

Métodos cuantitativos básicos

Sebastián Garrido de Sierra

Otoño 2020

H. de clases: Lun. y Mier.
Salón: <https://bit.ly/3h0Bhrl>
sebastian.garrido@cide.edu

H. de oficina: Lun. y Mier. 16-17:00 hrs.
Oficina: Piso 2 del LNPP
segasi.mx

¿De qué va este curso?

Éste es el primer curso de la secuencia de métodos en tu respectivo programa en el CIDE: la Licenciatura en Ciencia Política y Relaciones Internacionales, la Maestría o el Doctorado en Ciencia Política, y la Maestría en Periodismo sobre Políticas Públicas.

Su principal objetivo es dotarte de sólidas bases conceptuales y técnicas para que puedas transformar *datos* en *información* y *conocimiento* adecuado para el consumo humano.

Este proceso está compuesto de varias etapas que pueden resumirse gráficamente de la siguiente forma:

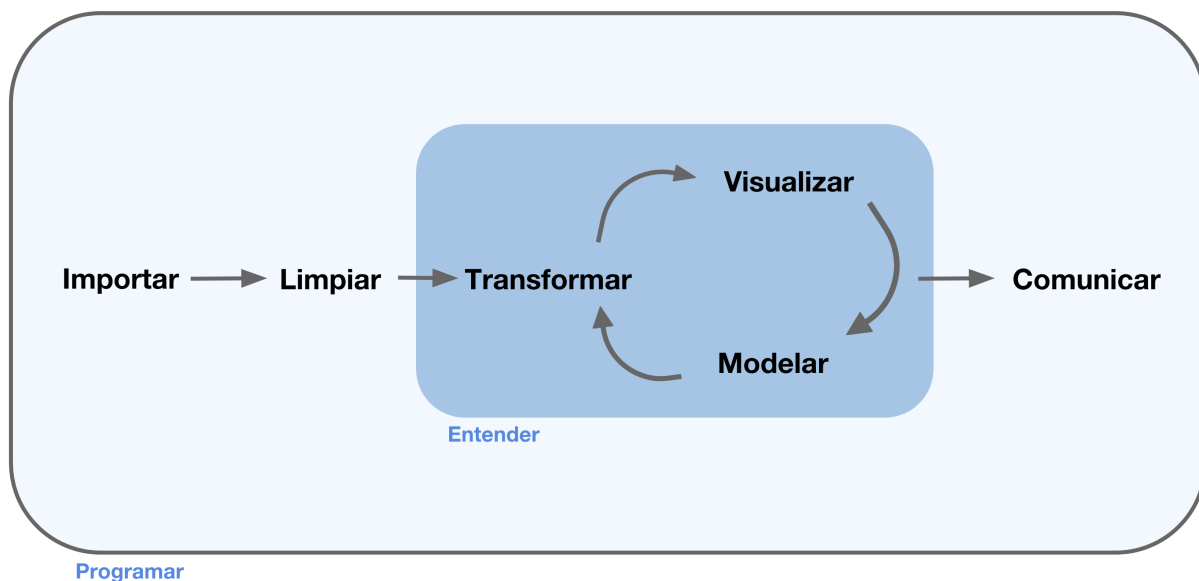


Figura 1: Basado en [Grolemund y Wickham, 2017](#)

Para lograr el objetivo recién planteado, durante las próximas 15 semanas (y cachito) cubriremos diversos temas relacionados con:¹

1. Programación (casi desde el comienzo aprenderán a usar R y RStudio)
2. Manejo de bases de datos (e.g., importar y preparar bases de datos)
3. Visualización de datos (e.g., cómo hacer gráficas eficientes y efectivos)
4. Estadística descriptiva (e.g., medidas de tendencia central y dispersión)

Formato del curso

Nos reuniremos dos veces a la semana, lunes y miércoles, de 9:45 a 11:15 hrs. (grupo 1) o 11:30 a 13:00 hrs. (grupo 2), a través de Zoom (<https://bit.ly/3h0Bhrl>).

Además, las y los alumnos del grupo 1 tendrán una sesión semanal de laboratorio con Helga Jáuregui (helga.jauregui@alumnos.cide.edu), y las y los alumnos del grupo 2 tendrán laboratorio. En ambos casos los laboratorios serán los viernes de 12:00 a 13:00 hrs. Helga y Juvenal les avisará con oportunidad qué plataforma usarán para el laboratorio.

Canales de comunicación

Fuera de las clases, toda la comunicación entre ustedes, Helga, Juvenal y yo la llevaremos a cabo a través de tres canales de comunicación.

- Las **horas de oficina** serán a través de Zoom con la misma liga asignada para las clases. Si deseas reunirte virtualmente conmigo, por favor reserva en [esta liga](#) un espacio de 10 minutos **antes** de que inicien las horas de oficina (Lun. y Miér., de 16 a 17 hrs.).
- Las **preguntas** relacionadas con cualquier tema del curso (R, RStudio, estadística, exámenes, etc.) **deben** ser hechas en la página del curso en Piazza. Si lo prefieres, podrás hacer tus preguntas de forma anónima (solo yo sabré quién la hizo).

Mientras más rápido comiences a plantear tus preguntas por esta vía, más pronto te podrás beneficiar del conocimiento colectivo de tus compañeras y compañeros, los laboratoristas (Helga y Juvenal) y el mío.

Para comenzar a usar Piazza, por favor inscríbete en [esta página](#) **antes** de que concluya la primera semana de clases.

- La gran mayoría de los anuncios relacionados con la clase los haremos a través de Piazza.

¹Si te interesa leer un poco sobre la motivación detrás de esta combinación de temas, te recomiendo leer los siguientes textos de [Nolan y Lang](#), [Cobb](#), [Cobb 1](#) y [American Statistical Association](#)).

- Otros asuntos, incluyendo las comunicaciones que tú consideres que así lo requieran, serán tratados a través de correo electrónico usando **siempre** nuestras cuentas institucionales.

Evaluación

A lo largo del semestre tu desempeño será evaluado **formativa y sumativamente**.

Las **evaluaciones formativas** incluirán breves ejercicios en clase y problemas que deberás resolver en *equipo antes* de cada laboratorio.² Este tipo de evaluaciones tienen como propósito permitir que:

- I. **Tú** entiendas qué tan bien vas con los objetivos de aprendizaje de un tema en particular; y,
- II. **Yo** evalúe qué tan bien he explicado un tema, para decidir si puedo seguir adelante o debo hacer ajustes y explicarlo nuevamente.

Tu desempeño en las evaluaciones formativas **no impactará tu calificación**.

Las **evaluaciones sumativas** serán realizadas a través de tareas y exámenes que deberás responder de manera *individual*. Los propósitos de estos instrumentos son que:

- I. **Tú** tengas una métrica objetiva de en qué medida alcanzaste los objetivos de aprendizaje **previamente** establecidos, y, en consecuencia, qué tan bien dominas un tema;
- II. **Yo** evalúe de forma objetiva en qué medida alcanzaste los objetivos de aprendizaje; y,
- III. **Yo** valore en qué medida las clases fueron exitosas para que tú y tus compañeras/os alcanzaran dichos objetivos.

Después de cada clase compartiré con el grupo los objetivos de aprendizaje del material cubierto ese día. Las tareas y exámenes evaluarán tu dominio de los objetivos correspondientes a cierto número de clases. Antes de cada examen, yo te indicaré los objetivos de aprendizaje que serán evaluados.

Tu calificación final dependerá de tu desempeño en las siguientes actividades:

Actividad	Porcentaje
Tareas	20 %
Primer examen parcial	20 %
Segundo examen parcial	20 %
Tercer examen parcial	20 %

²Cada jueves a las 13 hrs. recibirás un documento con ejercicios, mismos que deberás resolver colectivamente con dos de tus compañeras/os. Nuestro objetivo es que la solución de estos ejercicios no lleve más de 1-1:30 hrs. Los equipos serán generados aleatoriamente por tu laboratorista y cambiarán semanalmente.

Actividad	Porcentaje
Cuarto examen parcial	20 %

Importante: los cuatro exámenes se llevarán a cabo de las 11:30 a las 13 hrs. del día señalado en las siguientes páginas. Por favor, **reserva estos espacios desde ahora.**

URLs importantes

Va una recapitulación de todas las ligas que he mencionado hasta este punto para que las encuentres fácilmente:

- **Zoom** | para clases y horas de oficina | [url](#)
- **Calendly** | para reservar un espacio para horas oficina | [url](#)
- **Piazza** | para hacer todas las preguntas relacionadas con el curso | [url](#)

Lecturas

Una muy buena parte de las lecturas que harán este semestre corresponden a capítulos o secciones de los siguientes cinco libros, todos disponibles en línea de forma **gratuita**:

- Diez, David M, Christopher D. Barr, and Mine Cetinkaya-Rundel (2017). *OpenIntro Statistics*. OpenIntro.
- Golemund, Garrett and Hadley Wickham (2017). *R for Data Science*. Canada: O'Reilly.
- Ismay, Chester and Albert Y. Kim (2018). *Modern Drive. An Introduction to Statistical and Data Sciences via R*.
- Rodrigues, Bruno (2020). *Modern R with the tidyverse*.
- Phillips, Nathaniel D. (2018). *YaRrr! The Pirate's Guide to R*.

Además, al comienzo del semestre les compartiré un folder en línea con el resto de las lecturas.

Calendario

El curso consta de 29 sesiones (26 de clases y tres para exámenes), cada una de 1:30 hrs. A continuación detallo los temas que planeo cubrir en cada una.³

³Por favor toma en cuenta que el calendario del curso puede sufrir cambios dependiendo del progreso del grupo. En caso de que esto sea necesario, se los haré saber de forma anticipada.

Introducción

Semana 01 - 08/17: ¿Por qué están aquí?

- Cap. 2, pp. 47-73 - Silver, Nate (2002). *The Signal and the Noise: Why So Many Predictions Fail-but Some Don't*. Penguin Press.
- Tversky, Amos and Daniel Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases". In: *Science* 185.4157, pp.~1124-1131.)

Bloque 1: R y RStudio

Semana 01 - 08/19: R, RStudio y las bases para usarlos

- Capítulo 1 de Rodrigues, Bruno (2020). *Modern R with the tidyverse*.
- Secciones 2.2, 2.3 y 4.4 de Phillips, Nathaniel D. (2018). *YaRrr! The Pirate's Guide to R*.
- Capítulo 1 de Ismay, Chester and Albert Y. Kim (2018). *Modern Drive. An Introduction to Statistical and Data Sciences via R*.

Semana 02 - 08/24: Tipos de datos, tipos de estructuras de datos y tipos de variables

- Secciones 1.2.1 y 1.2.2 de Diez, David M, Christopher D. Barr, and Mine Cetinkaya-Rundel (2017). *OpenIntro Statistics*. OpenIntro.
- Capítulo 2 de Munzner, Tamara (2014). *Visualization Analysis and Design*. Nueva York, EEUU: CRC Press.

Semana 02 - 08/26: Tipos de variables en R

- Secciones 2.1 a 2.5 de Rodrigues, Bruno (2020). *Modern R with the tidyverse*.

Semana 03 - 08/31: Tipos de estructuras de datos en R

- Secciones 2.6 a 2.8 de Rodrigues, Bruno (2020). *Modern R with the tidyverse*.
- Secciones 5.2, 8.1, 8.2 y 8.3 de Phillips, Nathaniel D. (2018). *YaRrr! The Pirate's Guide to R*.

Semana 03 - 09/02: Primer examen parcial

Bloque 2: Importar y manejar datos

Semana 04 - 09/07: ¿Cómo importar bases de datos a R? + Proyectos

Importar datos

- Capítulo 11 de Grolemund, Garrett and Hadley Wickham (2017). *R for Data Science*. Canada: O'Reilly.
- Broman, Karl W. and Kara H. Woo (2018). "Data organization in spreadsheets". In: *The American Statistician* 72.1, pp.~2-10.
- Wickham, Hadley (2014). "Tidy Data". In: *Journal of Statistical Software* 59.10.

Proyectos en RStudio

- Capítulo 8 de Grolemund, Garrett and Hadley Wickham (2017). *R for Data Science*. Canada: O'Reilly.
- Sección 9.3 de Phillips, Nathaniel D. (2018). *YaRrr! The Pirate's Guide to R*.

Semana 04 - 09/09: Manejo de datos con `dplyr`, parte 1

- Secciones 5.1, 5.2, 5.3 y 5.4 de Grolemund, Garrett and Hadley Wickham (2017). *R for Data Science*. Canada: O'Reilly.

Semana 05 - 09/14: Manejo de datos con `dplyr`, parte 2

- Secciones 5.5, 5.6 y 5.7 de Grolemund, Garrett and Hadley Wickham (2017). *R for Data Science*. Canada: O'Reilly.

Semana 05 - 09/16: No hay clase

Semana 06 - 09/21: Manejo de datos con `dplyr`, parte 3

Semana 06 - 09/23: ¿Cómo unir bases de datos?

- Capítulo 13 de Grolemund, Garrett and Hadley Wickham (2017). *R for Data Science*. Canada: O'Reilly.

Semana 07 - 09/28: Fechas, esos animalitos extraños

- Capítulo 16 de Grolemund, Garrett and Hadley Wickham (2017). *R for Data Science*. Canada: O'Reilly.

Semana 07 - 09/30: Cadenas de texto

- Capítulo 14 de Grolemund, Garrett and Hadley Wickham (2017). *R for Data Science*. Canada: O'Reilly.

Semana 08 - 10/05: Segundo examen parcial

Bloque 3: Visualización de datos

Semana 08 - 10/07: Bases teóricas 1: Vista y percepción

- Capítulo 1 de Cairo, Alberto (2012). *The Functional Art: An Introduction to Information Graphics and Visualization*. California, EEUU: New Riders.

Semana 09 - 10/12: Bases teóricas 2: Marcas y canales

- Capítulo 5 de Munzner, Tamara (2014). *Visualization Analysis and Design*. Nueva York, EEUU: CRC Press.

Semana 09 - 10/14: ggplot2, parte 1

- Capítulo 3 de Grolemund, Garrett and Hadley Wickham (2017). *R for Data Science*. Canada: O'Reilly.
- Capítulo 2 de Ismay, Chester and Albert Y. Kim (2018). *Modern Drive. An Introduction to Statistical and Data Sciences via R*.

Semana 10 - 10/19: ggplot2, parte 2

- Capítulo 28 de Grolemund, Garrett and Hadley Wickham (2017). *R for Data Science*. Canada: O'Reilly.

Semana 10 - 10/21: ggplot2, parte 3

Semana 11 - 10/26: ggplot2, parte 4

Semana 11 - 10/28: Factores

- Capítulo 15 de Grolemund, Garrett and Hadley Wickham (2017). *R for Data Science*. Canada: O'Reilly.

Semana 12 - 11/02: No hay clase

Semana 12 - 11/04: Tercer examen parcial

Estadística descriptiva

Semana 13 - 11/09: Intro + distribuciones univariadas

- Secciones 1.6.2 y 1.6.3 de Diez, David M, Christopher D. Barr, and Mine Cetinkaya-Rundel (2017). *OpenIntro Statistics*. OpenIntro.

Semana 13 - 11/11: Valores (a)típicos y faltantes

- Secciones 3.1 y 3.2 de Agresti, Alan and Barbara Finlay (2009). *Statistical Methods for the Social Sciences*. 4th ed. Pearson.
- Sección 7.4 de Grolemund, Garrett and Hadley Wickham (2017). *R for Data Science*. Canada: O'Reilly.
- Apéndice A de Ismay, Chester and Albert Y. Kim (2018). *Modern Drive. An Introduction to Statistical and Data Sciences via R*.

Semana 14 - 11/16: No hay clase

Semana 14 - 11/18: Medidas de dispersión y posición

- Secciones 3.3 y 3.4 de Agresti, Alan and Barbara Finlay (2009). *Statistical Methods for the Social Sciences*. 4th ed. Pearson.
- Sección 1.6.4 de Diez, David M, Christopher D. Barr, and Mine Cetinkaya-Rundel (2017). *OpenIntro Statistics*. OpenIntro.
- Sección 7.3 de Grolemund, Garrett and Hadley Wickham (2017). *R for Data Science*. Canada: O'Reilly.

Semana 15 - 11/23: Análisis bivariado

- Sección 7.5 de Grolemund, Garrett and Hadley Wickham (2017). *R for Data Science*. Canada: O'Reilly.
- Sección 1.7 de Diez, David M, Christopher D. Barr, and Mine Cetinkaya-Rundel (2017). *OpenIntro Statistics*. OpenIntro.

Semana 15 - 11/25: Análisis de clusters

Lecturas por definir

Semana 16 - 11/30: Estudio de caso

Semana 16 - 12/02: Problemas comunes y cómo lidiar con ellos

Lecturas por definir

Semana 17 - 12/07: Cuarto examen parcial