

Datos y Derecho

Herramientas computacionales para la investigación empírica

Objetivo

El objetivo del curso de “Datos y Derecho. Herramientas computacionales para la investigación empírica” es introducir las herramientas de la ciencia de datos en ámbitos jurídicos, para facilitar la realización de análisis empíricos cuantitativos. La cada vez mayor existencia de bases de datos y la mayor disponibilidad de grandes cantidades de datos (*Big Data*), nos impone la necesidad de aprender a manejar las herramientas de procesamiento adecuada de acuerdo a nuestros objetivos para obtener la mayor información posible sobre la vigencia o no de los derechos.

Las ciencias jurídicas estuvieron en general desconectadas de las investigaciones empíricas cuantitativa, y del uso de datos para hacer análisis sobre políticas públicas o para generar evidencia para las discusiones jurídicas. El uso de computadoras para procesar datos, como herramienta de análisis, también estuvo relegada.

Es por ello que en este curso vamos a aprender a usar las herramientas de la Ciencia de Datos para la elaboración de investigaciones en materia de derechos. Usaremos el lenguaje “R” (<https://www.r-project.org/>) para descargar, ordenar, analizar, y visualizar datos. ¿Qué es R? es un lenguaje de programación para el manejo de bases de datos y la realización de análisis estadísticos de grandes bases de datos, de fuente abierta, muy popular en ámbitos académicos. Para ello, deberemos tener instalado el programa, junto con el entorno de trabajo RStudio, y aprender a usarlo. En el curso también aprenderemos algunas nociones sobre cómo hacer presentaciones e informes con el mismo lenguaje.

El curso está orientado a capacitar a investigadores, con el objetivo de brindar herramientas para el análisis de datos.

En esta edición, utilizaremos bases de datos, casos y ejemplos de derechos de las personas migrantes, para analizar y compararlas con el total de la población, o para diferenciar su goce de derechos.

Programa

Este curso constará de cinco clases prácticas. En estas clases les alumnos podrán adquirir una noción sobre qué pueden hacer con R, y las primeras nociones sobre cómo hacerlo. Una vez adquiridas estas nociones, cada alumne podrá avanzar en la realización de un análisis propio, con datos que sean de su interés, a modo de trabajo final.

Clase 1

¿Qué es la Ciencia de Datos? ¿Cuáles son los desafíos que nos impone la aparición de grandes cantidades de datos? Sobrevigilancia y control. ¿Cómo podemos aprovechar estos datos?

¿Qué es R? ¿Para qué sirve? ¿Por qué puede ser importante esto?

Instalación de R y Rstudio, y paquete Tidyverse

Primer proyecto con datos de mortalidad de la Ciudad de Buenos Aires por comuna.

Bibliografía:

- Capítulos 1 y 2 del libro de Antonio Vazquez Brust (AVB)
- Capítulo 1 del libro de Headley Wickham (HW)

Clase 2

Manipulación de datos. Verbos del Tidyverse: filter, mutate, select, arrange, group by, summarize.

Segundo proyecto con datos de la Encuesta Anual de Hogares, variable de país de nacimiento.

- Capítulo 3 del libro de AVB
- Capítulo 5 del libro de HW

Clase 3

Finalización de los proyectos anteriores. Introducción a las visualizaciones. GGplot. Gramática de los gráficos por capas. Graficos interactivos. Gapminder

- Capítulo 4 del libro de AVB
- Capítulo 3 del libro de HW

Clase 4

Ggplot y geolocalización. Paquete sf. Shapefiles abiertos.

- Capítulo 6 del libro de AVB
- primeros capítulos del libro *Geocomputation with R*
<https://geocompr.robinlovelace.net/>

Clase 5

Informes y presentaciones. Rmarkdown. Presentacion de paquetes Blogdown, Xaringan, Pagedown. Cómo hacer presentaciones actualizables.

- Capitulo 27, 28 y 29 del libro de Wickham.
- primeros capítulos del libro Rmarkdown the definitive guide
<https://bookdown.org/yihui/rmarkdown/>

Bibliografía

Si bien es un curso práctico, vamos a guianos principalmente en dos libros. El primero más operativo será el de Antonio Vázquez Brust, *Ciencia de Datos para Gente Sociable* disponible en el siguiente link

https://bitsandbricks.github.io/ciencia_de_datos_gente_sociable/

Y asimismo vamos a ir a las fuentes principales de Hadley Wickham (el *padre fundador* del paquete Tidyverse, y un referente de RStudio) que escribió este *R for Data Science*

<https://r4ds.had.co.nz/>.

Este último está en inglés, y recomiendo leerlo allí, siempre que lean inglés. La comunidad R en Latinoamérica comenzó un proyecto de traducción colectiva del libro de Wickham, disponible (al menos la primera parte) aquí: <https://es.r4ds.hadley.nz/>