A blue and yellow logo

Description automatically generated

M73 13 Taller de Programación[[1]](#footnote-1)

Primavera 2025

**María Noelia Romero**

***Profesora***

[25RO35480961@campus.economicas.uba.ar](mailto:25RO35480961@campus.economicas.uba.ar)

El **objetivo** de este curso es presentar las herramientas estadísticas y computacionales más utilizadas para hacer análisis descriptivos, predicciones y clasificaciones. El curso presenta casos aplicados de cada herramienta en el ámbito de investigación en ciencias sociales y políticas públicas. Mediante el entendimiento conceptual y práctico, los estudiantes desarrollan un pensamiento crítico de las oportunidades y limitaciones de cada herramienta computacional y descubren distintas bases de datos disponibles.

El curso desafía a los estudiantes a trabajar en equipo y:

1. Programar, manejar y explorar distintas bases de datos secundarias,
2. Realizar una presentación oral breve sobre una base de microdatos con una idea de exploración de datos,
3. Mostrar la aplicación una herramienta de visualización vistas en clase.

En resumen, estos objetivos del curso apuntan a que el alumno desarrolle gran versatilidad para comprender, utilizar y presentar datos y visualizaciones según la demanda en su futuro desarrollo profesional (sea académico o no académico).

**Clases**

El curso se desarrolla con una clase teórica y una tutorial a la semana.

|  |
| --- |
| Clase Online |
| Miercoles 19:00 a 22:00 h  Teams |

Mientras que la clase magistral suele tener un contenido más teórico y de aplicación económica; la clase tutorial guía a los estudiantes en el uso de cada herramienta en Python. Las tutoriales estarán disponibles en el Campus Virtual y en un repositorio de GitHub. En el campus, también compartiremos información relevante del curso y es un medio frecuente de comunicación (además de las clases, horas de consultas y emails). Por lo que se recomienda estar atentos a los anuncios en el campus.

*Asistencia*: como es práctica de UBA, se requiere asistir como mínimo al 75% de las clases.

**Contenidos:** El curso tiene fuerte carácter técnico, computacional, y de pensamiento crítico. El estudiante está motivado por el uso de datos, ya sea con recopilación de datos primarios o el cuestionamiento de usos de datos secundarios. Además, el curso es ideal para estudiantes con inclinación por sintetizar ideas complejas de manera sencilla para el público en general.

**Modalidad de trabajo:**

Este curso enfatiza el trabajo en equipo (de 3 integrantes en total) y espera una activa disciplina individual de estudio del material con pensamiento crítico.

**Mecanismo de evaluación:** La aprobación del curso se basa en las siguientes evaluaciones de **trabajos grupales.** El curso incentiva la colaboración en grupo de **tres** personas (a determinar la primera semana de clase).[[2]](#footnote-2) La nota del curso consiste en las siguientes partes e instancias:

* Cuatro **trabajos prácticos** (, 30% de la nota);
* **Presentación grupal** en clase (, 20% de la nota);
* Un **borrador** de ideas usando la base de datos introducidas en la presentación grupal (, 20% de la nota);
* Un **presentación grupal final** con una aplicación donde se destaque una herramienta de visualizaciónen la última de clase de manera presencial (, 30% de la nota);

Con los **trabajos prácticos** se introduce y vuelve al alumno en un experto en la base de microdatos del INDEC: Encuesta Permanente de Hogares ([link a EPH](https://www.indec.gob.ar/indec/web/Institucional-Indec-BasesDeDatos)). Estos trabajos requieren programación (entregar código de resolución de las consignas) y un reporte que interprete los resultados y discuta las limitaciones. Es requisito entregar y aprobar todos los trabajos prácticos.

La **presentación grupal** es breve de **10 minutos** en clase**.[[3]](#footnote-3)** El grupo debe encontrar una base de microdatos disponible y contar cómo piensa usar dicha base en las instancias del *borrador* y *poster*. El grupo debe explicar detalladamente la institución a cargo de la recolección de los datos y estructura general de los mismos. Debe contar brevemente el proceso de limpieza de datos y las variables de interés para ser analizadas en las siguientes etapas. La presentación debe concluir con ideas que les gustaría investigar. En la primera clase tutorial, se proveerán más instrucciones de las expectativas y sugerencias para dicha actividad.

La **aplicación final** debemostrar una o dos (como máximo) herramientas de visualización de las vistas en clase. En la mitad de la cursada (ver cronograma), se debe entregar un **borrador** de **3 (tres) páginas máximo** con una o más ideas preliminares para el poster final. Es opcional mostrar una estadística descriptiva en la instancia borrador (1 tabla y 2 gráficos de 1 o dos variables relevantes). A mejor visualización, mayor puntaje obtendrá el grupo. La **aplicación final** será en la última semana de clases. Las consignas de formato y expectativa de la aplicación final y cada instancia se pueden encontrar en el campus virtual. Además, en las clases tutoriales y horas de consultas se discutirá cualquier duda respecto a la entrega.

En resumen, la nota del curso () se calcula:

En todas las instancias, se espera un lenguaje profesional y/o académico en cada ítem, donde importa el contenido y la visualización de la información.

# **ESTRUCTURA DEL PROGRAMA Y LECTURAS**

Los libros de referencia del curso son:

**†** James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in Python.* Springer Nature. Descarga [gratis](https://www.statlearning.com/)

**†** Sosa Escudero, W., 2021, *Big data*, 7a edición, Siglo XXI Editores Argentina, Buenos Aires. Disponible físicamente en biblioteca (no en pdf).

**†** Sosa Escudero, W., 2024, *Viajar al futuro (y volver a contarlo)*, 1a edición, Siglo XXI Editores Argentina, Buenos Aires. Disponible físicamente en biblioteca (no en pdf).

**\*** Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer, Berlin: Springer series in statistics.

La lista de lecturas del curso consta de capítulos del libro de referencia del curso y una serie de artículos académicos y no académicos, todos detallados a continuación. La lista es bastante completa, pero sólo una pequeña proporción de ella, los textos identificados con una “daga” (**†**) califican como **lecturas obligatorias**. Aquellas con un “asterísco” (\*) califican como **lecturas recomendadas** más discutidas en clase.

El curso se organiza en cuatro grandes partes. Tras una breve introducción a la materia, clarificando la terminología comúnmente usada en aprendizaje automático (*machine learning*); primero, resumimos debates generales y una introducción a una buena visualización de datos. En la segunda parte, aprendemos métodos no supervisados y no paramétricos utiles en el analisis exploratorio y descriptivo de los datos. En la tercera parte, iniciamos el análisis de predicción propiamente revisando los métodos supervisados clásicos de regresión y clasificación. Además, discutimos el problema de sobreajuste de dichos modelos, el error de predicción fuera de la muestra y las soluciones con los métodos de remuestreos. En la tercera parte del curso, cubrimos modelos de predicción más complejos como los métodos de regularización, modelos de selección de variables, modelos no lineales y métodos basados en árboles.

Es importante destacar que en la última parte del curso (después de la semana de parciales), el nivel de dificultad comienza a aumentar. Las secciones con material más complejo están señaladas con el siguiente ícono:



# **INTRODUCCIÓN A CIENCIA DE DATOS Y VISUALIZACIÓN [[4]](#footnote-4)**

|  |
| --- |
| **Tema 1: Ciencia de Datos, terminología y debate. Visualización de datos.** |

\*Anderson, C. (2008). The end of theory. *Wired magazine, 16*(7), 16-07.

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, 355(6324), 483- 485. <http://science.sciencemag.org/content/sci/355/6324/483.full.pdf>

**†** James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in Python.* **Chap. 1 & 2**

Lambert, K. J., & Fegley, T. (2023). Economic Calculation in Light of Advances in Big Data and Artificial Intelligence. *Journal of Economic Behavior and Organization*, *206*, 243–250. <https://doi.org/10.1016/j.jebo.2022.12.009>

**†** Sosa Escudero, W., 2021, *Big data*, 7a edición, Siglo XXI Editores, Buenos Aires **Cap 1, pag 23 a 33**

**†** Schwabish, J. A. (2014). An economist's guide to visualizing data*. Journal of Economic Perspectives*, 28(1), 209-234.

**\*** Sosa Escudero, W., 2024, *Viajar al futuro (y volver a contarlo)*, 1a edición, Siglo XXI Editores Argentina, Buenos Aires. **Cap 3 pag 68 a 74.**

Nickerson, D., & Rogers, T. (2014). “Political Campaigns and Big Data”, *Journal of Economic Perspectives*, vol. 28(2), pp. 51-74.

Sosa Escudero, W. (2014). Big data: otra vez arroz?, Diario Clarín, 6/4/2014.

Sosa Escudero, W. (2016). Al infinito y más allá: Funes, Borges y big data, Diario La Nacion, 12/6/2016.

Sosa Escudero, W., Anauati, V y Brau, W. (2022), Poverty and inequality studies with machine learning, en Matyas, L. y Chen, F., *Econometrics with Machine Learning*, Springer, New York

# **ANÁLISIS EXPLORATORIOS Y DESCRIPTIVOS: METODOS NO PARAMETRICOS Y NO SUPERVISADOS**

|  |
| --- |
| **Tema 2: Métodos No Paramétricos: Histogramas y distribuciones de kernels** |

\* Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer, Berlin: Springer series in statistics. **Chap. 6.2**

Chen, Yen-Chi (2024) “STAT 425: Introduction to Nonparametric Statistics”

Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *Quarterly Journal of Economics*, 134(3), 1405-1454.

Jales, H. (2018). Estimating the effects of the minimum wage in a developing country: A density discontinuity design approach. *Journal of Applied Econometrics*, 33(1), 29-51.

John DiNardo, Nicole M. Fortin and Thomas Lemieux, 1996, “Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach,” *Econometrica*, Volume 64, Number 5 (September), pp. 1001- 1044.

|  |
| --- |
| **Tema 3: Métodos no supervisados I: Análisis de Componentes Principales (PCA)** |

**†** James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in Python.* **Chap. 6.3, 12.1, 12.2,**

**\*** Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer, Berlin: Springer series in statistics. **Chap 3.5.1, 10.2.3**

Diamond, R. (2016). The determinants and welfare implications of US Workers’ diverging location choices by skill: 1980-2000. *American Economic Review*, *106*(3), 479–524.

**\*** Gasparini, L., Sosa-Escudero, W., Marchionni, M., & Olivieri, S. (2013). Multidimensional poverty in Latin America and the Caribbean: new evidence from the Gallup World Poll. *Journal of Economic Inequality*, 11, 195-214.

|  |
| --- |
| **Tema 4: Métodos no supervisados II: Clúster** |

**†** James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in Python.* **Chap 12.4**

**†** Sosa Escudero, W., 2021, *Big data*, 7a edición, Siglo XXI Editores, Buenos Aires **Cap 3, pag 69 a 76**

Anselin, L., & Williams, S. (2015). Digital neighborhoods. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, 1-24.

\*Caruso, G., Sosa‐Escudero, W., & Svarc, M. (2015). Deprivation and the dimensionality of welfare: a variable‐selection cluster‐analysis approach. *Review of Income and Wealth*, 61(4), 702-722.

Levy-Yeyati, E. & Struzenegger F. (2023) Exchange Rate Regimes 20 years later: The prevalence of floats. *RedNIE Working Paper Series N182*

\*Lusk, J. L. (2017). Consumer research with big data: applications from the food demand survey (FooDS). *American Journal of Agricultural Economics,* 99(2), 303-320.

Lopez, Juan Cruz *Caracterización socioeconómica de clusters electorales* (Tesis grado de Lic. En economia, 2024)

# **ANÁLISIS DE PREDICCIÓN I: MÉTODOS SUPERVISADOS DE REGRESIÓN & CLASIFICACIÓN. PROBLEMA DE OVERFITTING & TÉCNICAS DE REMUESTREO**

|  |
| --- |
| **Tema 5: Modelo de Regresión lineal trade-off sesgo-varianza** |

**†** James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in Python.* **Chap 3**

**†** Sosa Escudero, W., 2024, *Viajar al futuro (y volver a contarlo)*, 1a edición, Siglo XXI Editores Argentina, Buenos Aires. **Cap 5 pag 112 a 140**

Brinatti, A., Cavallo, A., Cravino, J., & Drenik, A. (2021). The international price of remote work (No. w29437). *National Bureau of Economic Research.*

Cavallo, A. (2017) "Are Online and Offline Prices Similar? Evidence from Multi-Channel Retailers" *American Economic Review*. *107* (1). 283-303.

|  |
| --- |
| **Tema 6: Clasificación I: Introducción a Clasificación, Logit & Vecinos Cercanos (KNN)** |

**†** James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in Python.* **Chap 4.1** a **4.3**

**†** Sosa Escudero, W., 2024, *Viajar al futuro (y volver a contarlo)*, 1a edición, Siglo XXI Editores Argentina, Buenos Aires. **Cap 4 pag 89 a 98**

Mougenot, B., Amaya, E., Mezones-Holguin, E., Rodriguez-Morales, A. J., & Cabieses, B. (2021). Immigration, perceived discrimination and mental health: evidence from Venezuelan population living in Peru. *Globalization and health*, 17, 1-9.

​​Tollenaar, N., & van der Heijden, P. G. (2013). Which method predicts recidivism best?: a comparison of statistical, machine learning and data mining predictive models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 176(2), 565-584.

|  |
| --- |
| **Tema 7: Clasificación II: Análisis discriminante. Análisis ROC. Comparacion de metodos y regression de Poisson** |

**†** James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in Python.* **Chap 4.4** a **4.6**

**†** Sosa Escudero, W., 2024, *Viajar al futuro (y volver a contarlo)*, 1a edición, Siglo XXI Editores Argentina, Buenos Aires. **Cap 4, 98 a 111**

Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, 55(2), 107-120.

\* Baylé, Federico (2016) “Detección de villas y asentamientos informales en el partido de La Matanza mediante teledetección y sistemas de información geográfica” Tesis de Maestría.

Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264), 1073-1076.

Mullally, C., Rivas, M., & McArthur, T. (2021). Using Machine Learning to Estimate the Heterogeneous Effects of Livestock Transfers. *American Journal of Agricultural Economics*

|  |
| --- |
| **Tema 8: Métodos de Remuestreo: Cross-validation & Bootstrap** |

**†** James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in Python.* **Chap 5**

**†** Sosa Escudero, W., 2021, *Big data*, 7a edición, Siglo XXI Editores, Buenos Aires. **Cap 5**

\*Hastie, T., Tibshirani, T. Y Freedman, J. (2013) *The Elements of Statistical Learning.***Chap 7.2, 7.10**

\* Sosa Escudero, W., & Gasparini, L. (2000). A note on the statistical significance of changes in inequality. *Económica*, 46.

# **ANÁLISIS DE PREDICCIÓN II: MÉTODOS LINEALES Y NO LINEALES**Brain in head outline

|  |
| --- |
| **Tema 9: Modelos Lineales de Selección de Variables. Regularización: LASSO & RIDGE. Elastic Net, comparaciones & discusión de causalidad usando LASSO** |

**†** James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in Python.* **Chap 6.1** y **6.2**

Belloni, V. Chernozhukov, C. Hansen: “High-Dimensional Methods and Inference on Structural and Treatment Effects,” *Journal of Economic Perspectives*, 28 (2), Spring 2014, 29-50.

\*Bajari, P., Nekipelov, D., Ryan, S. P., & Yang, M. (2015). Machine learning methods for demand estimation. *American Economic Review*, 105(5), 481-485.

Chernozhukov, V., Demirer, M., Duflo, E., & Fernandez-Val, I. (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India (No. w24678). *National Bureau of Economic Research.*

\*Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. *American Economic Review*, 105(5), 491-495.

\* Wüthrich, K., & Zhu, Y. (2023). Omitted variable bias of Lasso-based inference methods: A finite sample analysis. *Review of Economics and Statistics*, 105(4), 982-997.

Zou, H. y Hastie, T., 2005, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society*, 67, 2, 301-320.

|  |
| --- |
| **Tema 10: Árboles (CART) & Métodos de ensamble basados en Árboles: Boosting, Bagging, Random Forest** |

**†** James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in Python.* **Chap 8**

**†** Sosa Escudero, W., 2021, *Big data*, 7a edición, Siglo XXI Editores, Buenos Aires **Cap 3, pag 85 a 94.**

Breiman, L. (2003). Statistical modeling: The two cultures. *Quality control and applied statistics*, 48(1), 81-82.

\* Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

Hothorn, T., & Zeileis, A. (2015). partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research*, 16(1), 3905-3909.

List, J. A., Muir, I., & Sun, G. (2024). Using machine learning for efficient flexible regression adjustment in economic experiments. *Econometric Reviews*, *44*(1), 2-40.

\* Keely, L. C., & Tan, C. M. (2008). Understanding preferences for income redistribution. *Journal of Public Economics*, 92(5), 944-961.

Riascos, A., & Serna, N. (2017, October). Predicting annual length-of-stay and its impact on health. *Medical Informatics and Healthcare* (pp. 27-34). PMLR.

Varian, H. R. (2014). Big data: New tricks for econometrics. *The Journal of Economic Perspectives*, 28(2), 3-27.

\* Wager, S. and Athey, S., 2018. Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests. *Journal of the American Statistical Association*, 113(523), pp.1228- 1242.

|  |
| --- |
| **Tema 11: Modelos basados en la reducción de la dimensionalidad y Modelos no lineales (polinomios, splines, local regressions)** |

**†** James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in Python.* **Chap 6.3** y **7.1** a **7.6**

\* Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer, Berlin: Springer series in statistics. **Chap. 3.6.2, 5.1, 5.2**

Bertrand, M., Kamenica, E., & Pan, J. (2015). Gender identity and relative income within households. *Quarterly Journal of Economics*, *130*(2), 571-614

Diamond, R. (2016). The determinants and welfare implications of US Workers’ diverging location choices by skill: 1980-2000. *American Economic Review*, *106*(3), 479–524.

Lee, D. S., Moretti, E., & Butler, M. J. (2004). Do voters affect or elect policies? Evidence from the US House. *Quarterly Journal of Economics*, 119(3), 807-859.

# **CRONOGRAMA TENTATIVO:** [**link a Teams**](https://teams.microsoft.com/l/message/19:gHrQ_WZCxoqpzOeqLGs3zR6LXrMk9QGFXkSeoS6drZk1@thread.tacv2/1758753017826?tenantId=4c818f79-ab84-4552-9b7c-2fe715b0d0d5&groupId=c5971737-d1c4-4a70-93fb-d394f52e7fb8&parentMessageId=1758753017826&teamName=Seminarios%20Optativos%20-%20Taller%20de%20Programaci%C3%B3n-4028687&channelName=General&createdTime=1758753017826)

# Clases, Lecturas Obligatorias con † y Fechas de Entrega Claves

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Miercoles 19:00 a 22:00** | |  |
| 24-sept | 1 | **Intro.Big Data & Visualizacion de Datos** |  |
|  |  | (videos) Introduccion a Python, webscrapping & jugando con APIs |  |
| 1-oct | 2 | **Analisis descriptivo II: Histogramas, Kernels** |  |
|  |  |  |  |
| 8-oct | 3 | **Metodos No Supervisados I PCA** |  |
|  |  | † James, et al (2023) Chap 6.3 |  |
| 8-oct | TP | Entrega TP 1: Cleaning & Descriptive statistics EPH |  |
| 15-oct | 4 | **Metodos No Supervisados II Cluster** |  |
|  |  | † Sosa Escudero, W. (2021) Cap 3, pag 69 a 76 |  |
| 22-oct | 5 | **Modelo Lineal. Regresion** |  |
|  |  | † Sosa Escudero, W. (2021) Cap 3, pag 76 a 85 |  |
| 22-oct | TP | Entrega TP 2: Unsurpervised exercise EPH |  |
| 29-oct | 6 | **Clasificacion I: Bayes, Logit. KNN neighbor. Curvas ROC** |  |
|  |  | † James, et al (2023) Chap 4.1 a 4.6 |  |
| 5-nov | 7 | **Remuestreo. Overfitting. CV. Bootstrapping** |  |
|  |  | † Sosa Escudero, W. (2021) Cap 5. Pag 121 a 130 |  |
| 5-nov |  | **1ra Instancia: Exposiciones grupales** |  |
| 12-nov | 8 | **Regularizacion. LASSO. Ridge. Elastic Net.PCR. Comparacion. Aplicaciones** | |
|  |  | † James, et al (2023) Chap 6 |  |
| 12-nov | TP | Entrega TP 3: EPH - Clasificación de hogares pobres |  |
| 19-nov | 9 | **CART. Bagging, Boosting** |  |
|  |  | † Sosa Escudero, W. (2021) Cap 3, pag 85 a 94 |  |
| 19-nov |  | **3ra Instancia: Borrador de ideas** |  |
| 26-nov | 10 | **Ensamble: Random Forest. Aplicaciones** |  |
|  |  | † James, et al (2023) . Chap 8.2.2., 8.2.5 |  |
| 26-nov | TP | Entrega TP 4: EPH - Regularización & CART de hogares pobres |  |
| 3-dic | 11 | **PCR & PLS. Modelos no lineales** |  |
|  |  | † James, et al (2023) Chap 7.1 a 7.6 |  |
| 10-dic | 12 | **2da instancia: Presentación de posters** |  |

# **APÉNDICE**

En este apéndice, hay más papers relacionados con “Big Data” y distintos usos con datos georeferenciados en economía. Estos son los “clásicos” a leer en esta temática.

|  |
| --- |
| **Otros papers simpáticos sobre Big Data** |

Donaldson, D., & Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, 30(4), 171-98. (este es un “buen literature review”)

Henderson, J. V., A. Storeygard, and D. N. Weil. A Bright Idea for Measuring Economic Growth. *American Economic Review* 101.3 (2011): 194-199.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google flu: traps in big data analysis. *Science*, 343(6176), 1203-1205.

Lazer, W. & Kennedy, R. (2015). What We Can Learn from the Epic Failure of Google Flu Trends, *Wired*, 10.01.15.

1. **Agradecimientos**: El presente programa está desarrollado e inspirado siguiendo el curso 2020 de Big Data por el Profesor Walter Sosa Escudero ([Website actualizado en 2024i](https://bigdataudesa.weebly.com/)) y se actualizó con la bibliografía del curso *ACE 592 Big Data in Empirical Economics* (Fall 2022) por el Profesor [Peter Christensen](https://www.uiuc-bdeep.org/christensen) (University of Illinois Urbana-Champaign). Cada semestre este programa se actualiza en función de las clases dictadas en las materias Big Data para Economistas (E337) y Ciencia de Datos (CC408), como el dictado en otras universidades (UBA y UNT). [↑](#footnote-ref-1)
2. Sólo se admitirá un número *limitado* de grupos con **2 (dos)** personas dependiendo del número de alumnos. Los grupos deben ser formados por alumnos dentro de cada tutorial. [↑](#footnote-ref-2)
3. Dependiendo de la cantidad de grupos, puede que algunas presentaciones sean en clase tutorial. [↑](#footnote-ref-3)
4. En el **Apéndice** de este programa, los alumnos pueden encontrar otras sugerencias bibliográficas de introducción a “Machine learning”, debates “Big Data”, y uso de datos georeferenciados. [↑](#footnote-ref-4)