

CARRERA: Licenciatura en Sociología

ASIGNATURA: Análisis de datos 1

DOCENTE (s) A CARGO: Lic. Betsy Cohen; Dr. Gaston Becerra

AÑO: 2024

I. FINALIDAD Y/U OBJETIVOS

La asignatura de *Análisis de Datos I* tiene como propósito introducir a lxs estudiantes a el análisis de datos en la cual la técnica se ve atravesada por discusiones epistemológicas y éticas.

Se realizará un recorrido por algunas de las herramientas y técnicas básicas de análisis de datos mediante el lenguaje R. Por último, se espera que la integración de estas herramientas y perspectivas permita comprender al campo de las ciencias sociales computacionales como un campo de estudio transdisciplinario.

Objetivos generales:

- a) Que lxs estudiantes se acerquen al fenómeno del big data y la “ciencia de datos” como manifestaciones tecnológicas, sociales y culturales, y puedan reconocer sus discusiones epistemológicas y éticas más relevantes;
- b) Que lxs estudiantes se introduzcan a la metodología de trabajo del análisis de datos y se familiaricen con algunas herramientas y técnicas en lenguaje R;
- c) Que lxs estudiantes aprendan a construir gráficos, en diferentes formatos como insumos para la comunicación en investigación;
- d) Que lxs estudiantes se acerquen a las ciencias sociales computacionales como un campo de estudio transdisciplinario.

II. CONTENIDOS

Unidad I: El análisis de datos y la Sociología.

Emergencia de las “ciencias de datos”. La sociología frente a las ciencias de datos y la “crisis metodológica”. La sociología y los datos. La mitología del data-mining y sus pretensiones epistemológicas.

Unidad 2: Introducción al análisis de datos con R y RStudio

- Por qué R. R y su comunidad. La sociología y R.
- Entornos, scripts y notebooks. Packages y librerías. Tipos de objetos en R. Convenciones y símbolos de R.
- El flujo de trabajo en un proyecto de análisis de datos. Importación de datos.
- ETL como proceso “iterativo” de conocimiento.

Unidad 3: Ordenar manipular, explorar volver a arrancar: introducción a ETL en acción y la filosofía “tidy”

Extraer, cargar, transformar

- Filosofía “tidy” para estructurar y manipular datos. Tidyverse.
- Transformación de datos (dplyr y sus verbos),
- uniones (joins) y pivoteos.
- La función across()

Análisis exploratorio

- Tipos de datos
- Análisis exploratorio y descriptivos para datos numéricos y datos categóricos.
- Uso de if_else() y case_when().
- Limpieza y normalización de datos.
- Tratamiento de casos perdidos.
- janitor() como complemento del universo tidyverse.
- "Empoderando" los verbos de dplyr: la función map() de la librería purrr

Unidad 4: Comunicar: gráficos y diferentes salidas de reportes con Quarto.

- Gráficos: Nociones generales sobre tipos de gráficos y sus usos. Introducción a ggplot2 y su gramática de gráficos. Esquise() como guía inicial para entender la gramática

- Tablas: Gramática de tablas y tablas con GT y DT::datatable()
- Mapas: tipos de mapas. Creación de mapas simples con leaflet
- RMarkdown y Quarto
 - Reportes y Presentaciones
 - Tableros
 - Blogs y Páginas web

Unidad 5: Recursos y experiencias de la comunidad.

Recorrido por paquetes de referencia de la comunidad: {EPH}, {polAr}, {datos}.
Proyecto rOpenSci. La comunidad de ciencia de datos en Argentina.

III. ACTIVIDADES y MODALIDAD

Durante el dictado semanal se combinarán clases sincrónicas y actividades asincrónicas. Las clases asincrónicas constarán de exposiciones sintéticas sobre los contenidos de cada unidad a cargo del docente donde se plantearán debates teóricos o se realizarán demostraciones de distintas herramientas prácticas. Dichas clases estarán acompañadas del seguimiento de actividades prácticas asincrónicas que incluirán la participación en foros, la elaboración de consignas breves a partir de materiales específicos y/o guías de lectura. A su vez, estas dos modalidades asincrónicas estarán acompañadas por clases sincrónicas donde se plantearán actividades de resolución de problemas, donde se resolverán dudas y donde se procurará conocer las trayectorias y momentos del aprendizaje de los estudiantes.

IV. EVALUACION Y PROMOCIÓN

La materia tendrá al menos 1 (un) examen, en el cual se espera que los estudiantes den cuenta de la comprensión de los principales debates introducidos en la asignatura respecto de la compleja relación entre las ciencias sociales y análisis de datos. A su vez, la materia también contempla la entrega de un proyecto de investigación final (grupal) que se

entregará al final de la cursada y proporcionará una segunda calificación, integrando todo el contenido visto en la cursada.

La materia contempla la posibilidad de promoción directa. Para ello, se requiere un 75% de asistencia a las clases o participación en actividades asincrónicas, y la obtención de un mínimo de 7 (siete) puntos en todas las instancias de evaluación.

Para poder rendir el examen final se deberá haber obtenido un mínimo de 4 (cuatro) puntos en todas las instancias de evaluación. La materia prevee una instancia de recuperación para el primer examen en caso de aplazo. La regularidad de la materia dura 2 años.

V. RECURSOS

Para las actividades sincrónicas se requiere una sala de reunión virtual que permita grabar videos y utilizar herramientas de chat y participación. Para las actividades asincrónicas se requerirá un aula virtual que permita el armado de foros, consignas, alojar materiales y recursos.

VI. BIBLIOGRAFIA OBLIGATORIA y BIBLIOGRAFÍA COMPLEMENTARIA

Unidad 1:

Bibliografía obligatoria:

Becerra, G. (2021). Los desafíos del big data para la formación sociológica. II Jornadas Institucionales de Innovación Educativa En La Universidad.

Meneses Rocha, M. E. (2018). Grandes datos, grandes desafíos para las ciencias sociales. Revista Mexicana de Sociología, 80(2), 415–444.

<https://doi.org/10.22201/iis.01882503p.2018.2.57723>

Kitchin, R. (2014). The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. Sage. <https://doi.org/10.4135/9781473909472>

Bibliografía complementaria:

boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662–679.

<https://doi.org/https://doi.org/10.1080/1369118X.2012.678878>

Savage, M., & Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology*, 41(5), 885–899. <https://doi.org/10.1177/0038038507080443>

Unidad 2:

Bibliografía obligatoria:

Becerra, G., & López Alurralde, J. P. (2020a). Analisis de datos con R. En *Introducción a la ciencia de datos con R: Una guía rápida para científicos sociales*.

https://bookdown.org/gaston_becerra/curso-intro-r/analisis-de-datos-con-r.html

Wickham, H., & Grolemund, G. (2019). Vectores. En *R para ciencia de datos*.

<https://es.r4ds.hadley.nz/>

Wickham, H., & Grolemund, G. (2019). Datos ordenados. En *R para ciencia de datos*.

<https://es.r4ds.hadley.nz/>

Bibliografía complementaria:

Wickham, H., & Grolemund, G. (2019). *R para ciencia de datos*.

<https://es.r4ds.hadley.nz/>

Grolemund, G. (2014). *Hands-on programming with R: Write your own functions and simulations*. " O'Reilly Media, Inc.". <https://rstudio-education.github.io/hopr/>

Unidad 3:

Bibliografía obligatoria:

Becerra, G., & López Alurralde, J. P. (2020). Entorno. En *Introducción a la ciencia de datos con R: Una guía rápida para científicos sociales*.

https://bookdown.org/gaston_becerra/curso-intro-r/entorno.html

Wickham, H., & Grolemund, G. (2019). Transformación de datos. En *R para ciencia de datos*. <https://es.r4ds.hadley.nz/>

Wickham, H., & Grolemund, G. (2019). Iteración. En *R para ciencia de datos*.

<https://es.r4ds.hadley.nz/>

Unidad 4:

Bibliografía obligatoria:

Wickham, H., & Grolemund, G. (2019). Visualización de datos. En *R para ciencia de datos*. <https://es.r4ds.hadley.nz/>

Cuevas-Fernandez, E. Introducción a Shiny. En Workshop CDSB 2021: Flujos de trabajo con RStudio y creación de Shiny apps.

https://comunidadbioinfo.github.io/cdsb2021_workflows/introducción-a-shiny.html

Bibliografía complementaria:

Wickham, H. (2021). *Mastering shiny*. O'Reilly Media, Inc. <https://mastering-shiny.org>

Unidad 5:

Bibliografía obligatoria:

Becerra, G., & López Alurralde, J. P. (2020). Procesamiento de lenguaje natural y análisis de sentimiento. En *Introducción a la ciencia de datos con R: Una guía rápida para científicos sociales*. https://bookdown.org/gaston_becerra/curso-intro-r/entorno.html