

Análisis Social con R

29 de enero al 2 de febrero

9:30 a 13:30h

Rodrigo Rodrigues-Silveira

Una breve inspección a los diferentes manuales de análisis de datos y de introducción al R revela que, en su mayor parte, estos guías se concentran en presentar conceptos estadísticos que van desde medidas descriptivas hasta modelos de regresión más o menos complejos. Este tipo de acercamiento al análisis de datos considera que el usuario o lector: (a) necesita un conocimiento básico de estadística antes de poder llevar a cabo cualquier análisis; (b) los programas son medios de implantar conceptos y técnicas estadísticas que pueden ser utilizadas para el análisis de datos y que ya son conocidos por los estudiantes; (c) no hace falta orientación adicional sobre qué técnicas utilizar en qué situación y objeto de estudio; y (d) los lectores/usuarios/estudiantes saben qué son datos, qué tipos de datos existen y qué técnicas deben ser aplicadas para su análisis.

Este enfoque técnico, y muy poco práctico, es quizás el principal obstáculo para un aprendizaje adecuado del análisis de datos. Su enfoque no se encuentra en el análisis de los datos en sí mismos, sino en enseñar técnicas estadísticas “de manual” y cómo implantarlas en algún paquete estadístico como el SPSS, el Stata o el R. Aún en los casos en los que se utiliza un problema como guía para la enseñanza, este problema se elige para ilustrar un concepto o técnica estadística. De ese modo, lo que falta es un acercamiento que ponga énfasis en el análisis de datos propiamente dicho y que utilice las técnicas como instrumentos que pueden ayudar en ciertos casos (y molestar en otros).

El presente curso pretende ofrecer un enfoque alternativo al tradicional, haciendo hincapié en los conocimientos necesarios para llevar a cabo un análisis de datos efectivo que, en términos concretos, significa: (a) el desarrollo de una inquietud intelectual genuina; (b) la formulación de cuestiones e hipótesis de trabajo; (c) la elección de métodos de análisis adecuados para el problema a ser analizado; (d) el ejercicio de creatividad para el desarrollo de nuevos métodos o técnicas; (e) la elección de los instrumentos adecuados para el tipo de dato trabajado; y (f) conocer los límites y las potencialidades de cada técnicas de análisis adoptada.

Se espera que los estudiantes adquieran las habilidades para utilizar datos, seleccionar las técnicas de análisis y visualización adecuadas. Sin embargo, resulta aún más importante que sean capaces de desarrollar *insights*, preguntas de investigación e hipótesis originales, que sirvan de forma práctica en su desarrollo académico y profesional posterior. Como competencias específicas se espera que sean capaces de: (a) buscar datos para sus investigaciones; (b) tratarlos y limpiarlos para evitar problemas e interferencias; (c) realizar distintos análisis con el objetivo de generar preguntas de investigación e hipótesis; (d) utilizar estadísticas, gráficos y otros recursos analíticos para explorar los datos y comunicar resultados obtenidos; (e) definir y testar relaciones causales.

Este curso, de 20 horas de duración, está dividido en cinco temas de cuatro horas cada uno. Sin embargo, para un aprendizaje adecuado, se recomienda que los estudiantes realicen las lecturas recomendadas en la bibliografía para cada tema y lleven a cabo los ejercicios propuestos. Estas actividades ayudarán a que consoliden los conocimientos adquiridos en clase y facilitarán un aprendizaje más denso y completo de los contenidos y propuestas cubiertos en este curso.

Local: aula de informática 2 de la Facultad de Derecho (Planta Jardín).

TEMARIO

Tema 1. Analizar datos, pero ¿qué son datos? y ¿cómo el R puede ayudar?

- Qué son datos
- Tipos de datos (nominales, ordinales, intervalo, continuos)
- Recolección de datos (datos primarios, secundarios, de Internet)
- Formas de organización de datos (tabulares, árboles, redes)
- El R: una plataforma de análisis
- Abriendo una plantilla de Excel en el R

Tema 2. Antecedentes: preparación y transformación de datos

- La perfección cuesta trabajo: limpiando la basura y corrigiendo errores
- Entendiendo la estructura de los datos
- Filtrar, reordenar y reestructurar bases de datos
- Calculando indicadores sociales: transformar y rescalonar variables

Tema 3. Generando *insights* por el análisis de patrones y anomalías

- *Insights* a partir de patrones y anomalías en los datos
- Haciendo preguntas y formulando hipótesis a partir de los datos
- El Análisis Exploratorio de Datos: receta vs. saber práctico
- Clasificando casos semejantes: *clusters* y componentes principales
- Buscando respuestas a las preguntas: asociación y teste de hipótesis

Tema 4. Visualización: entre explorar, presentar e interaccionar

- Explorar y Comunicar: la doble finalidad de la visualización de datos
- Eligiendo el tipo de gráfico adecuado a la información deseada
- Explorando patrones y anomalías utilizando gráficos
- Comunicación visual de los resultados de investigación
- Gráficos interactivos

Tema 5. Explicando los resultados: la inferencia

- Adivinando el futuro y explicando el pasado: la predicción y la inferencia
- Por qué asociación no es causalidad
- Diferentes formas de llevar a cabo la inferencia causal
- Simplificar y generalizar: los modelos de regresión
- Modelos: cuándo utilizarlos y cuál es el más adecuado

BIBLIOGRAFÍA

Lecturas recomendadas

Tema 1.

Kabacoff, Robert I. 2011. "Introduction to R". En: *R in Action: Data Analysis and Graphics with R*. New York: Manning Publications Co. pp. 3-20.

- Munzner, Tamara. 2014. “What: Data Abstraction” y “Why: Task Abstraction”. En: *Visualization Analysis and Design*. Har/Psc edition. Boca Raton: A K Peters/CRC Press. pp.20-40 y 42-65.
- Yau, Nathan. 2013. “Understanding data”. *Data Points: Visualization That Means Something*. Edición: 1. Indianapolis, IN: Wiley John + Sons. Pp. 1-41.

Tema 2.

- Kabacoff, Robert I. 2011. “Creating a Dataset”, “Basic Data Management”, “Advanced Data Management”, “Advanced Methods for Missing Data”. En: *R in Action: Data Analysis and Graphics with R*. New York: Manning Publications Co. pp. 21-45; 73-90; 91-116; 352-371.

Tema 3.

- Kabacoff, Robert I. 2011. “Basic Statistics”, “Analysis of Variance”, “Principal Component and Factor Analysis”. En: *R in Action: Data Analysis and Graphics with R*. New York: Manning Publications Co. pp. 141-170; 219-245; 331-350.

Tema 4.

- Yau, Nathan. 2013. “Exploring Data Visually” y “Designing for an Audience”. *Data Points: Visualization That Means Something*. Edición: 1. Indianapolis, IN: Wiley John + Sons. Pp. 135-199 y 235-275.
- Tufte, Edward R. 2001. “Graphical Excellence” y “Data-Ink and Graphical Redesign”. En: *The Visual Display of Quantitative Information*. Cheshire: Graphic Press. Pp. 13-51 y 91-105.
- Kabacoff, Robert I. 2011. “Getting started with graphs”, “Basic graphs”, “Intermediate Graphs”, “Advances Graphs”. En: *R in Action: Data Analysis and Graphics with R*. New York: Manning Publications Co. pp. 45-71; 119-140; 263-290; 373-399.

Tema 5.

- Kabacoff, Robert I. 2011. “Regression”, “Generalized Linear Models”. En: *R in Action: Data Analysis and Graphics with R*. New York: Manning Publications Co. pp. 171-218; 313-330.

Manuales de estadística

- Levine, David, and David Stephan. 2009. *Even You Can Learn Statistics: A Guide for Everyone Who Has Ever Been Afraid of Statistics*. 2 edition. Upper Saddle River, N.J: FT Press.