

Ciencias Sociales y Humanidades también se escriben con R

Macarena Sol Quiroga (CIIPME-CONICET/UNaHur)
macarenaquiros@conicet.gov.ar

Noelia Ayelén Stetie (CONICET/UBA)
nstetie@conicet.gov.ar

Palabras clave: uso de R en Ciencias Sociales, uso de R en Humanidades, formación en R, investigación en R

Resumen

El lenguaje de programación R es ampliamente utilizado en diversas áreas de investigación, especialmente en el análisis de datos, conocimiento cada vez más necesario en la actualidad (Jordan et al., 2018). Aunque comúnmente se asocia su uso con las Ciencias Exactas y Naturales, R también desempeña un papel significativo en las Ciencias Sociales y Humanidades (Arnold & Tilton, 2015). Sin embargo, las investigaciones y los datos sobre su uso en estos ámbitos son escasos, al igual que los cursos diseñados específicamente para esta población (Sirkis, 2006; Vance et al., 2022). Esto motivó la realización de un trabajo de campo, que consistió en una encuesta para investigar cómo se utiliza R en las Ciencias Sociales y Humanidades. La investigación se centró en responder dos preguntas centrales: ¿cómo se utiliza R en las Ciencias Sociales y Humanidades?, ¿cómo aprendieron (y continúan aprendiendo) los usuarios a manejar este lenguaje de programación?

La encuesta fue realizada en *Google forms* y difundida por redes sociales; por lo tanto, se utilizó un muestreo por conveniencia (Hernández Sampieri et al., 2014). Todavía seguimos en proceso de recolección de datos, pero, hasta el momento, participaron 116 personas de 12 países de Latinoamérica, con un 66% de respuestas de Argentina. En cuanto a la edad, realizamos una segmentación por franjas etarias y contamos con respuestas de 8 personas que tienen entre 19 y 25 años, 65 personas que tienen entre 26 y 35 años, 30 personas entre 36 y 45 años, 10 personas entre 46 y 50 años, 1 persona entre 51 y 60 años y 2 personas mayores de 60. Respecto del género, participaron 43 mujeres, 71 varones y 2 personas que se identificaron como no binarias. Los participantes provienen de 14 carreras distintas, que incluyen áreas tan diversas como antropología, filosofía, derecho y geografía. Sin embargo, la mayor cantidad de respuestas se concentran en las áreas de sociología, psicología, economía y, en menor medida, lingüística.

Para evaluar el nivel de conocimiento sobre R presentamos una escala Likert de 5 puntos. La mayoría de las respuestas estuvo concentrada en los valores 3 ($n = 45$, 38.79%) y 2 ($n = 22$, 18.97%). Además, la mayoría de las personas declaró haber aprendido en cursos en instituciones públicas ($n = 70$, 60.34%).

De las 116 respuestas, casi la mitad de las personas trabajan simultáneamente en docencia e investigación ($n = 40$, 34.5%), seguido por investigación ($n = 29$, 25%). El resto de las respuestas se

dividen en docencia, investigación y desarrollo ($n = 11$, 9.48%), investigación y desarrollo ($n = 5$, 4.31%) y respuestas varias, tales como consultoría, administración pública, IT, docencia ($n = 2$, 1.72%) y coordinación de proyectos, clínica, recursos humanos ($n = 1$, .86%). Más allá de los perfiles de respuesta, la investigación fue la respuesta mayoritaria ($n = 79$, 68.1%), seguida por la docencia ($n = 56$, 48.3%) y más atrás por investigación y desarrollo ($n = 18$, 15.5%). En cuanto a los ámbitos de trabajo, la mayoría trabaja exclusivamente en el sector público ($n = 61$, 52.6%), en segundo lugar el sector privado ($n = 35$, 30.2%) y finalmente en ambos sectores ($n = 20$, 17.2%).

Respecto del uso de R, la mayoría realiza exclusivamente análisis de datos o análisis estadístico ($n = 63$, 54.3%); en segundo lugar, análisis de datos o estadístico junto con creación de *dashboards* o informes ($n = 16$, 13.8%). El resto de las respuestas es una combinación de análisis de datos o estadístico y otros usos, como *business intelligence*, creación de algoritmos o programación de rutinas. Solo dos respuestas no incluyeron análisis de datos: *business intelligence* junto con aplicación de algoritmos, y creación de *dashboards* o informes junto con desarrollo de aplicaciones o sitios web ($n = 1$, .86%). La opción análisis de datos o estadístico fue seleccionada en 113 respuestas (97.4%). En este sentido, los paquetes más utilizados se vinculan a la limpieza de datos y a la realización de gráficos; casi todos los participantes declaran usar tidyverse, dplyr o ggplot2.

La encuesta incluye también tres preguntas de respuesta abierta acerca de las experiencias personales con el uso de R, los aspectos que más dificultad les trajeron y la forma habitual de resolver problemas o dudas. Si bien estas preguntas requieren de un análisis cualitativo que será completado para la presentación final de este trabajo, en general la idea que más surge es la dificultad de aplicar lo visto en los cursos sobre sus propios datos. Por otro lado, dentro de los aspectos que al día de hoy les generan dificultades aparecen algunos conceptos que suelen ser fundacionales en la experiencia de programación, como la creación de funciones específicas o bucles.

Al momento de la presentación de este resumen, continuamos recolectando respuestas; de todos modos, estos resultados nos permiten delinejar algunas conclusiones preliminares. En primer lugar, el hecho de que la mayoría de las personas que provienen de las Ciencias Sociales y Humanidades trabajen con análisis estadístico nos lleva a preguntarnos si las carreras de base efectivamente proveen herramientas para este trabajo. En segundo lugar, y relacionado con lo anterior, la observación de una fuerte presencia de aprendizaje autodidacta plantea la posibilidad de que los cursos y capacitaciones no sean un espacio amigable para personas que provengan de estas áreas o que no se dicten temáticas específicas para el trabajo en estas disciplinas. Es posible que el tipo de trabajo de las Ciencias Sociales y Humanidades tenga un grado de especificidad que requiera otro acercamiento al provisto por la mayoría de los cursos de enseñanza de R. Al mismo tiempo, nos preguntamos si la formación en estas carreras no provee también a las personas de una serie de fortalezas sobre la naturaleza sintáctica y la experiencia iterativa de la lógica de la programación que quizás es desaprovechada por las capacitaciones generales. Finalmente, el hecho de que muchas personas que trabajan cotidianamente con R no hayan incorporado ciertos aspectos básicos de la programación sugiere que quizás el uso de R en estos contextos laborales esté tan asociado al análisis estadístico que no siempre deje lugar al aprendizaje de R como lenguaje de programación. Esto nos lleva a preguntarnos si quizás las personas que provienen de las Ciencias Sociales y Humanidades no se beneficiarían de capacitaciones específicas sobre la lógica de la programación.

Referencias

- Arnold, T., & Tilton, L. (2015). *Humanities Data in R: Exploring Networks, Geospatial Data, Images, and Text*. Springer.

- Hernández Sampieri, R., Fernández-Collado, C., & Baptista Lucio, P. (2014). *Metodología De La Investigación 6ta edición*. McGraw-Hill.
- Jordan, K. L., Corvellec, M., Wickes, E. D., Zimmerman, N. B., Duckles, J. M., & Teal, T. K. (2018). Short-format workshops build skills and confidence for researchers to work with data. En *2018 ASEE Annual Conference & Exposition*.
- Sirkin, M. (2006). *Statistics for the Social Sciences* (3rd edition). Sage Publications.
- Vance, E. A., G limp, D. R., Pieplow, N. D., Garrity, J. M., & Melbourne, B. A. (2022). Integrating the Humanities into Data Science Education: Reimagining the Introductory Data Science Course. *Statistics Education Research Journal*, 21(2).