



Doctorado en  
Políticas Públicas

**Visualización y  
Manejo de datos**

Clases: Jueves, 10-13 hrs.  
Horas de oficina: Miér., 11-13 hrs.

Sebastián Garrido de Sierra  
sebastian.garrido@cide.edu

## Introducción

Este curso está diseñado para fortalecer tus habilidades en dos áreas cada vez más importantes en el mercado laboral y que pocas veces se enseñan en un curso de licenciatura o posgrado: la **visualización** y el **manejo** de datos.

La **visualización de datos** es el área del conocimiento que combina principios de diseño, teoría del color y nuestras capacidades perceptivas para elaborar gráficas y mapas que transmiten de forma eficiente y efectiva los resultados sustantivos derivados del análisis de bases de datos.

En el contexto de esta clase, el **manejo de datos** incluye la construcción, limpieza y transformación<sup>1</sup> de bases de datos para su posterior uso. Dado que es inevitable que tengamos que dedicar una enorme cantidad de tiempo a estas tres tareas antes de poder analizar y/o visualizar nuestros datos, es crucial aprender principios y herramientas que nos permitan realizarlas de forma eficiente, ordenada y replicable/auditable.

## Objetivos

A lo largo de este curso los alumnos aprenderán, entre otras cosas:

1. Cómo funcionan y se relacionan el sistema de la vista, nuestras percepciones y la memoria;

---

<sup>1</sup> En términos estrictos, el manejo de datos sólo incluye la transformación de los mismos. Sin embargo, en este curso aprenderás principios teóricos y habilidades prácticas para construir y limpiar bases de datos.

2. Principios teóricos básicos para visualizar datos de forma efectiva y eficiente;
3. Diversas técnicas y sistemas de visualización, así como cuándo es conveniente utilizar cada uno de ellos;
4. Cómo diseñar visualizaciones a través de un proceso estructurado, incorporando los principios de la percepción humana; y,
5. Cómo generar visualizaciones estáticas e interactiva utilizando diversos programas y plataformas web. La lista incluye, entre otros, *R*, Excel, LibreOffice, Tableau, CartoDB y Gephi.
6. Principios básicos para el diseño y construcción de bases de datos;
7. Cómo usar herramientas para diagnosticar si una base de datos tiene problemas y cómo corregirlos;
8. Qué es *R*, *RStudio* y los conocimientos básicos para utilizar estos programas;
9. Cómo manejar y transformar una base de datos en *RStudio* para, entre otras cosas:
  - a. Crear nuevas variables construidas a partir de variables existentes;
  - b. Renombrar variables;
  - c. Seleccionar un subconjunto de renglones o columnas;
  - d. Unir dos o más bases de datos;
  - e. Calcular estadísticas descriptivas; y,
  - f. Un largo etcétera...

## **Público objetivo**

El curso está diseñado para personas con o sin experiencia en la manipulación de bases de datos y/o en la elaboración de visualizaciones. Los únicos prerrequisitos son:

1. Que los asistentes tengan conocimientos básicos de estadística;

2. Que sepan utilizar hojas de cálculo (Excel, LibreOffice, etc.);
3. Que estén dispuestos a aprender a programar en *R* y que hayan tomado este curso (<http://bit.ly/1FORUxq>) en línea gratuito **antes** de que comencemos a trabajar con *R* y *R Studio* en la sesión 8 (marzo 23).
4. Que hayan descargado e instalado R Studio (<http://bit.ly/1Hllr0q>), un entorno de desarrollo integrado que nos facilitará (dentro de lo que cabe) la vida al usar *R*.

## Evaluación

Dada la naturaleza de este curso, la práctica cotidiana es clave. Por ello, tendrán al menos una tarea a la semana. Estas tareas aportarán el **50%** de su calificación semestral. El resto de su calificación (**50%**) dependerá de su trabajo final. En la sexta semana del curso (marzo 9) les daré los lineamientos para este trabajo.

## Sesiones

A lo largo del semestre tendremos 15 sesiones. Nos reuniremos cada jueves de 10 a 13 horas en el Teatro de Decisiones del Laboratorio Nacional de Políticas Públicas (LNPP), mismo que está ubicado en el tercer piso de la biblioteca. En caso de que el Teatro sea requerido para algún evento del LNPP, la clase será en alguno de los salones del Doctorado de Políticas Públicas. Si esto ocurriera, se los notificaré con anticipación.

La mayoría de las 15 sesiones estarán divididas en dos partes. En la primera discutiremos aspectos teóricos del tema en cuestión y en la segunda los estudiantes aprenderán a usar diversas herramientas.

A continuación enlisto los temas que cubriremos en cada sesión. Mientras que las sesiones en **rojo** corresponden a clases en las que abordaremos temas relacionados con la **visualización de datos**, en las sesiones en **azul** cubriremos uno o más temas relacionados con el **manejo de datos** y *RStudio*.

### **Sesión 1** (feb. 2)

- Introducción al curso
- Introducción a la visualización de datos

### **Sesión 2** (feb. 9)

- Entendiendo al ojo y el “cerebro visual”
- Reconociendo al recordar

### **Sesión 3** (feb. 16)

- ¿Qué?
- ¿Por qué?

### **Sesión 4** (feb. 23)

- ¿Cómo?
- Organización gráfica I: tablas

### **Sesión 5** (mar. 2)

- Organización gráfica II: redes
- Organización gráfica III: mapas

### **Sesión 6** (mar. 9)

- Principios de diseño
- Consejos prácticos

### **Sesión 7** (mar. 16)

- Introducción a la segunda parte del curso
- Construcción de bases de datos

### Sesión 8 (mar. 23)

- Limpieza de bases de datos
- Introducción a *R* y *RStudio*

### Sesión 9 (mar. 30)

- Visualización de datos en *R*, parte I: la gramática de las gráficas

### Sesión 10 (abr. 6)

- Manejo y transformación de datos, parte I

### Sesión 11 (abr. 20)

- Manejo y transformación de datos, parte II

### Sesión 12 (abr. 27)

- Reordenando bases de datos
- Unión y división de bases de datos

### Sesión 13 (may. 4)

- Visualización de datos en *R*, parte II: los detalles
- Variables de texto y tiempo

Para las **sesiones 14** (may. 11) y **15** (may. 18), les pediré que voten por dos de los siguientes temas:

- Gráficas interactivas en *R*
- Mapas en *R*
- Extracción de datos de páginas web con *R* (scraping)
- R Markdown

Cubriré las dos opciones ganadoras, una en cada sesión.