

Doctorado en Sociología
Escuela Interdisciplinaria de Altos Estudios Sociales - Universidad
Nacional de San Martín

Seminario: Visualización de Datos

Docente: Germán Rosati

Fechas: 3, 17 y 31 de mayo y 14 y 28 de junio

Justificación

Suele pensarse que el proceso de generación de información visualizable en una investigación es un aspecto meramente estético (o sea, secundario) y delegable en otras disciplinas. Sin embargo, la construcción de visualizaciones efectivas ayuda a explorar y comprender diversos aspectos de la estructura subyacente en los datos.

A su vez, son muy relevantes en la comunicación de resultados tanto a audiencias técnicas como no técnicas. La comunicación pública de la ciencia se beneficia de visualizaciones e infografías diseñadas de forma clara. Un buen despliegue visual de la información ayuda a transmitir conceptos complejos.

Este curso corto de doctorado es una introducción práctica a la visualización de datos. Se enfoca en el análisis y presentación de datos reales. A su vez, se cubren algunos aspectos de la dimensiones estética y cognitiva de la percepción gráfica. Se abordará el flujo de trabajo en el proceso de generación de visualizaciones y se trabajarán diferentes tipos de datos: categóricos, cuantitativos, georreferenciados y algunas herramientas para generar visualizaciones interpretables de modelos de regresión.

Objetivos

El seminarios se propone que les asistentes

1. desarrollen e incorporen algunos criterios operativos e intuiciones sobre la construcción de visualizaciones;
2. puedan identificar las características de una visualización útil;
3. adquieran algunos fundamentos sobre la producción de visualización de datos sólida y efectiva utilizando el lenguaje R.

Vale aclarar que no se trata de un curso de capacitación en R, sino que el lenguaje será utilizado como herramienta para generar visualizaciones de forma replicable.

Contenidos y cronograma

- Clase 1. Explorando los datos. Introducción a R y al tidyverse. Nociones conceptuales: forma, color, tamaño, color. Sobre la percepción y su relación con la visualización. ¿Cómo reconocer una mala visualización?
- Clase 2. Traduciendo los datos a elementos visuales. Gramática de gráficos. Nociones de datos ordenados (tidy data). Introducción a ggplot. Geometrías, estéticas, capas y facetados.
- Clase 3. Puliendo los gráficos. Leyendas, líneas de tendencia, modificando ejes.
- Clase 4. Mapeando datos. Nociones sobre cartografía (proyecciones, sistemas de coordenadas). Coropletas, cartogramas y afines. ¿Cuántas variables pueden representarse en un mapa?
- Clase 5. Visualizando modelos. Repaso de regresión lineal y logística. ¿Cómo salir de las tablas de coeficientes? La post-estimación como herramienta para visualizar efectos. Introducción a broom.

Software

Todas el trabajo en clase se hará utilizando [el lenguaje R](#). Se usará [RStudio](#) para administrar el código y los proyectos.

Para instalar R:

- [Instalar la versión más reciente de R](#). R es libre para sistemas operativos basados en Windows, Mac y Linux. Descargar la versión de R compatible con su sistema operativo.
- Una vez que R está instalado, [descargar e instalar RStudio](#). RStudio es una interfaz gráfica (o IDE) que ayuda a administrar el código, los datos y los proyectos. Es decir, hace a R mucho más fácil de usar. RStudio también es gratis y está disponible para plataformas Windows, Mac y Linux.
- Instalar las librerías a utilizar en el curso (*tidyverse* y otras). Estas librerías proveen funciones sumamente útiles que nos servirán a lo largo del curso. Para instalar *tidyverse*, deberán asegurarse de estar conectados a internet y abrir RStudio. Luego, en la línea de comando de RStudio (en el panel llamado "Console") tipear las siguientes líneas y presionar Enter:

```
my_packages <- c("tidyverse", "broom", "cowplot", "drat",  
                "gapminder", "GGally", "ggforce", "ggrepel", "ggribbles",  
                "gridExtra", "here", "interplot", "margins", "maps",  
                "mapproj", "mapdata", "MASS", "quantreg", "rlang",  
                "scales", "survey", "srvyr", "devtools")  
  
install.packages(my_packages, repos = "http://cran.rstudio.com")
```

RStudio comenzará a bajar e instalar los paquetes. Esto podría tardar un rato.

Dinámica de las clases

Las clases constan de dos partes:

- exposición teórica a cargo del docente
- práctica guiada realizada en conjunto

Al finalizar la clase, se asignan consignas de ejercicios que deberán ser resueltas para la clase siguiente.

Condiciones de aprobación y modalidad de evaluación

- 80% de asistencia a las clases (en modalidad virtual sincrónica)
- Exposición oral grupal de la resolución de al menos una guía de consignas de ejercicios
- Entrega y aprobación de un trabajo final monográfico

Bibliografía

- Healy, Kieran (2019). *Data Visualization: A Practical Introduction*, Princeton: Princeton University Press.
 - Versión online borrador gratis <http://socviz.co/>
- Wickham, Hadley y Golemund, Garret (2017). *R for Data Science Import, Tidy, Transform, Visualize, and Model Data*, California: O'Reilly Media.
 - Versión online gratis en castellano: <https://es.r4ds.hadley.nz/>
 - Versión online gratis en inglés: <http://r4ds.had.co.nz/>
- Tufte, Edward (1983). *The Visual Display of Quantitative Information*, Cheshire, CT: Graphics Press.
- Vázquez Brust, Antonio (2020). *Ciencia de datos para gente sociable*, <https://bitsandbricks.github.io/ciencia-de-datos-gente-sociable/>