

Programa de Metodología Aplicada a la Gestión Pública

Política Pública con Big Data

Prof. @Ernesto Calvo
Universidad de Maryland
ecalvo@umd.edu

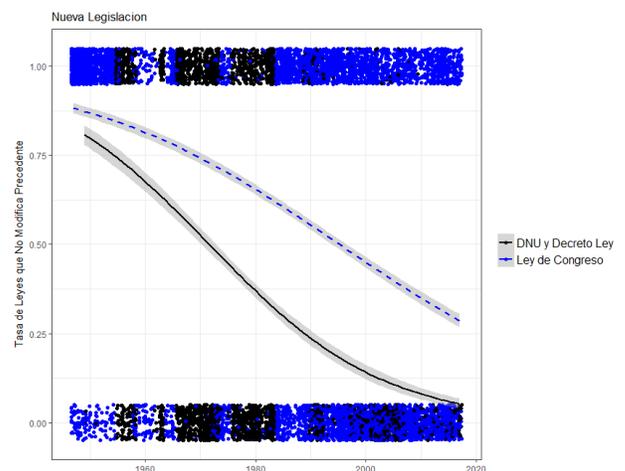
Universidad Nacional de San Martín
Agosto 5 a agosto 11, 2021

Este seminario tiene como objetivo introducir a sus participantes al uso de grandes datos para procesamiento de grandes bases de datos en R/R-Studio; el uso de ggplot, stargazer y rvest; el procesamiento de filas JSON; el uso de APIs de organizaciones como Facebook, Twitter y el NYT. El seminario está estructurado en 12 módulos, concluyendo con un hackaton de datos y la presentación de resultados. El curso será dictado integralmente en R/R-Studio.

Objetivos del Curso:

Los **@participantes** adquirirán las herramientas para:

1. Entender los distintos tipos de bases disponibles para el análisis de la política pública (Big Data).
2. Entender el tipo de patologías presentes en grandes datos cuando son re-utilizados para la investigación académica o para entender la política pública.
3. Entender las herramientas básicas del análisis estadístico de grandes bases de datos en R/R-Studio. Listados, arrays, vectores, matrices, modelos estadísticos lineales (lm) y modelos lineales generalizados (glm);



4. Graficar resultados en ggplot y producir tablas de estimados con stargazer.

5. Finalmente, los participantes realizarán un diseño de investigación para la utilización de una base de datos relevante de grandes datos.

Textos de apoyo:

- Healy, K. (2018). *Data visualization: a practical introduction*. Princeton University Press.
- Salganik, M. J. (2019). *Bit by bit: Social research in the digital age*. Princeton University Press.
- Eleccion Argentina 2017: Documento, [Eleccion PASO 2017](#).
- Eleccion Bolivia 2019: Documento Oficial: [Bolivia-OEA](#), Revision: [Kronick et al.](#)
- [Alcañiz, I., Calvo, E., & Rubio, J. \(2016\). Educadamente desiguales: género y salario en el sector público argentino \(2003-2010\). *Desarrollo económico*, 55\(217\), 343-357.](#)

Software:

En el seminario utilizaremos R/R-Studio (cualquiera de las dos opciones).

- Para bajar R-Studio: <https://www.rstudio.com/>
- Para bajar R: <https://cran.r-project.org/>
- Para instalar los paquetes requeridos para la clase correr en R o R-Studio el código:

```
sapply(c("rvest","httr","knitr","tm","igraph","stargazer"),install.packages)
```

Programa

Módulo 1, Día I: Introducción a bases de datos en el sector publico

- ¿Dónde hay un dato? Datos, meta-datos, data estructurada y no estructurada. Algunos ejemplos de datos compilados por el Estado y sobre sus posibles usos (EPH, InfoLeg, Datos Legislativos, Corte Suprema).

Módulo 2, Día I: Que podemos hacer con los datos públicos y como se generan nuevos datos desde el estado.

- Descripción, modelización, identificación. ¿Cómo se vincula el trabajo académico en Big Data y las necesidades del sector público? *Salganik (2019) Capitulo 2.*

Módulo 3, Día II: Introducción a R-Studio, 1½ horas

- Una introducción a R/R-Studio. Tipos de datos, vectores, matrices, arrays, listas, loops, apply, tapply/sapply. Apoyo: *Healy (2018), Capítulo 2.*

Módulo 4, Día II: Tidyverse, 1½ horas

- Manejo de datos. Un ejemplo usando los datos de InfoLeg. *Healy (2018), Capítulo 2.*

Módulo 5, Día III: Trabajando con Modelos en R, 1 ½ horas

- LM y GLM. Estimados y gráficos. Una revisión de modelos estadísticos utilizando R. Algunas reglas para trabajar con Big Datasets. *Healy (2018), Capítulo 6.*

Módulo 6, Día III: Un ejemplo de Modelización, 1 ½ horas

- Salarios, entradas y salidas del sector público en la Argentina.

[Alcañiz, L., Calvo, E., & Rubio, J. \(2016\). Educadamente desiguales: género y salario en el sector público argentino \(2003-2010\). *Desarrollo económico*, 55\(217\), 343-357.](#)

Módulo 7, Día IV: Usando ggplot para explorar nuestras variables dependientes y nuestros covariados.

- La semántica de gráficos de ggplot. Un modelo, un gráfico, una idea. *Healy (2018), Capítulo 1.*

Módulo 8, Día IV: ggplot y COVID-19

- Visualizar y modelar la evolución de la pandemia. <https://kjhealy.github.io/covdata/>

Módulo 9, Día V: Autopsia Electoral, 1 ½ horas.

- Big Data aplicado a la evaluación electoral:

Argentina: [Eleccion PASO 2017](#),

Bolivia: Documento Oficial: [Bolivia-OEA](#), Revision: [Kronick et al.](#)



Escuela de
Política y Gobierno
EPyG_UNSAM

Módulo 10, Cierre: Donde formarse y para que, 1 ½ horas.

- Recursos humanos para tomar ventaja de los recursos en el estado.