar responder a las posibles demandas, presentaremos a continuación los resultados del cuestionario obtenido al inicio de la primera sesión del curso formativo gratuito en R y otras herramientas de *software* libre, con un total de 36 respuestas de las 50 plazas ofertadas y ocupadas en el curso⁴ (Aula de Innovación Tecnológica para el Uso de Aplicaciones Informáticas Accesibles en las Ciencias Sociales, 2024), lo que indica que fue cumplimentado por la mayor parte las personas inscritas. En términos muestrales, se estaría ante un muestreo no probabilístico de conveniencia (Kish, 1965), al ser una muestra seleccionada por estar presente y accesibles en la primera sesión de los talleres. Es por ello por lo que los datos recogidos se limitan a tener un carácter puramente exploratorio, siendo identificables sesgos de autoselección (Campbell y Stanley, 1963) por parte del alumnado, tanto en la respuesta, como en la inscripción al curso. En particular, es atribuible que se inscribieran al curso y respondieran aquellas personas que más interés tuviera hacia aprender análisis de datos, lo que nos presenta la limitación de no tener la radiografía del conjunto del alumnado de Sociología y áreas afines. Por su parte, la administración del cuestionario fue de manera autoadministrada y accesible a través de un formulario web, con un tiempo estimado de cumplimentación de 10 minutos máximos.

El análisis de los datos y la creación de gráficos ha sido posible gracias al *software* R y a los paquetes *dplyr* (Hadley *et al.*, 2023) y *ggplot2* (Wickham, 2016), siendo coherente con los argumentos esgrimidos en este artículo. Debe adelantarse que se presenta una limitación añadida en los datos recogidos: no se recogió en este cuestionario el género de las personas encuestadas. Por consiguiente, resulta complicado detectar un posible sesgo de género entre las personas asistentes a esta primera sesión. Así mismo, ante la posibilidad de obtenerlo mediante el cruce con otras bases de datos, hemos considerado no llevarlo a cabo dado que podría comprometer el carácter anónimo de los datos.

De acuerdo con el cuestionario, se preguntó al alumnado del curso inscrito «¿Qué grado de provecho consideras que puedes sacar a R en tu trayectoria profesional y/o académica futura?». En este sentido, el 52,8% indicó que R influirá mucho, seguido de un 41,7% que señalaba que bastante, y un escaso 5,6% que piensa que solo le influirá algo. En líneas generales, estos porcentajes dan pistas acerca de la creciente preocupación del alumnado sobre el papel de R, conscientes de una mayor demanda en la práctica socio-

⁴ A modo de información previa, un 41,67% señalaba tener Bachillerato, FP y/o Acceso a la Universidad como nivel más alto finalizado. Como es obligatorio que estén matriculados en la Universidad, es esperable que este porcentaje corresponda a personas en vías de obtención del título de Grado, lo que sugiere un gran interés por aprender R en estadios iniciales. Un 38,88% declaraba haber obtenido un título de Grado, Máster o equivalente, y un 19,44% estudios de Doctorado. Aunque no se puede comprobar con seguridad, estos últimos es posible que correspondan mayormente con profesorado o personal investigador doctor de la Universidad.

lógica. En cuanto a las expectativas del curso en general (ver Tabla I), cubierto mayormente por sesiones de R, pero complementadas con algunas de Python y PostgreSQL, la mayoría del alumnado indicó que lo cursaba para aprender a hacer análisis estadístico con R, introducirse en la herramienta, así aprender a hacer análisis cualitativo. Como comentaremos en el apartado de potencialidades, llamará la atención esta última categoría de respuesta mayoritaria, en tanto a que hoy día no hay ninguna extensión actualizada que permita hacer un análisis cualitativo, de la misma forma que se hace con *software* propietario.

Tabla I. ¿Qué esperas obtener de este curso? [respuesta múltiple] (n = 36)

OPCIÓN DE RESPUESTA	FRECUENCIA
Aprender a realizar análisis estadístico en R	77,78 %
Introducirme en el conocimiento de R	61,11 %
Aprender a realizar análisis cualitativo en R	55,56 %
Tener otra alternativa a los programas que empleo habitualmente	27,78 %
Me gustaría trabajar con software libre	22,22 %
Gráficos y mapas	2,78 %

Fuente: elaboración propia en base a cuestionario previo del curso.

En cuanto a las expectativas del curso a partir de la formación en R (ver Tabla II), se aprecia como la mayor parte del alumnado tiene un interés en aprender R para cuestiones de análisis de datos, más que de automatización en la recolección de estos. Dicha predisposición puede deberse a que el análisis suele ser la única funcionalidad que suele implementar el *software* tradicional de corte propietario, por lo que es posible que no conocieran la potencialidad de poder extraer datos gracias al *software*, o la elaboración de cuadros de mando para compartir los resultados a terceros de manera interactiva.

Tabla II. ¿Tienes algunos usos concretos que te gustaría implementar en tu trabajo diario a partir de R? [respuesta múltiple] (n = 36)

OPCIÓN DE RESPUESTA	FRECUENCIA
Realizar análisis descriptivo de datos	86,11 %
Realizar representación gráfica de datos	77,78 %
Trabajar con grandes bases de datos	75,00 %
Elaborar modelos multivariantes avanzados	61,11 %
Realizar análisis cualitativo	50,00 %
Análisis de redes sociales	41,67 %
<i>Social media analysis</i>	36,11 %
Automatización de procesos de descarga de datos	33,33 %
Automatización de procesos de generación de informes	33,33 %
Elaboración de cuadros de control/mando/ <i>dashboard</i>	22,22 %

Fuente: elaboración propia en base a cuestionario previo del curso.

Como última demanda, y no constatable a través del cuestionario, debe destacarse que cada vez más el aprendizaje y la utilización de R y Python se presenta tanto como requisito como mérito valorable en las distintas ofertas de trabajo, ya sea del ámbito profesional o científico (Kesari *et al.*, 2024). Es por ello por lo que debemos considerar esto como una demanda añadida, que desarrollaremos en el apartado de conclusiones como una de las posibles limitaciones de este artículo.

4. Usos potenciales de R

Expresado todo lo anterior, R presentará los siguientes usos potenciales, si atendemos a su uso tanto para la investigación sociológica como para la docencia. En primer lugar, debemos destacar su carácter orientado al análisis estadístico. En este sentido, R se plantea como una herramienta capaz de leer los datos de las principales entidades de producción de datos sociológicos, como el Centro de Investigaciones Sociológicas, y analizarlas ya sea de manera univariada, bivariada o multivariada. En principio, R permite trabajar con lo que se denomina factor, donde se ubicarían las respuestas de tipo categórico, nominales (por ejemplo, las categorías de género, ciudad de residencia, etc.) u ordinales (la variable nivel de estudios, grado de afirmación con respecto a una afirmación, etc.). También puede trabajar con variables de tipo numérico, donde se encuentran las habituales de corte cuantitativo (la edad concreta, la cantidad de ingresos, etc.). Ello nos permitirá la posibilidad de contrastar hipótesis relacionadas con los objetos de investigación clásicos de la sociología, lo que hace que se muestre muy afín su implementación con todas las asignaturas de Sociología con cierto carácter empírico, donde el alumnado puede cruzar variables y elaborar modelos relacionados con el análisis sociológico.

No obstante, y a diferencia de la situación actual de los programas propietario, también extenderá sus funcionalidades a campos que, si bien no son nuevos en la sociología, si han sido poco frecuentes hasta ahora. Será el caso del uso de datos geoespaciales, el análisis de redes sociales, el análisis de sentimientos a partir de palabras de texto, etc. Es por ello por lo que R se presenta como un programa versátil para atender a los nuevos retos teóricos y empíricos a la hora de atender a investigaciones actualizadas dentro del marco de la Sociología, donde estos últimos objetos de estudios empiezan a cobrar más relevancia. De esta manera, se lograría además conectar mejor con el alumnado, más familiarizado con el entorno digital y tener objetos de estudios que les puedan resultar más próximos (por ejemplo, las interacciones en redes sociales digitales o cómo se distribuye visualmente el nivel educativo alcanzado por coordenadas de una ciudad).

En estrecha relación con la anterior, la segunda funcionalidad que nos aportará la implementación de R para la docencia tendrá que ver con la visualización de los datos. Dicha funcionalidad se muestra como una de las principales prácticas en lo que respecta a la comunicación científica, ya sea para demostrar que se ha analizado los datos adecuadamente, para realizar presentaciones en case, así como en todo lo que respecta a la comunicación y divulgación científica dentro y fuera de la universidad. Acudiendo al caso concreto de R, el hecho de entornos que demanden un código para la producción de imágenes se mostrará ampliamente productivo en términos de conocimiento (Healy y Moody, 2014). En esta línea, esto se dará en el hecho de tener un código de referencia del cual aprender o replicar los datos, así como la posibilidad de aportar al equipo docente por el alumnado un código que refleje demostrar con mayor conocimiento sobre los datos. Así mismo, se pueden extender funcionalidades más allá de la que suelen presentar los programas propietarios más extendidos: la posibilidad de producir mapas que representen datos geoespaciales con el paquete *sf* (Pebesma, 2018; Pebesma y Bivand, 2023), la producción de datos interactivos mediante *plotly* (Sievert, 2020), o la producción de cuadros de mando que mezclen datos y gráficos a través de *shiny* (Chang, 2024). Como plus añadido, desde la docencia pueden elaborarse gráficos didácticos que puedan hacer uso fácilmente de distribuciones y modelados estadísticos comunes en Sociología. Por ejemplo, se puede representar mediante código y su correspondiente visualización el Teorema Central del Límite, siendo uno de los principios esenciales en toda estadística inferencial, multivariante y en el muestreo en Sociología, o visualizar las áreas críticas al llevar a cabo una distribución y prueba chi-cuadrado.

No estrictamente relacionado con el análisis, R (y también Python) ofrecerán también potenciales en términos de recolección de datos. En esta línea, mediante paquetes ofrecen la posibilidad de recoger datos de distintas páginas webs de manera automática. Ello puede ser o bien facilitado por la propia web, a través de lo denominado como API, el cual permite la descarga automática de datos, o bien mediante prácticas denominadas como web *scraping*, donde se extrae de manera automatizada una información seleccionada de una página web (Robles *et al.*, 2022). En el caso de esta último, el paquete de *rvest* (Wickham, 2022) ofrece la funcionalidad de llevarlo a cabo. No obstante, requiere cierto conocimiento de lenguajes de programación web, así como del aseguramiento por parte del docente o equipo investigador de que es legítimo obtener datos de dicha web.

Finalmente, la última funcionalidad que podemos destacar de R en torno a la docencia reside en su capacidad a la hora producir tanto un entorno de aprendizaje autónomos como entornos de aprendizaje cooperativos. Por un lado, los entornos de aprendizaje autónomos permiten ser una extensión práctica de las habituales soluciones en problemas con ejercicios, donde el alumnado puede comprobar si el código que ha escrito es correcto. Por otro lado, la constitución de entornos de aprendizaje colaborativos permite la interacción en tiempo real del alumnado y el profesorado. De esta manera, se permite la supervisión y corrección de las prácticas en directo por parte del docente. Así mismo, también se facilita la elaboración de trabajos en grupo de esta forma, ya que el trabajo se puede ejecutar en línea y sin recurrir a la descarga y ejecución del archivo local, tal y como suele ser frecuente en *software* propietario. Esta cuestión se planteará con mayor detalle en el apartado de recomendaciones.

5. Desafíos en la docencia a través de R

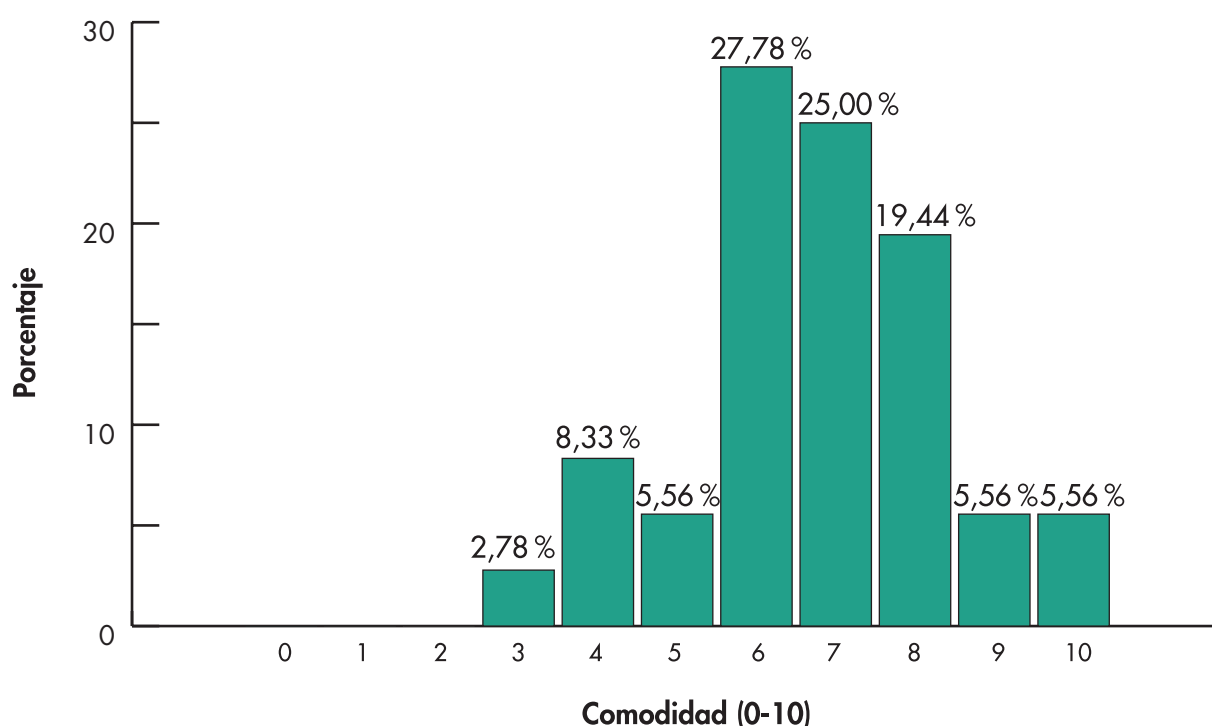
Si atendemos a los principales desafíos de la docencia con R, en primer lugar, deberemos señalar el filtro socioeconómico y el de género. Como cualquier otra herramienta informática, y especialmente patente en el entorno educativo con la pandemia que forzó el aprendizaje en línea, la tenencia de ordenadores y la conexión a Internet se encuentra determinada por el origen socioeconómico y de clase social del alumnado (Müller *et al.*, 2023). De esta manera, la carencia de un ordenador personal, así como de que cumpla los requisitos técnicos suficientes, hace que el alumnado dependa exclusivamente de los dispositivos presentes en su universidad o centro educativo. A pesar de ello, herramientas como R facilitan un mayor acceso con respecto a las alternativas propietarios, fundamentalmente, porque no requiere de la compra de una licencia para su uso. De igual forma, las expectativas de género también se construyen como barreras a la hora de acceder a lenguajes de programación como R. No tanto por su potencial acceso, sino por las fronteras multifactoriales que se producen en el hecho de que haya menos mujeres que hombres en los aspectos relacionados con la informática, fundamentalmente la visión de la informática como un entorno casi exclusivamente masculino (Cheryan *et al.*, 2017).

Con carácter más curricular, debe destacarse que el alumnado de Ciencias Sociales no suele estar familiarizado con el hecho de programar, situado en el imaginario como una competencia más afín a otras disciplinas como las ingenierías técnicas. Un reto añadido dentro del alumnado y personal de Ciencias Sociales recae en que no necesariamente ha recibido previamente conocimientos de programación. En este sentido, tanto la atribución de las competencias informáticas al ámbito tecnológico en los currículos preuniversitarios, como la utilización de *software* estadístico basado exclusivamente en interfaces visuales, ha llevado a que por lo general el terreno de la programación se muestre como un terreno árido, donde el

alumnado y profesionales de Ciencias Sociales no es de su competencia. De esta manera, se espera que, al no estar familiarizado con dichas competencias, la barrera inicial de entrada y la curva de aprendizaje sea más inclinada para la gente no familiarizada por la programación.

Atendiendo a los datos del cuestionario del proyecto de Innovación Docente, la mayor parte de la muestra indica que su formación previa en matemáticas y estadística es regular (44,44%) o buena (41,67%). Con gran distancia en puntos porcentuales, un 8,33% indica que ha sido mala, mientras que un 2,78% señala estar en los extremos de muy mala o buena. Con un argumento similar, en el Gráfico 1 se presentará la escala 0-10 de comodidad del alumnado del curso a la hora de trabajar con datos estadísticos.

Gráfico 1. En una escala de 0 a 10, donde 0 es 'nada cómodo/a' y 10 'totalmente cómodo/a', ¿qué comodidad sientes al trabajar con datos estadísticos?



Fuente: elaboración propia en base a cuestionario previo del curso (n = 36).

Así pues, de acuerdo con dicho gráfico, destacará como los mayores porcentajes se encuentran en los valores 6, 7 y 8 de la escala. Esto sugiere que, aunque no todo el alumnado de Ciencias Sociales tiene suficiente competencia al respecto, sería importante reflexionar si el hecho de apuntarse al curso es una elección que ha tomado el alumnado que más competente en dicho asunto, en relación con el filtro de autoselección mencionado anteriormente. Es por ello por lo que este resultado se tendrá en cuenta en futuros cuestionarios e investigaciones, de manera que podamos comprobar de manera efectiva si esto responde a un sesgo de autoselección en los participantes del curso.

Finalmente, el último reto que queremos destacar es la dificultad en R para llevar a cabo un análisis más próximo a la metodología cualitativa, de manera que pudiera integrarse en la docencia de asignaturas relacionadas con este tipo de metodología. Si bien históricamente existieron paquetes como RQDA (Huang, 2018), que extendían las funcionalidades a la posibilidad de hacer un análisis al estilo de los

programas CAQDA (Computer Assisted/Aided Qualitative Data Analysis Software), dicho paquete se encuentra actualmente discontinuado. Por ello, se hace difícil su ejecución hoy día para la mayoría de los usuarios, además de que presentaba una interfaz gráfica menos pulida visualmente que las alternativas de software propietario. Para un análisis de contenido o de los significados asociados a las palabras encontraríamos paquetes como Quanteda (Benoit *et al.*, 2018) en R, u el software Orange (Demšar *et al.*, 2013) basado en Python. No obstante, será más afín a la metodología cuantitativa que al análisis con Teoría Fundamentada (Carrero *et al.*, 2012) o del Discurso común en la Sociología (Conde, 2009).

6. Conclusiones y recomendaciones de futuro

El lenguaje de programación R, junto Python, es cada vez uno de los programas de análisis de datos más demandados en la formación para el alumnado y profesionales de Sociología y Ciencias Sociales afines. En este artículo, hemos propuesto los principales resultados obtenidos a través de un cuestionario previo a la puesta en acción de un curso formativo en herramientas de uso de *software* libre, principalmente R, dirigido a alumnado, personal técnico y docente de la Facultad de Ciencias Políticas y Sociología de la Universidad Complutense de Madrid.

Por nuestra parte, hemos señalado como limitación el hecho de que no se encuentra recogida la variable género en el cuestionario, o al menos, no sin comprometer la anonimización de los datos. Como se han señalado desde distintas investigaciones que han incidido en la brecha de género al respecto de la formación en informática, se sigue atribuyendo que las formaciones y trabajos computacionales son generalmente para el género masculino (Botella *et al.*, 2019; Cheryan *et al.*, 2017; Jacinto *et al.*, 2020). Es por ello por lo que indagar en esta cuestión en la siguiente edición se muestra como uno de los ejes cruciales, de manera que se pueda complementar los datos asociados a dicha cuestión, más teniendo en cuenta que el alumnado total en Ciencias Sociales a los que se dirige en mayor o menor medida el curso, es mayormente femenino. En este sentido, en 2023, las mujeres representaban el 61,2% del total de estudiantes matriculadas en la Universidad Complutense de Madrid en el área de Ciencias Sociales y Jurídicas en general, así como el 59,3% en la Facultad de Ciencias Políticas y Sociología en particular (Observatorio de Igualdad de la Universidad Complutense de Madrid, 2024).

También, en este artículo hemos señalado como R puede llegar a observarse como un elemento ajeno a la Sociología, lo que acaba llevando a que el alumnado finalmente inscrito en el curso pueda tener un autofiltro de selección y considere que tenga una cierta competencia a nivel técnico con el uso de la informática. No obstante, para abordar los principales desafíos con los que nos encontramos, en el marco de la docencia en titulaciones de Sociología, propondremos las siguientes recomendaciones:

En primer lugar, debemos señalar la necesidad de un material didáctico de programación orientado a los objetos de investigación más comunes en Sociología y Ciencias Sociales en general. De esta manera, si se trabaja con objetos que resulten familiares, sobre todo si se ha estudiado previamente en otras asignaturas teóricas, los sentimientos de logro y la motivación inicial se verán significativamente reforzados. Ejemplos de ello puede ser el análisis de datos orientado a analizar actitudes, datos sociodemográficos, y desigualdades como la brecha salarial por género, la brecha digital o el origen socioeconómico. Es en estos asuntos donde la Sociología y la interdisciplinariedad con Ciencias Sociales afines donde tendrán que hablar, al comprender mejor el contexto y las problemáticas sociales, de manera que el análisis de datos no se plantee con una interpretación única y limitada al dato. De esta manera, se debe tener en cuenta no

solo cómo se produce el dato, sino también su efecto en la población al poder acentuar las desigualdades (O'Neil, 2017). Relacionado con ello, es posible que una falta de especificidad acerca de los objetos de investigación sociológicos, así como cierta asociación de la disciplina como ajena a la programación, es lo que esté detrás de un menor desarrollo que otras disciplinas sociales con atribuciones empíricas y cuantitativas más extendidas, como la economía (Hudson, 2017).

En segundo lugar, y en estrecha relación con lo planteado anteriormente, la formación del personal docente se muestra como un hito indispensable. Aunque la formación autodidacta puede favorecer su uso en aquellos casos que cuenten con una gran motivación intrínseca, la formación por parte de departamentos, proyectos de innovación docente y planes de formación del profesorado dentro de los entornos universitarios se muestran indispensables para una implementación satisfactoria en términos de usos de la herramienta. Será por ello importante la disposición de recursos digitales a través de los cuales tanto el profesorado como el alumnado de las distintas materias pueda formarse. Así pues, cada vez es mayor el número de cursos gratuitos o de pago en el mercado y a través de distintas plataformas, así como distintos tutoriales y materiales de apoyo en la red. A nivel de manuales, la mayoría de estos se encuentran en inglés (Davies, 2016; Matloff, 2011), si bien en español cada vez también aparece un mayor número de manuales al respecto aplicados a las Ciencias Sociales, como los manuales de Bouso (2018) y de Rey-Sáez y Estrada (2024).

En tercer lugar, se sugiere que el proceso de adaptación a *software* libre debe ser una cuestión gradual, tanto para el alumnado, como para el personal docente en Sociología. Por un lado, para el alumnado deberá presentarse cierta secuenciación para que este no se vea abrumado o frustrado ante una cantidad demasiado elevada de conocimiento y práctica. Por otro lado, y no menos importante, la transición a programas de libre acceso como R por parte de docentes deberá plantearse también de manera progresiva. Ello se deberá a que las ventajas de utilizar un nuevo programa para el análisis estadístico y de datos no se mostrará como una ventaja a corto plazo en el marco de la preparación de la docencia. No obstante, una implementación paulatina puede amortiguar esta cuestión, venciendo progresivamente la resistencia hacia el cambio de programa. Como factor añadido, una mayor concienciación sobre las ventajas del *software* en la corrección, así como de la gratuidad y adaptabilidad, puede facilitar dicha cuestión. Es esperable que según más se extienda su uso dentro y fuera de la universidad, más fácil sea sumarse a la iniciativa. Como hemos señalado, también queda pendiente que aparezca entre las extensiones más extendidas un paquete que ayude para la realización de análisis cualitativo para atraer al alumnado y profesionales más afines a este.

Así mismo, este reto puede solventarse parcialmente mediante procesos mixtos de uso de interfaces gráficas y código, lo que puede hacer menos agresiva dicha transición. De esta manera, encontraremos entornos como el presentado por el paquete Rcmdr (Fox, 2005; Fox, 2017; Fox y Bouchet-Vatat, 2023) y su interfaz RCommander, similar a la de los *software* propietarios clásicos de la estadística, o el programa Jamovi (*The jamovi project*, 2024), el cual tiene como motor base R, pero que funciona sin necesidad de instalarlo de forma previa. También serán de ayuda entornos de programación interactiva, como Jupyter Notebook y JupyterLab (Project Jupyter, 2015; Project Jupyter, 2018). Siendo la segunda una versión más avanzada y flexible de la primera, archivos de este tipo asociados a R y Python suponen entornos que permiten acompañar código ejecutable de texto explicativo, elementos visuales como gráficos y resultados en el propio documento. De esta manera, el alumnado puede ejecutar y explorar fragmentos parciales de

código para ver los resultados. Como beneficio añadido, dichos documentos son fácilmente distribuibles, lo que permite tanto compartir material didáctico de manera descentralizada, como compartir el código y resultados en aras de transparencia en un contexto académico cada vez más abierto. Aunque menos interactivo y más orientado a informes, en R encontraríamos además los ficheros RMarkdown (Allaire *et al.*, 2023), con también posibilidades didácticas al respecto (Dvorak *et al.*, 2019). A este respecto, estos tres entornos generaran archivos locales, situados en el ordenador de quien lo ejecuta. Como elemento complementario o alternativo, la empresa Google ofrece hoy día de manera gratuita acceso a Google Colab (Google, 2024), siendo esta una plataforma gratuita que permite ejecutar tanto código de R como de Python de manera similar a los archivos locales de Jupyter, pero con la ventaja de ejecutarse en línea sin instalación local. De esta manera, se permite el trabajo colaborativo entre alumnado y profesorado en tiempo real, si bien debemos destacar que no está garantizado su uso gratuito al depender de una empresa privada centralizada.

Por último, queremos finalizar este artículo señalando la importancia de la universidad como un espacio clave para la formación de la ciudadanía del futuro. El fomento del uso de software no solo permitirá superar las barreras económicas asociadas a el software propietario. También contribuirá a la democratización del aprendizaje y al desarrollo de conocimiento. Se estará por tanto ante un enfoque emancipador hacia la accesibilidad, la libertad y la equidad, alcanzando cada vez a una población más amplia, con independencia de la ubicación geográfica o contexto socioeconómico del sujeto. Es por todo ello que disciplinas como la Sociología, así como otras disciplinas científicas, deberán velar por integrar estos enfoques en sus prácticas para fomentar un análisis crítico en torno a las implicaciones sociales en el acceso al conocimiento. Al utilizar herramientas de *software* libre, se estará impulsando la inclusión educativa, así como la capacidad de la ciudadanía para participar en la producción y distribución del conocimiento. Se sugiere, por consiguiente, la indagación en futuras investigaciones acerca de la difusión y uso de aplicaciones de *software* libre, como R u aplicaciones ofimáticas, extrapolables al conjunto de población.

Reseñas bibliográficas

Allaire, Joseph J.; Xie, Yihui; McPherson, Jonathan; Luraschi, Javier; Ushey, Kevin; Atkins, Aron; Wickham, Hadley; Cheng, Joe; Chang, Winston y Iannone, Richard (2023). *rmarkdown: Dynamic Documents for R*. Versión 2.20.

Apache Software Foundation (2004). *Apache License*. Version 2.0. Disponible en: <https://www.apache.org/licenses/LICENSE-2.0>, consultado por última vez 19 de enero de 2025.

Aula de Innovación Tecnológica para el Uso de Aplicaciones Informáticas Accesibles en las Ciencias Sociales (2024). *Base de Datos de Cuestionarios del proyecto 2023-2024*.

Botella, Carmen; Rueda, Silvia; López-Iñesta, Emilia y Marzal, Paula (2019): “Gender Diversity in STEM Disciplines: A Multiple Factor Problem”. *Entropy*, 21 (1), 30. DOI: <https://doi.org/10.3390/e21010030>.

Benoit, Kenneth; Watanabe, Kohei; Wang, Haiyan; Nulty, Paul; Obeng, Adam; Müller, Stefan y Matsuo, Akitaka (2018): “quanteda: An R Package for the Quantitative Analysis of Textual Data”. *Journal of Open Source Software*, 3 (30), 774. DOI: <https://doi.org/10.21105/joss.00774>.

- Bouso, Jesús (2018). *El paquete estadístico R*. (2.^a edición revisada). Madrid: Centro de Investigaciones Sociológicas.
- Carrero, Virginia; Soriano, Rosa y Trinidad, Antonio (2012). *Teoría Fundamentada Grounded Theory*. Madrid: Centro de Investigaciones Sociológicas.
- Campbell, Donald y Stanley, Julian (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally & Company.
- Chang, Winston; Cheng, Joe; Allaire, Joseph; Sievert, Carson; Schloerke, Barret; Xie, Yihui; Allen, Jeff; McPherson, Jonathan; Dipert, Alan; Borges, Barbara (2024). *shiny: Web Application Framework for R*. Versión 1.8.1.1.
- Cheryan, Sapna, Ziegler, Sianna A., Montoya, Amanda K. y Jiang, Lily (2017): “Why Are Some STEM Fields More Gender Balanced than Others?”. *Psychological Bulletin*, 143 (1), 1-35. DOI: <https://doi.org/10.1037/bul0000052>.
- Conde, Fernando (2009). *Análisis sociológico del sistema de discursos*. Madrid: Centro de Investigaciones Sociológicas.
- Davies, Tilman M (2016). *The Book Of R: A First Course In Programming And Statistics*. San Francisco: No Starch Press.
- Demšar, Janez; Curk, Tomaž; Erjavec, Aleš; Gorup, Črt; Hočevar, Tomaž; Milutinović, Mitar; Možina, Martin; Polajnar, Matija; Toplak, Marko; Starič, Anže; Štajdohar, Miha; Umek, Lan; Žagar, Lan; Žbontar, Jure; Žitnik, Marinka; Zupan, Blaž (2013): “Orange: Data Mining Toolbox in Python”. *Journal of Machine Learning Research*, 14(Aug): 2349-2353.
- Díaz de Rada, Vidal; Domínguez, Juan Antonio y Pasadas, Sara (2019). *Internet como modo de administración de encuestas*. Madrid: Centro de Investigaciones Sociológicas.
- Dietz, Tracy L. (2006): “The Effect of Data Analysis Modules in the Introductory Sociology Course: Lessons for the Social Sciences”. *Innovative Higher Education*, 31 (1), 27-42. DOI: <https://doi.org/10.1007/s10755-006-9006-8>.
- Dvorak, Tomas; Halliday, Simon; O'Hara, Michael y Swoboda, Aaron (2019): “Efficient Empiricism: Streamlining Teaching, Research, and Learning in Empirical Courses”. *The Journal of Economic Education*, 50 (3), 242-257. DOI: <https://doi.org/10.1080/00220485.2019.1618765>.
- Fielding, Nigel G.; M. Lee, Raymond y Blank, Grant (2017). *The SAGE Handbook of Online Research Methods*. (Second edition). Electronic Version. Londres: SAGE.
- Fox, John (2005): “The R Commander: A Basic Statistics Graphical User Interface to R”. *Journal of Statistical Software*, 14 (9), 1-42.
- Fox, John (2017). *Using the R Commander: A Point-and-Click Interface to R*. Boca Raton FL: Chapman and Hall/CRC Press.
- Fox, John y Bouchet-Valat, Milan (2023). *Rcmdr: R Commander*. Versión 2.9-1.

- Google (2024). *Google Colab*. Sitio web. Disponible en: <https://colab.google/>, consultado por última vez 19 de enero de 2025.
- Healy, Kieran y Moody, James (2014): “Data Visualization in Sociology”. *Annual Review of Sociology*, 40 (1), 105-128. DOI: <https://doi.org/10.1146/annurev-soc-071312-145551>.
- Huang, Ronggui (2018). *RQDA: Qualitative Data Analysis*. Versión 0.3-1.
- Hudson, John (2017): “Identifying Economics’ Place Amongst Academic Disciplines: A Science or a Social Science?”. *Scientometrics* 113:735-750. DOI: <https://doi.org/10.1007/s11192-017-2519-3>.
- Jacinto, Claudia; Millenaar, Verónica; Roberti, Eugenia; Burgos, Alejandro y Sosa, Mariana (2020): “Mujeres estudiantes en Programación: entre la reproducción y las nuevas construcciones de género. El caso de la formación en el nivel medio técnico en la Ciudad de Buenos Aires”. *Revista de Sociología de la Educación-RASE*, 13 (3):432-450. DOI: <https://doi.org/10.7203/RASE.13.3.16605>.
- Kay, Robin H. (2007): “Learning Performance and Computer Software: An Exploration of Knowledge Transfer”. *Computers in Human Behavior*, 23 (1), 333-352. DOI: <https://doi.org/10.1016/j.chb.2004.10.029>.
- Kesari, Aniket; Kim, Jae Yeon; Shah, Sono; Brown, Taylor; Ventura, Tiago y Law, Tina (2024): “Training Computational Social Science PhD Students for Academic and Non-Academic Careers”. *PS: Political Science & Politics*, 57 (1):101-106. DOI: <https://doi.org/10.1017/S1049096523000732>.
- Kish, Leslie (1965). *Survey Sampling*. New York: Wiley.
- Matloff, Norman (2011). *The Art of R Programming: A Tour of Statistical Software Design*. San Francisco: No Starch Press.
- Müller, Fabian; Goudeau, Sébastien; Stephens, Nicole M.; Aelenei, Cristina y Sanitioso, Rasyid Bo (2023): “Social-Class Inequalities in Distance Learning During the COVID-19 Pandemic: Digital Divide, Cultural Mismatch, and Psychological Barriers”. *International Review of Social Psychology*, 36 (1), 3. DOI: <https://doi.org/10.5334/irsp.716>.
- Observatorio de Igualdad de la Universidad Complutense de Madrid (2024). *Distribución del estudiantado de la UCM según género evolución desde el curso 2000-2001 hasta el curso 2023-2024*. Disponible en: <https://www.ucm.es/unidaddeigualdad/distribucion-de-estudiantes-de-la-ucm-segun-genero->, consultado por última vez 19 de enero de 2025.
- Open Source Initiative (2006). *The Open Source Definition*. Página web. Disponible en: <https://opensource.org/osd>, consultado por última vez 19 de enero de 2025.
- O’Neil, Cathy (2017). *Armas de destrucción matemática: cómo el Big Data aumenta la desigualdad y amenaza la democracia*. Madrid: Capitán Swing.
- Pebesma, Edzer (2018): “Simple Features for R: Standardized Support for Spatial Vector Data”. *The R Journal*, 10 (1), 439-446. DOI: <https://doi.org/10.32614/RJ-2018-009>.
- Pebesma, Edzer y Bivand, Roger (2023). *Spatial Data Science: With Applications in R*. New York: Chapman and Hall/CRC. DOI: <https://doi.org/10.1201/9780429459016>.

- R Core Team (2022). *R: A Language And Environment For Statistical Computing*. R Foundation For Statistical Computing. Software. Vienna, Austria. Disponible en: <https://www.R-project.org/>, consultado por última vez 19 de enero de 2025.
- PostgreSQL Global Development Group (2024). *PostgreSQL 17.2 Documentation*. Disponible en: <https://www.postgresql.org/files/documentation/pdf/17/postgresql-17-A4.pdf>, consultado por última vez 19 de enero de 2025.
- Project Jupyter (2015). *Jupyter Notebook*. Software. Disponible en: <https://jupyter.org/>, consultado por última vez 19 de enero de 2025.
- Project Jupyter (2018). *JupyterLab*. Software. Disponible en: <https://jupyter.org/>, consultado por última vez 19 de enero de 2025.
- Rey-Sáez, Ricardo y Estrada, Eduardo (2024). *Análisis de datos con R en ciencias sociales y de la salud*. Madrid: Síntesis.
- Robles, José Manuel; Rodríguez, J. Tinguaro; Caballero, Rafael y Gómez, Daniel (2022): *Big data para científicos sociales. Una introducción*. Madrid: Centro de Investigaciones Sociológicas.
- Sievert, Carson (2020). *Interactive Web-Based Data Visualization with R, Plotty, and Shiny*. Florida: Chapman and Hall/CRC.
- Stack Overflow (2024). *Developer Survey*. Disponible en: <https://survey.stackoverflow.co/2024/>, consultado por última vez 19 de enero de 2025.
- Stallman, Richard (2002). *Free Software, Free Society: Selected Essays of Richard M. Stallman*. Boston: GNU Press. Disponible en: <https://www.gnu.org/philosophy/fsfs/rms-essays.pdf>, consultado por última vez 19 de enero de 2025.
- Tinati, Ramine; Halford, Susan; Carr, Leslie y Pope, Catherine (2014): “Big Data: Methodological Challenges and Approaches for Sociological Analysis”. *Sociology*, 48 (4), 663-681. DOI: <https://doi.org/10.1177/0038038513511561>.
- The jamovi project (2024). *Jamovi*. Software. Versión 2.5. Disponible en: <https://www.jamovi.org>, consultado por última vez 19 de enero de 2025.
- Van Rossum, Guido y Drake, Fred (2009). *Python 3 Reference Manual*. Scotts Valley: CreateSpace.
- Wallace, Byron C.; Dahabreh, Issa J.; Trikalinos, Thomas A.; Lau, Joseph; Trow, Paul y Schmid, Christopher H. (2012): “Closing the Gap between Methodologists and End-Users: R as a Computational Back-End”. *Journal of Statistical Software*, 49 (5), 1-15. DOI: <https://doi.org/10.18637/jss.v049.i05>.
- Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Wickham, Hadley (2022). *rvest: Easily Harvest (Scrape) Web Pages*. Versión 1.0.3.
- Wickham, Hadley; François, Romain; Henry, Lionel; Müller, Kirill; Vaughan, Davis (2023). *dplyr: A Grammar of Data Manipulation*. Versión 1.1.0.

Nota biográfica

David Moreno Alameda es Personal Investigador en Formación en el Departamento de Antropología Social y Psicología Social de la Universidad Complutense de Madrid. Actualmente, se encuentra realizando una tesis doctoral sobre la medición y variación de la Identidad de Clase Social en España. Sus líneas de investigación son los Métodos de Encuesta, las Ciencias Sociales Computacionales, los experimentos en Ciencias Sociales, las desigualdades sociales y las percepciones subjetivas de la desigualdad.

ORCID: <https://orcid.org/0000-0003-4989-9507>.