

---

**PROGRAMA DE ESTUDIOS**  
***Reproducibilidad y Comunicación de Resultados***

---

**I. IDENTIFICACIÓN**

Duración: 5 sesiones de 3 horas 30 minutos y 1 sesión un sábado de 5 horas

Profesores: René Lagos Barrios, Amaru Agüero Jiménez, Andrés González Santa Cruz

Ayudante: –

Fechas de clases: Diciembre 2024

Datos de contacto para asuntos administrativos: Alejandra Carmona

(alejandra.carmona@udp.cl)

**II. DESCRIPCIÓN**

Aborda estrategias, formatos y plataformas que buscan asegurar la reproducibilidad y transparencia de análisis de datos, junto con comunicar resultados de manera clara y efectiva a diferentes audiencias.

Este curso está diseñado para enseñar a los/as estudiantes la importancia de la reproducibilidad en la ciencia de datos y cómo aplicar prácticas de reproducibilidad en R. Los/as estudiantes aprenderán a organizar y documentar su código y análisis de datos de manera que otras personas puedan entender y replicar sus resultados. Se enfatizará en el uso de herramientas y técnicas para garantizar la transparencia y fiabilidad de los análisis, como la creación de informes dinámicos con R Markdown y una introducción al control de versiones con Git y máquinas virtuales (*virtual machines*).

**Se requiere conocimientos previos en R y RStudio, además de contar con un computador habilitado para instalar R (4.2+), RStudio, Rtools, GitHub Desktop ([enlace](#)) y Quarto ([enlace](#)).**

**III. OBJETIVOS**

Al finalizar este curso, los/las estudiantes serán capaces de:

- Desarrollar un esquema de trabajo reproducible y replicable.
- Identificar contextos de aplicación de herramientas básicas y generales de reporte y diseminación de resúmenes y análisis de datos.
- Aplicar los conceptos aprendidos sobre arquitectura y modelamiento de datos en un contexto práctico.
- Diseñar un modelo conceptual de datos utilizando diagramas de Entidad-Relación (ER).
- Explicar y argumentar la importancia de la reproducibilidad y la ciencia abierta en proyectos de análisis de datos.
- Estructurar y planificar proyectos de manipulación de datos desde su concepción hasta su implementación.

**IV. DESCRIPCIÓN DE CONTENIDOS**

Durante el módulo se abordarán los conceptos fundamentales de la reproducibilidad en el análisis de datos, distinguiendo entre reproducibilidad y replicabilidad, y se presentarán principios básicos para garantizarla mediante proyectos de datos, así como el concepto de ciencia abierta y sus instituciones promotoras. Se explorará la arquitectura y el modelamiento de datos, destacando su definición, relevancia y componentes clave como bases de datos relacionales y no relacionales, procesos ETL, almacenamiento y acceso. Se profundizará en los fundamentos del modelamiento, incluyendo almacenamiento (warehousing), modelos conceptuales mediante diagramas ER, y modelos lógicos y físicos. Además, se enseñará la creación de informes dinámicos con R Markdown y Quarto, generando documentos en formatos

como HTML y PDF. Se introducirá Shiny para la elaboración de informes altamente interactivos. Finalmente, se explicarán los repositorios y máquinas virtuales, sus ventajas, desventajas y contextos de aplicación, ofreciendo un panorama general de las principales opciones disponibles en el mercado como AWS Cloud, Code Ocean, Google Colab y Saturn Cloud. Por último, se introducirá al sistema de control de versiones mediante una aplicación concreta en la integración de GitHub con RStudio.

## V. IMPORTANCIA DEL CURSO EN EL PLAN DE ESTUDIOS

El curso permitirá desarrollar la capacidad de presentar esquemas de trabajo, resumir datos y exponer análisis de manera efectiva que impacte a la audiencia y estimule la reflexión crítica. En entornos donde se opera con datos cada vez más complejos, es esencial fomentar habilidades y el uso de herramientas que aseguren la transparencia, la replicabilidad y la reproducibilidad de los resultados.

## VI. METODOLOGÍA

El desarrollo del módulo considera las siguientes estrategias metodológicas:

- Presentaciones en-línea y sincrónicas con apoyo de medios audiovisuales de contenido
- Asignación de trabajos grupales
- Talleres prácticos donde los/as estudiantes aprenden a utilizar herramientas de R como R Markdown, knitr para crear análisis reproducibles.
- Sesiones prácticas en las que los/as estudiantes utilizan R Markdown y GitHub para crear informes y presentaciones dinámicas e interactivas.

## VII. FORMAS DE EVALUACIÓN

Dos talleres prácticos desarrollados en grupos de mínimo dos o tres personas, dependiendo del tamaño del curso (**80%**).

- **Taller 1:** Elaboración de un esquema conceptual de arquitectura de datos, aplicado a un caso real o hipotético (**40%**). De 5 a 10 páginas (sin contar portadas y anexos).
  - Presentar el contexto y los objetivos del caso seleccionado.
  - Justificar la relevancia del estudio en el ámbito de las ciencias sociales, principios detrás de la reproducibilidad, replicabilidad y ciencia abierta
  - Identificar y describir componentes clave de la arquitectura de datos
  - Definir un Modelo Conceptual y traducir a un Modelo Lógico, introduciendo consideraciones respecto al almacenamiento de datos
    - Incorporar filtros y definir subconjuntos para estratificar según variables de interés de manera dinámica.
- **Taller 2:** Aplicación de la propuesta en un informe interactivo alojado en un repositorio o máquina virtual (**40%**).
  - Debe contener un elemento gráfico interactivo (i.e., plotly, highchart, altair u obtenida a través de htmlwidgets)
  - Debe contener una tabla personalizable (i.e., DT, kable, rpivotTable u otra utilizando D3.js)
  - Subir y publicar proyecto de R a Github (Github Pages)

El **20%** de la nota considera **participación** e involucramiento directo en las actividades de clases.

- *Todas las evaluaciones deberán ser sometidas en la fecha y hora indicada a través de la plataforma del curso. Considerando una escala de 1.0 a 7.0, la penalización por atraso no justificado es de 0.3 puntos por día sobre la nota obtenida, con un máximo de 7 días de retraso.*
- *Luego de los 7 días el/la estudiante será calificada/o con la nota mínima en dicha evaluación.*
- *Condiciones o reparos adicionales deben ser comunicados a l@s profesores con al menos una semana de anticipación a la evaluación en cuestión.*

## VIII. CRONOGRAMA

Sesión/ Día	Hora	Tema	Lecturas (complementarias)
<b>Martes 3-12</b>	18:00- 21:30	<b>Introducción y repaso de conceptos relevantes (AGS):</b> <ul style="list-style-type: none"> <li>- Definición y relevancia de la reproducibilidad en el análisis de datos.</li> <li>- Principios básicos para garantizar la reproducibilidad en proyectos de datos.</li> <li>- Diferencias entre la reproducibilidad y la replicabilidad</li> <li>- Ciencia abierta: definiciones, alcances, instituciones promotoras.</li> <li>- Protocolos</li> <li>- Libro de códigos y documento de campo</li> <li>- Herramientas: <code>renv()</code>, <code>dateback()</code>, <code>checkpoint()</code>, <code>rang()</code>, <code>packrat()</code>, <code>codebook()</code></li> </ul>	<ul style="list-style-type: none"> <li>- Abbasi, A., Chiang, R. &amp; Xu, J. (2023). Data Science for Social Good. <a href="https://arxiv.org/abs/2311.14683">https://arxiv.org/abs/2311.14683</a></li> <li>- Trisovic, Ana, Matthew K. Lau, Thomas Pasquier, and Mercè Crosas. "A Large-scale Study on Research Code Quality and Execution." <i>Scientific Data</i> 9, no. 1 (2022): 1-16. Accessed November 8, 2024. <a href="https://doi.org/10.1038/s41597-022-01143-6">https://doi.org/10.1038/s41597-022-01143-6</a>.</li> <li>- Bishop, D. (2019). Rein in the four horsemen of irreproducibility. <a href="https://doi.org/10.1038/d41586-019-01307-2">https://doi.org/10.1038/d41586-019-01307-2</a></li> <li>- Simmons, J. P., Nelson, L. D., &amp; Simonsohn, U. (2011). False-Positive Psychology. <i>Psychological Science</i>. <a href="https://doi.org/10.1177/0956797611417632">https://doi.org/10.1177/0956797611417632</a></li> <li>- BMJ (2015, 22 de Septiembre). Introducing "How to write and publish a study protocol" using BMJ's new eLearning programme, Research to Publication. BMJ Blogs. <a href="https://blogs.bmj.com/bmjopen/2015/09/22/introducing-how-to-write-and-publish-a-study-protocol-using-bmjs-new-elearning-programme-research-to-publication/">https://blogs.bmj.com/bmjopen/2015/09/22/introducing-how-to-write-and-publish-a-study-protocol-using-bmjs-new-elearning-programme-research-to-publication/</a></li> <li>- Arslan, R. C. (2019). How to Automatically Document Data With the codebook Package to Facilitate Data Reuse. <i>Advances in Methods and Practices in Psychological Science</i>. <a href="https://doi.org/10.1177/2515245919838783">https://doi.org/10.1177/2515245919838783</a></li> <li>- Figueiredo Filho, D., Lins, R., Domingos, A., Janz, N., &amp; Silva, L.. (2019). Seven Reasons Why: A User's Guide to Transparency and Reproducibility. <i>Brazilian Political Science Review</i>, 13(2), e0001. <a href="https://doi.org/10.1590/1981-3821201900020001">https://doi.org/10.1590/1981-3821201900020001</a></li> </ul>
<b>Jueves 5-12</b>	18:00- 21:30	<b>Arquitectura y Modelamiento de datos (RL):</b> <ul style="list-style-type: none"> <li>- Definición y relevancia</li> <li>- Componentes: bases de datos relacionales y no-relacionales, procesos ETL (preparación BBDD)</li> <li>- Almacenamiento y acceso</li> </ul>	<ul style="list-style-type: none"> <li>- Granger, B. E., &amp; Perez, F. (2021). Jupyter: Thinking and Storytelling With Code and Data. <i>Computing in Science &amp; Engineering</i>, 23(2), 7–14. <a href="https://doi.org/10.1109/MCSE.2021.3059263">https://doi.org/10.1109/MCSE.2021.3059263</a></li> <li>- Giebler, C., Gröger, C., Hoos, E., Schwarz, H., &amp; Mitschang, B. (2019). Leveraging the data lake: Current state and challenges. <i>Big Data Analytics and Knowledge Discovery: 21st International Conference, DaWaK 2019, Linz, Austria, August 26–29, 2019, Proceedings</i> 21, 179–188.</li> <li>- Kimball, R., &amp; Ross, M. (2013). The data warehouse toolkit: The definitive guide to dimensional modeling. John Wiley &amp; Sons. (Capítulo 1)</li> <li>- Kaur, H., &amp; Kaur, G. (2019). Comprehensive Survey of OLAP Models. En N. Yadav, A. Yadav, J. C. Bansal, K. Deep, &amp; J. H. Kim (Eds.), <i>Harmony Search and Nature Inspired Optimization Algorithms</i> (Vol. 741, pp. 415–422). Springer Singapore. <a href="https://doi.org/10.1007/978-981-13-0761-4_40">https://doi.org/10.1007/978-981-13-0761-4_40</a></li> </ul>
<b>Sábado 7-12</b>	09:00- 12:00	<b>Modelamiento (RL):</b> <ul style="list-style-type: none"> <li>- Fundamentos</li> <li>- Almacenamiento (warehousing)</li> <li>- Modelo conceptual (diagrama e identificación ER)</li> <li>- Modelos lógicos y físicos</li> </ul>	
		<b>Taller N°1: Propuesta</b>	<b>Domingo 14 de Diciembre</b>

<b>Martes 10-12</b>	18:00-21:30	<b>Informes de Interactividad Limitada (AA):</b> Creación de informes dinámicos con R Markdown y Quarto Generación de documentos en diferentes formatos (HTML, PDF)	<ul style="list-style-type: none"> <li>- Close, G. (2022). Technical Writing and Publishing Data-Rich Articles with Quarto. Unpublished. <a href="https://doi.org/10.13140/RG.2.2.14862.43846/1">https://doi.org/10.13140/RG.2.2.14862.43846/1</a></li> <li>- Matei, S. A., &amp; Hunter, L. (2021). Data storytelling is not storytelling with data: A framework for storytelling in science communication and data journalism. The Information Society, 37(5), 312–322. <a href="https://doi.org/10.1080/01972243.2021.1951415">https://doi.org/10.1080/01972243.2021.1951415</a></li> </ul>
<b>Jueves 12-12</b>	18:00-21:30	<b>Informes de Interactividad Limitada (continuación) (AA):</b> Creación de informes dinámicos con R Markdown y Quarto Generación de documentos en diferentes formatos (HTML, PDF)	<ul style="list-style-type: none"> <li>- Xie, Y., Allaire, J.J. y G. Golemund. 2019. R Markdown: The definitive guide. CRC Press. <a href="https://bookdown.org/yihui/rmarkdown">https://bookdown.org/yihui/rmarkdown</a></li> </ul>
<b>Martes 17-12</b>	18:00-21:30	<b>Informes de flujo de información dinámico y alta interactividad (AGS):</b> Introducción a Shiny (concepto, estructura de archivos, aplicaciones generales).	<ul style="list-style-type: none"> <li>- Wickham, H., &amp; Safari, a. (2021). Mastering Shiny. O'Reilly Media, Incorporated. <a href="https://mastering-shiny.org/">https://mastering-shiny.org/</a></li> <li>- Laderas, T., Minnier, J., Pozhidayeva, D. &amp; Lo, P. (2021). A gRadual intRoduction to Shiny." <a href="https://laderast.github.io/edu/2021-03-20-a-gradual-introduction-to-shiny">https://laderast.github.io/edu/2021-03-20-a-gradual-introduction-to-shiny</a></li> </ul>
<b>Jueves 19-12</b>	18:00-21:30	<b>Repositorios, máquinas virtuales y contenedores (AGS):</b> Explicación de máquinas virtuales, alojamiento en la nube, ventajas y desventajas y contextos de aplicación, recomendaciones. Panorama general de las principales máquinas virtuales disponibles en el mercado: AWS Cloud, Code Ocean, Google Colab, Saturn Cloud, Rstudio Cloud, Github Actions, entre otras. Diferencias con contenedores (ej., Docker). Mi primer repositorio GitHub <ul style="list-style-type: none"> <li>- Contexto: control de versiones</li> <li>- Flujo de trabajo</li> <li>- Elementos</li> <li>- Acciones</li> <li>- GitHub Pages</li> </ul>	<ul style="list-style-type: none"> <li>- Hsing, C., &amp; Gennarelli, V. (2019). Using GitHub in the Classroom Predicts Student Learning Outcomes and Classroom Experiences: Findings from a Survey of Students and Teachers. In Proceedings of the 50th ACM Technical Symposium on Computer Science Education (pp. 672–678). Association for Computing Machinery</li> <li>- Bolivar, R.G.(2019). Collaborative Learning Using Git with GitLab in Students of the Engineering Programming Course. In Proceedings of the International Congress on Educational and Technology in Sciences 2019, Arequipa, Peru, December 10-12, 2019 (pp. 92–101). CEUR-WS.org.</li> <li>- Ali Awan, D. (2024, 19 de Abril). Tutorial de GitHub y Git para principiantes. DataCamp. <a href="https://www.datacamp.com/es/tutorial/github-and-git-tutorial-for-beginners">https://www.datacamp.com/es/tutorial/github-and-git-tutorial-for-beginners</a></li> <li>- Plaza-Vega, F. (2024). Mini curso: Quarto y GitHub Pages. III Jornadas de Ingeniería Estadística 202, 11 y 12 de Noviembre 2024, Auditorio DMCC Universidad de Santiago de Chile. <a href="https://franplaza.github.io/Quarto-Github/slides/quarto-github-tutorial.html#/title-slide">https://franplaza.github.io/Quarto-Github/slides/quarto-github-tutorial.html#/title-slide</a></li> </ul>
<b>Taller N°2: Aplicación</b>			<b>Domingo 29 de Diciembre</b>

## IX. BIBLIOGRAFÍA OBLIGATORIA

- No hay

## X. CLÁUSULA ÉTICA

La Facultad de Ciencias Sociales, en coherencia con la normativa de la Universidad, establece severas sanciones para las faltas graves que los estudiantes realicen, tales como plagio o falsificación de documentos, señaladas en su reglamento. Son definidas como conductas que vician una evaluación, todos aquellos actos que implican una ruptura o violación de los estándares éticos que un estudiante debe resguardar en una evaluación, distorsionando la apreciación de los conocimientos y habilidades reales que el estudiante posee (por ejemplo, copia y uso indebido de documento y/o medios electrónicos, modificación o falsificación de documentos, plagio en trabajos escritos). Se consideran como fuentes de información que se deben citar, los textos e imágenes en formato electrónico o en papel, en todas sus modalidades: libros, diarios, documentos de trabajo, artículos académicos, etc. Para el caso del curso, se seguirán los canales y procedimientos definidos por la Facultad para enfrentar una situación de este tipo.

## **XI. Normativa sobre Prevención y Sanción de acciones de discriminación, violencia sexual y/o de género (extracto).**

Ésta surge de la necesidad de reducir brechas de desigualdad existentes entre hombres y mujeres, marco en el cual se produce violencia de género en sus diversas formas. Esta normativa tiene por propósito asegurar estándares de convivencia dentro de la institución y en todas aquellas actividades vinculadas a su quehacer, proporcionando mecanismos para enfrentar y sancionar situaciones de violencia sexual y brindar apoyo y orientación a quienes han sido víctimas de ella.

**Artículo 1.** Serán consideradas conductas de violencia sexual todos aquellos comportamientos, palabras o gestos de connotación sexual no consentidos y que atentan contra la dignidad o integridad psicológica y/o física de los miembros de la Universidad.

**Artículo 2.** La institución considerará acciones de violencia sexual, y sin que la enumeración sea taxativa, el acoso sexual, la agresión sexual, la exhibición, exposición y voyerismo sexual y el acecho sexual y, en general cualquier interacción con significado sexual no consentida.

**Artículo 3.** Se denominará acoso u hostigamiento sexual a las insinuaciones y requerimientos de carácter sexual, no deseados y no consentidos que afectan la dignidad e integridad de quien los recibe. El acoso sexual ocurre cuando se sugiere explícita o implícitamente que la aceptación o el rechazo de esas acciones tendrá incidencia en las condiciones laborales y/o académicas de una persona o grupo al interior de la Universidad, o bien cuando la conducta tiene el propósito o el efecto de interferir la convivencia, el rendimiento académico o laboral, creando un ambiente intimidante u hostil para el o los afectados. No es necesaria una acción sistemática de acoso u hostigamiento ya que una sola acción puede constituir acoso sexual.

**Artículo 4.** La agresión sexual es un acto que se comete mediante el uso de la fuerza, la amenaza del uso de la fuerza o cualquier otra forma de coacción o intimidación. En el marco de esta normativa, la agresión sexual incluye, sin que esté limitada, la penetración y el contacto sexual sin consentimiento.

**Artículo 5.** La exhibición, exposición y el voyerismo sexual buscan violar la privacidad de una persona. Incluye conductas como presenciar u observar relaciones o actividades sexuales de otra(s) persona (s); desnudarse parcial o totalmente delante de otros; grabar, fotografiar, y distribuir sin el consentimiento de los involucrados y por uno o más medios (correo electrónico y/o redes sociales) imágenes de personas semidesnudas o desnudas, teniendo relaciones sexuales y/o desarrollando conductas sexuales.

**Artículo 6.** El acecho sexual se define como el conjunto de acciones con connotación sexual, dirigidas hacia una persona sin su consentimiento y que generan temor e inseguridad, causando perturbación emocional significativa. Implica conductas como seguir, observar o vigilar a una persona, presentarse en su lugar de residencia, trabajo o estudio, llamar por teléfono, enviar mensajes, grabaciones, fotos, videos u objetos de connotación sexual.

**Artículo 7.** Las acciones de violencia sexual ocurren sin el consentimiento del afectado. El consentimiento se entiende como el acuerdo o aceptación de participar en una actividad sexual específica. Este se manifiesta de manera inequívoca mediante palabras o actos concluyentes de una persona capaz que está en condiciones de consentir y se encuentra en pleno uso de sus facultades.

El consentimiento para un acto sexual particular no puede interpretarse como consentimiento para actos sexuales posteriores. Del mismo modo, los actos sexuales previos o la existencia de una relación de pareja tampoco constituye por sí misma expresión de consentimiento.

Para más información consultar la página: <https://genero.udp.cl/por-una-cultura-libre-de-discriminacion-y-violencia-2-2/normativa-de-prevencion-y-sancion-de-acciones-de-discriminacion-violencia-sexual-y-o-de-genero/>