

PROGRAMA DE ASIGNATURA		
1. Nombre de la Actividad Curricular		
Métodos computacionales para las ciencias sociales		
2. Nombre de la Actividad Curricular en Inglés		
Computational methods for social science		
3. Área a la que corresponde la Actividad Curricular		
Profundización Metodológica		
4. Descripción de la Actividad Curricular (No más de 200 palabras)		
<p>Este curso revisa herramientas computacionales para las ciencias sociales en tres áreas: extracción de información a partir de fuentes no tradicionales, procesamiento de lenguaje natural y análisis de datos geoespaciales. Es un curso aplicado, con énfasis en el abordaje de problemas cotidianos del quehacer profesional. Se utilizará principalmente el lenguaje de programación R y, en menor medida, el lenguaje Python. Las y los estudiantes formarán grupos y realizarán tres proyectos a lo largo del semestre. Al terminar el curso, los y las estudiantes sabrán extraer información mediante código desde la web y obtener conclusiones a partir de fuentes de datos no tradicionales. Asimismo, contarán con una visión amplia respecto a las potencialidades que tienen las nuevas fuentes de información para el quehacer sociológico.</p>		
5. Nombre Completo del Docente(s) Responsable(s)		
Klaus Ignacio Lehmann Meléndez		
6. Unidad Académica / organismo de la unidad académica que lo desarrolla		
Facultad de Ciencias Sociales / Departamento de Sociología		
7. Semestre Académico en que se dicta		
Ciclo de Profundización		
8. Ámbito		
Investigación / intervención		
9. Horas de trabajo	Horas semanales de trabajo presencial	Horas semanales de trabajo no presencial



10. Tipo de créditos SCT	3 horas	3 horas
11. Número de Créditos SCT – Chile 4 créditos		
12. Requisitos	Teoría sociológica clásica ; Análisis de información cualitativa ; Estadística correlacional ; Estrategias de investigación cuantitativa	
13. Propósito general del curso	<p>Introducir a los y las estudiantes en el uso de métodos computacionales para las ciencias sociales, poniendo énfasis en problemas habituales del quehacer profesional.</p> <p>El curso muestra a los y las estudiantes un abanico de nuevas fuentes de información y técnicas de análisis, posibles de ser utilizadas en el quehacer sociológico, tanto desde un punto de vista académico como profesional.</p>	
14. Competencias	<p>1a Delimitar, conceptualizar y analizar diversos objetos de investigación social, con especial énfasis en aquellos relacionados con los procesos de transformación del país y Latinoamérica</p> <p>1c Manejar diversas estrategias metodológicas de las ciencias sociales</p> <p>1d Manejar un conjunto de herramientas para el procesamiento y análisis de información</p> <p>1e Transmitir los conocimientos derivados de la práctica investigativa, así como aquellos adquiridos durante el proceso formativo.</p>	
15. Subcompetencias	1.4 Contribuir a generar conocimiento	



	<p>sociológico en el marco de estudios y/o procesos de investigación donde se articulen creativamente las dimensiones teórica, metodológica y práctica.</p> <p>1.5 Comunicar los saberes disciplinares de manera pertinente a las características de distintos contextos y audiencias, utilizando diversas estrategias y formatos.</p>
<p>16. Resultados de Aprendizaje</p> <p>Obtener información web programáticamente, para el abordaje de problemas relevantes de las ciencias sociales.</p> <p>Utilizar técnicas computacionales de análisis mediante el lenguaje R, para la obtención de conocimiento accionable en el marco del quehacer profesional.</p> <p>Resolver desafíos de automatización mediante escritura de código, para la optimización de procesos en el marco del quehacer profesional.</p>	
<p>17. Saberes / Contenidos</p> <p>UNIDAD 1: Introducción a los métodos computacionales en ciencias sociales y reforzamiento en programación</p> <ul style="list-style-type: none">• Métodos computacionales y el surgimiento de unas nuevas Ciencias Sociales• Reforzamiento en programación<ul style="list-style-type: none">◦ Tipos y estructuras de datos básicos◦ Operadores lógicos◦ Controles de flujo (if, else, else if)◦ Loops (for, while)◦ Declaración de funciones◦ Manipulación de listas <p>UNIDAD 2: Herramientas para el procesamiento de grandes volúmenes de información</p> <ul style="list-style-type: none">• Estrategias básicas de carga de datos	



- Procesamiento con data.table
- Procesamiento con apache arrow
- Procesamiento con duckdb

UNIDAD 3: Extracción de datos

- ¿Qué es una API?: extrayendo datos de twitter y spotify
- Web scraping: Extracción de texto desde la web
 - Extracción de información de páginas estáticas
 - Extracción de información de páginas dinámicas

UNIDAD 4: Procesamiento de lenguaje natural

- Expresiones regulares para el preprocesamiento de datos
- Exploración básica de textos
- Herramientas de la ciencia política
- Convirtiendo el texto en vectores mediante TF-IDF
- Operaciones algebraicas sobre vectores de texto
- Recuperación de información
- Construcción de un clasificador de textos
- Word embeddings
- Transformers y modelos de atención (Huggingface)

UNIDAD 5: Análisis de datos geoespaciales

- Relevancia de los datos geoespaciales para las ciencias sociales
- Formatos y herramientas en R para trabajar con datos geoespaciales
- Elaboración de mapas con información geoespacial
- Cálculo de distancias para la estratificación socioeconómica



- Introducción al procesamiento de imágenes satelitales

18. Metodología

El curso tendrá una metodología mixta que combinará sesiones expositivas y prácticas. El abordaje teórico y práctico de las herramientas se realizará considerando problemas reales del quehacer profesional y artículos científicos relevantes en el área de las ciencias sociales computacionales. Durante las sesiones expositivas se describirán los métodos en términos teóricos y conceptuales, invitando a los y las estudiantes a reflexionar sobre las aplicaciones para sus temas de interés.

Dado que el énfasis del curso está puesto en el aprendizaje de herramientas, aproximadamente el 40% del tiempo estará destinado a poner en práctica los contenidos relatados durante las sesiones expositivas. Mediante ejercicios, los y las estudiantes aplicarán en R las herramientas revisadas, bajo la guía del profesor.

19. Evaluación

Se realizarán 4 evaluaciones. La primera de ellas corresponde a una tarea individual. La segunda, consiste en la presentación de un artículo científico relacionado con los contenidos del curso. La tercera y cuarta corresponden a proyectos grupales, en los que los y las estudiantes presentarán los resultados obtenidos a partir de la aplicación de las herramientas revisadas en el curso.

- (1) Tarea de programación (20%)
- (2) Presentación artículo (10%)
- (3) Presentación grupal 1 en R (30%).
- (4) Presentación grupal 2 en R (40%).

Adicionalmente, habrán incentivos para promover la participación en clases.

20. Requisitos de aprobación

Se requiere un 80% de asistencia como requisito mínimo para aprobar el curso

Escala de evaluación: de 1,0 a 7,0, con incrementos de un decimal

Nota de aprobación mínima: 4,0

Notas arriba de 5 eximen examen

21. Palabras Clave

Ciencias sociales computacionales (CSC); ciencia de datos; datos geoespaciales;



procesamiento de lenguaje natural; machine learning.

22. Bibliografía Obligatoria

Lazer, David, Eszter Hargittai, Deen Freelon, Sandra Gonzalez-Bailon, Kevin Munger, Katherine Ognyanova, and Jason Radford. 2021. *Meaningful Measures of Human Society in the Twenty-First Century*. *Nature* 595 (7866): 189–96. <https://doi.org/10.1038/s41586-021-03660-7>. (*)

Lazer, David et al 2009. *Computational Social Science*. *Science* 323(5915):721–23.

Hilbert, M & López, P 2011. *The World's Technological Capacity to Store, Communicate, and Compute Information*. *Science*. (*)

Salganik, Matthew J. 2018. *Bit by Bit : Social Research in the Digital Age*. Princeton, New Jersey: Princeton University Press. Capítulos 1 y 2. (*)

Gennaro, Gloria, y Elliott Ash. 2021. *Emotion and Reason in Political Language*. *The Economic Journal* 132 (643): 1037-59. En línea: <https://doi.org/10.1093/ej/ueab104> (*)

Freire-Vidal, Y.; Graells-Garrido, E.; Rowe, F. *A Framework to Understand Attitudes towards Immigration through Twitter*. *Appl. Sci.* 2021, 11, 9689. En línea: <https://www.mdpi.com/2076-3417/11/20/9689> (*)

Nguyen, Dong, A. Seza Doğruöz, Carolyn P. Rosé, and Franciska de Jong. 2015. *Computational Sociolinguistics: A Survey*. ArXiv:1508.07544 [Cs]. En línea: <https://arxiv.org/abs/1508.07544> (*)

Mäntylä, Mika V., Daniel Graziotin, and Miikka Kuuttila. 2018. *The Evolution of Sentiment Analysis—A Review of Research Topics, Venues, and Top Cited Papers*. *Computer Science Review* 27:16–32. En línea: <https://www.sciencedirect.com/science/article/abs/pii/S1574013717300606> (*)

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. *Fake News on Twitter during the 2016 U.S. Presidential Election*. *Science* 363(6425):374–78. doi: 10.1126/science.aau2706. (*)

Cartes, C., Asahi, K. & Fernández, R. *Riots and subways, a relationship moderated by the neighborhood's income level*. *Sci Rep* 12, 10557 (2022). <https://doi.org/10.1038/s41598->



022-14859-7 (*)

Luis Fuentes Arce, María Inés Ramírez, Sebastián Rodríguez & Andrés Señoret (2023) *Socio-spatial differentiation in a Latin American metropolis: urban structure, residential mobility, and real estate in the high-income cone of Santiago de Chile*, International Journal of Urban Sciences, 27:2, 195-214, DOI: 10.1080/12265934.2022.2116087 (*)

González-Espejo F, Astroza S, Hurtubia R. 2022. *On the relation between school and residential location choice: Evidence of heterogeneous strategies from Santiago de Chile*. Journal of Transport Geography. <https://doi.org/10.1016/j.jtrangeo.2022.103359>

* Corresponden al listado de lecturas que podrían ser elegidas por los y las estudiantes para presentar al curso.

23. Bibliografía Complementaria

Bojanowski, Piotr, Edouard Grave, Armand Joulin, y Tomás Mikolov. 2016. *Enriching Word Vectors with Subword Information*. CoRR abs/1607.04606.
<http://arxiv.org/abs/1607.04606>.

Poole, Keith, y Howard Rosenthal. 1983. *A Spatial Model for Legislative Roll Call Vote Analysis*. American Journal of Political Science 29 (agosto). <https://doi.org/10.2307/2111172>.

Fábrega, Jorge. 2022. *Ordenamiento Ideológico en la Convención Constitucional Chilena*. Revista de Ciencia Política 42 (1): 127-51. <https://doi.org/10.4067/s0718-090x2022005000106>.

James Lo, Sven-Oliver Proksch and Jonathan B. Slapin *Ideological Clarity in Multiparty Competition: A New Measure and Test Using Election Manifestos*. British Journal of Political Science, Available on CJO 2014 doi:10.1017/S0007123414000192.

Mejia-Arbelaez, C.; Sarmiento, O.L.; Mora Vega, R.; Flores Castillo, M.; Truffello, R.; Martínez, L.; Medina, C.; Guaje, O.; Pinzón Ortiz, J.D.; Useche, A.F.; et al. *Social Inclusion and Physical Activity in Ciclovía Recreativa Programs in Latin America*. Int. J. Environ. Res. Public Health 2021, 18, 655. <https://doi.org/10.3390/ijerph18020655>

Benoit, K., Munger, K., & Spirling, A. (2019). *Measuring and Explaining Political Sophistication through Textual Complexity*. American Journal of Political Science, 63(2), 491–508.



24. Recursos Web

Araya, S. & Cruz, A. Carga de Bases. En F. Urdinez y A. Cruz (Eds.), *Analizar Datos Políticos*. En línea: <https://arcruz0.github.io/libroadp/>

Cruz, A. 2019. Manejo de Datos. En F. Urdinez y A. Cruz (Eds.), *Analizar Datos Políticos*. En línea: <https://arcruz0.github.io/libroadp/>

Escobar, A., y Ortiz, G. 2019. Manejo y Mapeo de Datos Espaciales. En F. Urdinez y A. Cruz (Eds.), *Analizar Datos Políticos*. En línea: <https://arcruz0.github.io/libroadp/>

Urdinez, F., y Cruz, A. 2019. Manejo Avanzado de Datos Políticos. En F. Urdinez y A. Cruz (Eds.), *Analizar Datos Políticos*. En línea: <https://arcruz0.github.io/libroadp/>

Wickham, H. y Grolemund, G. 2017. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Sebastopol, CA: O'Reilly Media. En línea: <https://r4ds.had.co.nz/>

Wickham, H. 2019. *Advanced R*. Second Edition. Chapman and Hall. En línea: <https://advr.hadley.nz/index.html>

INE 2020. Curso de R básico. En línea: <https://github.com/clases-r-basico>

INE 2022. Curso de R intermedio. En línea: <https://github.com/clases-r-intermedio>

25. Programación por sesiones

Semana 1:

- Presentación del curso (contenidos y logística)
- Estado actual de las ciencias sociales computacionales
- Nuevas competencias profesionales en el quehacer sociológico
- Tipos de datos en R
- Operadores lógicos

Semana 2:

- Estructuras de datos básicas
- Controles de flujo



- Loops (for y while)
- Sesión práctica

Semana 3:

- Declaración de funciones
- Manipulación de listas
- Sesión práctica

Semana 4:

- Carga de archivos de gran tamaño
- Procesamiento con data.table
- Procesamiento con apache arrow
- Procesamiento con duckdb

Semana 5:

- ¿Qué es una API?
- Extrayendo datos de spotify
- Extrayendo datos de redis
- Sesión práctica

Semana 6:

- Web scraping: Extracción de texto desde la web
 - Extracción de información de páginas estáticas
 - Extracción de información de páginas dinámicas
- Sesión práctica

Semana 7: Vacaciones de septiembre



Semana 8: Sesión de presentaciones trabajo 1

Semana 9:

- Expresiones regulares para el preprocesamiento de datos
- Exploración básica de textos
- Convirtiendo el texto en vectores mediante TF-IDF

Semana 10:

- Sesión de presentaciones
- Operaciones algebraicas sobre vectores de texto
- Recuperación de información

Semana 11:

- Construcción de un clasificador de textos
- Breve historia de los modelos de lenguaje
- Word embeddings

Semana 12:

- Transformers y modelos de atención
- Relevancia de los datos geoespaciales para las ciencias sociales

Semana 13: Trabajo autónomo

Semana 14: Formatos y herramientas en R para trabajar con datos geoespaciales

Semana 15:

- Relevancia de los datos geoespaciales para las ciencias sociales
- Formatos de datos geoespaciales
- Elaboración de mapas con información geoespacial

Semana 16:

- Cálculo de distancias para la estratificación socioeconómica



FACULTAD DE
**CIENCIAS
SOCIALES**
UNIVERSIDAD DE CHILE

Escuela de
PREGRADO

- Introducción a las imágenes satelitales.

Semana 17:

- Sesión de presentaciones trabajo 2