

PROGRAMA DE ASIGNATURA CURSOS DE FORMACIÓN GENERAL (CFG)

2° semestre 2025

Departamento de Pregrado
Vicerrectoría de Asuntos Académicos
Universidad de Chile

1. NOMBRE Y CÓDIGO DEL CURSO

Nombre	Ciencia abierta: Análisis crítico de datos públicos mediante software libre
Código	VA-01-0327-200

2. NOMBRE DEL CURSO EN INGLÉS

Open Science: Critical analysis of public data using open software

3. EQUIPO DOCENTE

Docentes responsables	Unidad académica
Sandra Andrea Flores Alvarado	Instituto de Salud Poblacional (Escuela de Salud Pública), Facultad de Medicina
Felipe Andrés Medina Marín	Instituto de Salud Poblacional (Escuela de Salud Pública), Facultad de Medicina

Docentes invitados/as	Unidad académica
María Begoña Carroza Escobar	Departamento de Promoción de la Salud de la Mujer y el Recién Nacido, Facultad de Medicina
René Alonso Lagos Barrios	Escuela de Ingeniería y Ciencias, Facultad de Ciencias Físicas y Matemáticas, Área de Ingeniería e Innovación
María Pía Muñoz Quezada	Instituto de Salud Poblacional, Escuela de Salud Pública, Facultad de Medicina
Mauricio Fuentes Alburquenque	Instituto de Salud Poblacional, Escuela de Salud Pública, Facultad de Medicina
Álvaro Besoain Saldaña	Departamento de Kinesiología y Dirección de Asuntos Estudiantiles y Comunitarios, Facultad de Medicina

Daniel Larenas Rosa	Departamento de Fonoaudiología, Facultad de Medicina
Natalia Tamara Santis Alay	Departamento de Tecnología Médica, Facultad de Medicina
María Isabel Villanueva	Dirección de Servicios de Información y Bibliotecas (SISIB)
Cristián Calabrano Torres	Dirección de Servicios de Información y Bibliotecas (SISIB)
Rafael Castillo Guerrero	Dirección de Servicios de Información y Bibliotecas (SISIB)
Además, el curso contará con la colaboración del Grupo de Ciencia de Datos para la Salud Pública, integrado por la y el docente responsable y algunas/os de las/os docentes colaboradores; y también por el Grupo de Ciencia de Datos SISIB.	

Ayudante	
-----------------	--

4. CARGA ACADÉMICA Y CRÉDITOS SCT-CHILE

Duración total del curso	17 semanas
Nº máximo de horas de trabajo sincrónico / presencial semanal	1,5 horas
Nº máximo de horas de trabajo asincrónico / no presencial semanal	1,5 horas
Nº de créditos SCT	2 SCT

5. MODALIDAD, DÍA Y HORARIO

Modalidad	Semipresencial. El curso tendrá sesiones presenciales y sesiones remotas asincrónicas (ver cronograma semanal).
Día	Miércoles
Horario	15:30 – 17:00
Lugar	Clases presenciales en el Campus Andrés Bello (Santiago Centro)

6. COMPETENCIAS SELLO A LAS QUE CONTRIBUYE EL CURSO

X	1. Capacidad de investigación, innovación y creación
X	2. Capacidad de pensamiento crítico y autocrítico
	3. Capacidad para comunicarse en contextos académicos, profesionales y sociales
X	4. Compromiso ético y responsabilidad social y ciudadana
	5. Compromiso con el desarrollo humano y sustentable
	6. Compromiso con el respeto por la diversidad y multiculturalidad
	7. Compromiso con la igualdad de género y no discriminación

7. PROPÓSITO FORMATIVO DEL CURSO

El tema principal a abordar en este CFG es el **análisis de datos públicos y abiertos** disponibles en internet como ejercicio para adquirir un conjunto de habilidades del siglo XXI que permita a sus estudiantes abordar de manera informada, crítica, colaborativa y transdisciplinaria, los desafíos que supone un mundo sobrecargado de información que se genera a cada segundo.

El alcance del curso es la introducción a sus estudiantes a un abordaje crítico de la ciencia de datos a través del uso de **software libre** y gratuito para su aplicación en el análisis de datos. Para ello, sus estudiantes aprenderán a buscar, descargar e integrar datos de diferentes repositorios públicos, principalmente gubernamentales, con el sentido de proponer y ejecutar un proyecto de descripción y visualización de información de interés ciudadano que refleje la diversidad de las poblaciones en estudio, con la finalidad de generar hipótesis de investigación.

Este curso contribuirá a la formación de sus estudiantes en el **trabajo colaborativo inter y transdisciplinario** en la formulación y respuesta a preguntas de investigación de interés público. Además, el aprendizaje de los **aspectos básicos de programación** les aportará al desarrollo de un pensamiento algorítmico transferible a otras esferas de su vida. Por otro lado, aprender los aspectos básicos de la ciencia de datos y la descripción cuantitativa y visual de la información, ayudará al desarrollo del pensamiento estadístico, orientado al reconocimiento de la variabilidad de los fenómenos estudiados. Todo esto les permitirá cuestionar cómo se presentan resúmenes y gráficos, concientizando sobre cómo diferentes perspectivas pueden influir en los resultados comunicados dentro y fuera del ámbito académico, moldeando las percepciones de la ciudadanía.

8. RESULTADOS DE APRENDIZAJE DEL CURSO

Al finalizar este curso, el/la/le estudiante será capaz de:

1. **Analizar críticamente datos públicos:** El/la/le estudiante analizará conjuntos de datos públicos obtenidos de repositorios abiertos utilizando software libre, aplicando técnicas de análisis de datos con un enfoque crítico, que les permita identificar patrones, tendencias y posibles sesgos en la información.
2. **Desarrollar y presentar proyectos de visualización de datos:** El/la/le estudiante desarrollará y presentará proyectos de visualización de datos que reflejen la diversidad de las poblaciones en estudio, utilizando herramientas de software libre para crear visualizaciones claras y efectivas que respondan a preguntas de investigación de interés público.
3. **Reflexionar sobre el uso ético de datos y la responsabilidad ciudadana:** El/la/le estudiante reflexionará sobre las implicancias éticas del uso de datos públicos, abordando temas como la privacidad, la representación justa de las poblaciones y el impacto social de la visualización de datos, con el fin de promover un uso responsable y ético de la información en contextos ciudadanos.

Estos resultados están alineados con las competencias sello de pensamiento crítico, investigación, innovación, responsabilidad social y ética, y buscan contribuir a la formación integral y transversal de los estudiantes en un entorno interdisciplinario.

9. SABERES FUNDAMENTALES O CONTENIDOS

El curso se estructura en **cuatro ejes temáticos**, cada uno diseñado para desarrollar las competencias necesarias para el análisis crítico de datos públicos mediante software libre. Sus contenidos incluyen:

1. Introducción a la ciencia abierta y ciudadana: Conceptos básicos, importancia de la ciencia abierta y la participación ciudadana en la investigación.
2. Obtención y preparación de datos públicos: Métodos documentados y reproducibles de búsqueda, descarga, limpieza y preparación de datos provenientes de repositorios públicos.
3. Análisis de datos y visualización: Técnicas fundamentales de análisis de datos utilizando R, incluyendo la creación de tablas, gráficos y mapas.
4. Consideraciones éticas y responsabilidad social: Reflexión sobre las implicancias éticas del uso de datos, la privacidad, la equidad en la representación y el impacto social.

Estos contenidos se integran para proporcionar a los estudiantes una comprensión completa y práctica de la ciencia de datos en contextos ciudadanos y abiertos, enfocándose en habilidades transferibles y en la responsabilidad ética en el manejo de información pública.

10. METODOLOGÍA Y EVALUACIÓN: PLAN DE TRABAJO

A) Descripción general de la metodología y la evaluación del curso:

Se utilizará una estrategia de **aprendizaje basado en proyectos**, que se llevarán a cabo de manera colaborativa en **equipos de trabajo interdisciplinarios** compuestos por 5 o 6 estudiantes provenientes de diferentes carreras de pregrado (máximo, 2 estudiantes de una misma carrera por grupo), a lo largo de todo el semestre.

El curso considerará **sesiones presenciales y no presenciales** (ver cronograma semanal). Las 7 sesiones no presenciales se enfocarán en el **aprendizaje del uso de software R** para la administración de datos, creación de tablas resumen y para la visualización de datos públicos abiertos, guiadas a través de cápsulas grabadas con demostraciones paso a paso del uso del software, y guías de trabajo para su aplicación en el contexto del proyecto grupal, apoyado a través de discusiones en el **foro** del curso. Como **requisito** de aprobación, se requerirá la entrega de todas las **evaluaciones formativas asincrónicas**.

Las 9 sesiones presenciales se enfocarán en la **discusión grupal guiada** de los resultados obtenidos a partir de los análisis realizados en sesiones no presenciales. Estas sesiones tendrán tres etapas. La primera etapa será una **exposición breve** (20 a 30 minutos) por parte del equipo docente en que se problematizará el uso crítico de los datos públicos para considerar perspectivas de género y equidad en los análisis, entregando ejemplos útiles para la discusión posterior. La segunda etapa consistirá en una **discusión intra grupo**, en que se realizará una autoevaluación formativa del trabajo usando pautas proporcionadas por el equipo docente. La tercera, consistirá en la **exposición frente a otro grupo**, que evaluará a través de una rúbrica de evaluación formativa por pares y entregará retroalimentación, para luego intercambiar roles. Como requisito de aprobación del curso, la **asistencia mínima** a las sesiones presenciales será del 65%.

Además, 3 sesiones presenciales tendrán **evaluaciones sumativas**, enfocadas en los **avances del proyecto** de trabajo, correspondientes a las sesiones 5, 12 y 17. La sesión 5 consistirá en la presentación de la propuesta de trabajo, la sesión 12 en un primer avance, y la sesión 17 en la presentación del trabajo final. Cada una de estas evaluaciones, incluida la última, considerará la entrega de **retroalimentación** por parte del equipo docente. Para cada una de estas sesiones se hará una invitación abierta al grupo de Ciencia de Datos para la Salud Pública de la Escuela de Salud Pública, un grupo abierto y transdisciplinario, para que haga de público participativo entregando retroalimentación con base en sus diversos saberes.

B) Resumen del esquema de evaluaciones calificadas del curso:

Actividad evaluada	Tipo de actividad	Ponderación en nota final	Semana estimada de entrega
Presentación de propuesta proyecto	grupal	20%	Semana 5
Presentación de avance de proyecto	grupal	30%	Semana 12
Presentación de proyecto realizado	grupal	50%	Semana 17

C) Planificación y cronograma preliminar del curso:

Semana	Fecha	Actividades		Evaluación
		Presenciales	No presenciales - Asincrónicas Trabajo autónomo	
1	13-ago	<p>Clase: Introducción al curso.</p> <p>Clase: Presentación de repositorios de datos públicos de diferentes áreas.</p> <p>Dinámica: <i>speed dating</i> para encontrar grupos de trabajo afines.</p>	<p>Trabajo grupal: Elección del tema y base de datos a trabajar.</p>	Inscripción de grupo de trabajo vía U-Cursos. ¹
2	20-ago	No hay sesión presencial.	<p>Clase: Introducción a R y RStudio. Instalación del software. Elementos básicos del uso de R para el análisis de datos. Reproducibilidad y documentación de códigos.</p>	Publicación en foro de captura de pantalla del programa instalado y funcionando correctamente.
3	27-ago	<p>Clase breve (ejemplos): Análisis de género y equidad con datos públicos.</p> <p>Trabajo grupal: Planteamiento de problema de investigación: ¿qué pregunta nos interesa resolver con nuestros datos? ¿por qué es un problema relevante? ¿cómo podríamos resolverla?</p>	<p>Trabajo grupal: avance en presentación de clase 5.</p>	Envío de problema y propuesta de resolución y bases de datos a utilizar.

¹ Se aceptarán incorporaciones a grupos preexistentes hasta la 3ra semana del curso.

4	03-sep		<p>Clase: Carga de bases de datos y análisis descriptivos básicos en R.</p> <p>Trabajo grupal: análisis descriptivos básicos de las variables de interés para el tema de estudio.</p>	Publicación en el foro de alguno(s) de sus primeros resúmenes numéricos y gráficos para el problema en investigación.
5	10-sep	Presentación del proyecto de investigación.	<p>Trabajo grupal (previo a la actividad sincrónica): preparación de presentación de proyecto.</p>	Evaluación sumativa: Presentación.
6	17-sep	SEMANA DE PAUSA TRANSVERSAL		
7	24-sep		<p>Clase: Administración de datos en R (dplyr, tidyr).</p> <p>Trabajo grupal: selección, filtrado y limpieza de datos a utilizar en el análisis.</p>	Publicación en el foro de captura de pantalla de resumen de sus datos antes y después de la administración.
8	01-oct	<p>Clase breve (ejemplos): ¿Cómo evalúo mi selección de datos desde una perspectiva inclusiva?</p> <p>Trabajo grupal: autoevaluación grupal de la administración de datos desarrollada respecto a su suficiencia para evaluar el problema de interés y propuestas de mejora para su presentación en el avance.</p> <p>Trabajo inter-grupal: presentación de administración de datos y propuestas de mejora a otro grupo para su retroalimentación. Luego intercambian roles.</p>	<p>Trabajo grupal: avance en presentación de clase 12.</p>	Autoevaluación y evaluación por pares en las dinámicas de grupo.
9	08-oct		<p>Clase: Tablas de datos en R (dplyr, tidyr, kable).</p> <p>Trabajo grupal: tablas de resúmenes de datos del proyecto.</p>	Publicación en el foro de alguna(s) de sus tablas.

10	15-oct	<p>Clase breve (ejemplos): ¿Cómo evaluó mis tablas desde una perspectiva inclusiva?</p> <p>Trabajo grupal: autoevaluación grupal de las tablas desarrolladas respecto a su capacidad de representar el problema de interés y propuestas de mejora para su presentación en el avance.</p> <p>Trabajo inter-grupal: presentación de tablas y propuestas de mejora a otro grupo para su retroalimentación. Luego intercambian roles.</p>	<p>Trabajo grupal: avance en presentación de clase 12.</p>	<p>Autoevaluación y evaluación por pares en las dinámicas de grupo.</p>
11	22-oct		<p>Clase: Visualización de datos en R (ggplot2).</p> <p>Trabajo grupal: análisis gráfico de datos del proyecto.</p>	<p>Publicación en el foro de alguno(s) de sus gráficos.</p>
12	29-oct	<p>Presentaciones de avance de proyecto de investigación.</p>	<p>Trabajo grupal (previo a la actividad sincrónica): preparación de presentación de avance.</p>	<p>Evaluación sumativa: Presentación.</p>
13	05-nov		<p>Clase: Visualización avanzada de datos en R: series temporales (ggplot2, lubridate).</p> <p>Clase: Visualización avanzada de datos en R: mapas (ggplot2, tmap).</p> <p>Trabajo grupal: análisis de datos temporales o espaciales del proyecto.</p>	<p>Publicación en el foro de alguno(s) de sus gráficos temporales o mapas.</p>
14	12-nov	<p>Clase breve (ejemplos): ¿Cómo evaluó mis gráficos desde una perspectiva inclusiva?</p> <p>Trabajo grupal: autoevaluación grupal de los gráficos clásicos, temporales y mapas desarrollados respecto a su capacidad de representar el problema de interés, selección de gráficos más pertinentes y propuestas de mejora para su presentación final.</p>	<p>Trabajo grupal: avance en presentación de clase 17.</p>	<p>Autoevaluación y evaluación por pares en las dinámicas de grupo.</p>

		Trabajo inter-grupal: presentación de gráficos y propuestas de mejora a otro grupo para su retroalimentación. Luego intercambian roles.		
15	19-nov		Clase: visualización dinámica en R: shiny. Trabajo grupal: transformación de uno de los gráficos previamente desarrollados en un visualizador.	Publicación en el foro de su visualizador.
16	26-nov	Clase: Presentación de cierre de curso recapitulando los problemas de investigación tratados por estudiantes y los aciertos y mejoras a lo largo del proceso. Trabajo grupal: reflexión sobre lo expuesto correspondiente a otros grupos, las limitaciones de los análisis y los datos, así como imaginar posibles soluciones. Trabajo inter-grupal: compartir reflexiones y concluir puntos claves a considerar cuando se consume y produce información.		
17	03-dic	Presentaciones de proyectos de investigación.	Trabajo grupal (previo a la actividad sincrónica): preparación de presentación final.	Evaluación final: Presentación.

11. REQUISITOS DE APROBACIÓN

- Calificación final igual o superior a 4,0, en una escala de 1 a 7.
- Asistencia mínima del 65% de las sesiones presenciales.
- Entrega de todas las evaluaciones formativas asincrónicas.

12. RECURSOS DE APRENDIZAJE O BIBLIOGRAFÍA BÁSICA OBLIGATORIA

Wickham, Hadley, Mine Çetinkaya-Rundel, and Garrett Golemund. *R para Ciencia de Datos*. 2023. Traducción de R for Data Science. Disponible en <https://es.r4ds.hadley.nz/>

D'ignazio, Catherine, and Lauren F. Klein. *Feminismo de datos*. 2023. Traducción de Data Feminism. Disponible en <https://data-feminism.mitpress.mit.edu/bienvenida>

13. RECURSOS ADICIONALES O BIBLIOGRAFÍA COMPLEMENTARIA

Abadal, Ernest. *Ciencia abierta: un modelo con piezas por encajar*. Arbor, 197(799), a588. 2021. <https://doi.org/10.3989/arbor.2021.799003>

Carmi, Elinor, and Yates, Simeon J.. *What do digital inclusion and data literacy mean today?* Internet Policy Review 9.2. 2020.

Perez, Caroline Criado. *Invisible women: Data bias in a world designed for men*. Abrams, 2019.

Teranishi, Robert T., et al., eds. *Measuring race: why disaggregating data matters for addressing educational inequality*. Teachers College Press, 2020.