

Visualización de datos longitudinales en R

Elisa Salas (esalas1@uc.cl), Isidora Didier
(isidora.didier@ug.uchile.cl) y Cristóbal Ortiz
(cristobal.ortiz.v@ug.uchile.cl)

ELSOC, COES

20/7/2021

Contenidos

En este curso veremos:

Bloque 1

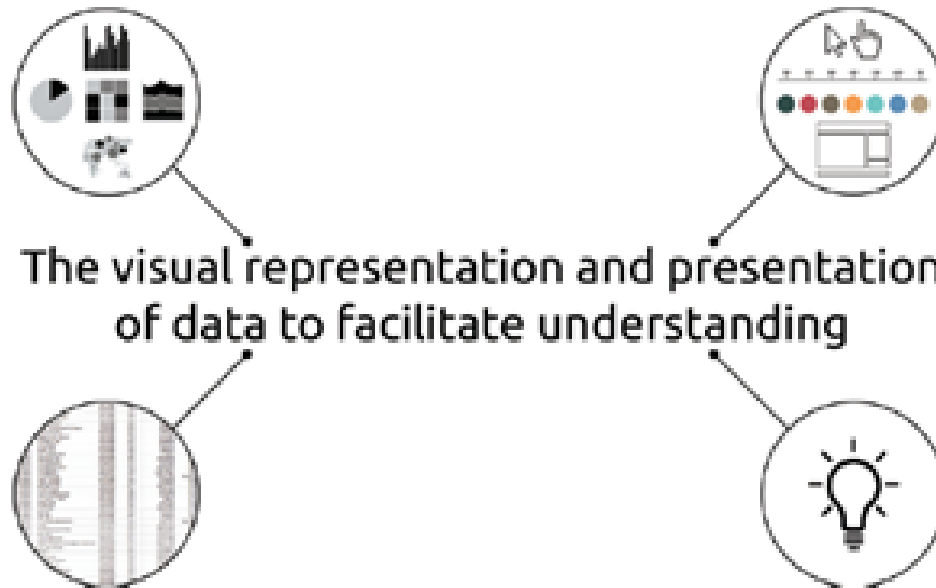
1. ¿Por qué visualizamos datos?: Comprender la lógica de la visualización de datos longitudinales.
2. Tratamiento de base de datos: Preparar y depurar una base de datos longitudinales para el análisis.

Bloque 2

1. ¡A visualizar!: Graficar datos longitudinales en base a muestras complejas.
 - Gráficos de Barra
 - Gráficos de Puntos/Líneas
 - Gráficos Alluvial

1. ¿Qué es la visualización de datos?

- “En términos simples, se trata de gráficos y el acto de seleccionar el gráfico correcto para mostrar las características de los datos que se cree que son más relevantes.” (Kirk, 2018, p.17)
- Se debe ser fiel a los datos, por lo que es muy importante un **buen tratamiento de base de datos**.
- Por ende, los gráficos tienen que estar a merced de los datos.



1.1. Fases de la Visualización de Datos.



The Four Stages of the Data Visualisation Design Process. Fuente: Kirk (2018)

1. **Formula tu plan de trabajo:** planifica, define e inicia tu proyecto.
2. **Trabajando con la base de datos:** produce, maneja y prepara tus datos.
3. **Establece tu pensamiento editorial:** define qué le mostrarás a tu audiencia.

1.2. Tablas VS Gráficos

¿Por qué no ambos?

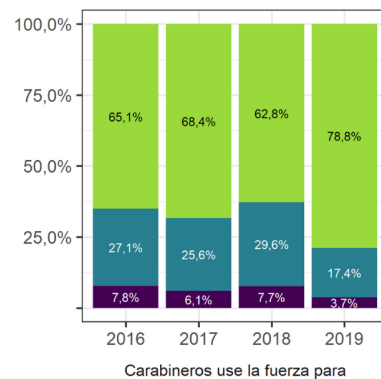
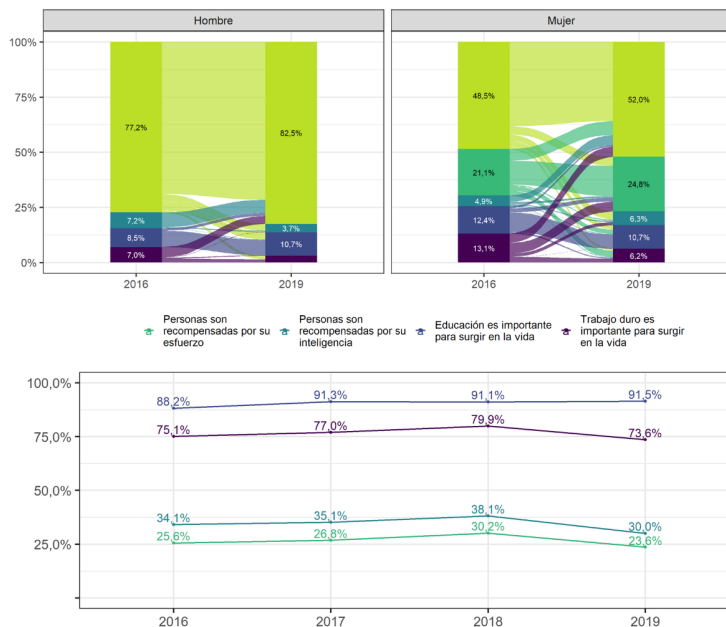
- Ambas presentan objetivos totalmente distintos. Dependerá del contexto la elección de una de ellas.
- Cuando hacemos una tabla, hay que preguntarse ¿Cuánto tiempo tardé en encontrar los elementos importantes en ésta?

Bondades de los gráficos

- Facilita el **entendimiento** de los datos, lo cual fomenta abrir la ciencia.
- Tiene la capacidad de **resumir** datos, mostrando los elementos más relevantes.
- Es **atractivo** y capta mejor la atención de los y las lectoras.

1.3. Visualizando datos longitudinales ELSOC

- Difícil representación de una gran cantidad de datos en tablas debido a la temporalidad de éstos.
- Un adelanto de lo que veremos hoy:



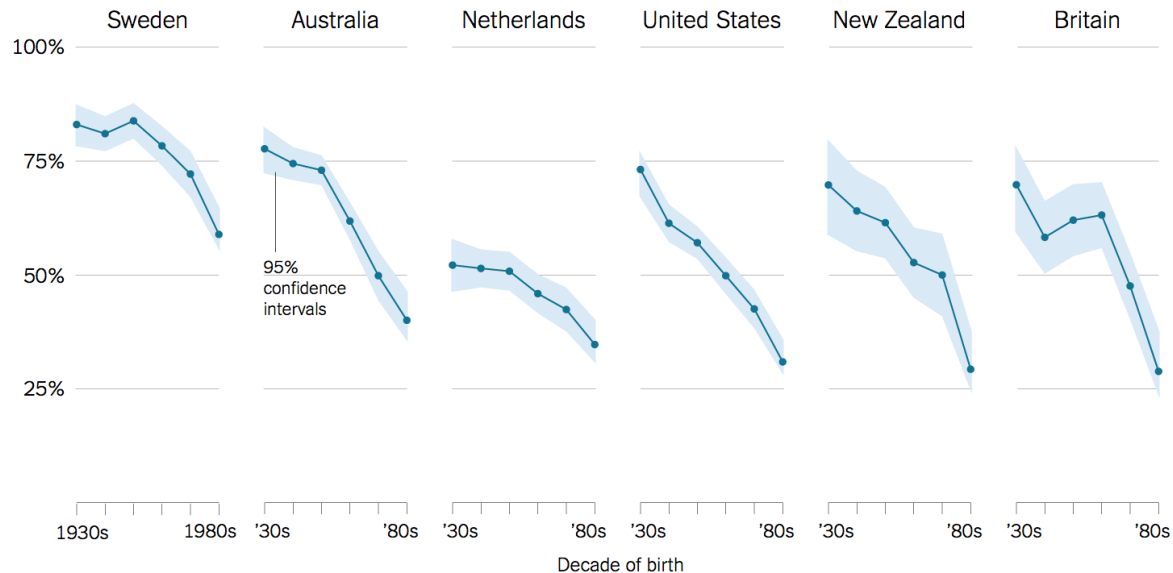
ELSOC (2019). Resultados longitudinales 2016-2019

1.4. Cómo **NO** visualizar datos.

- Más allá de lo estético, el error más grande que se puede cometer es la **mala representación de los datos**, lo que probablemente se debe a un mal manejo de la base de datos
- Un ejemplo de esto es el gráfico ¿Crisis de fe en la democracia? (New York Times), que veremos a continuación.

¿Crisis de la fe en la democracia?

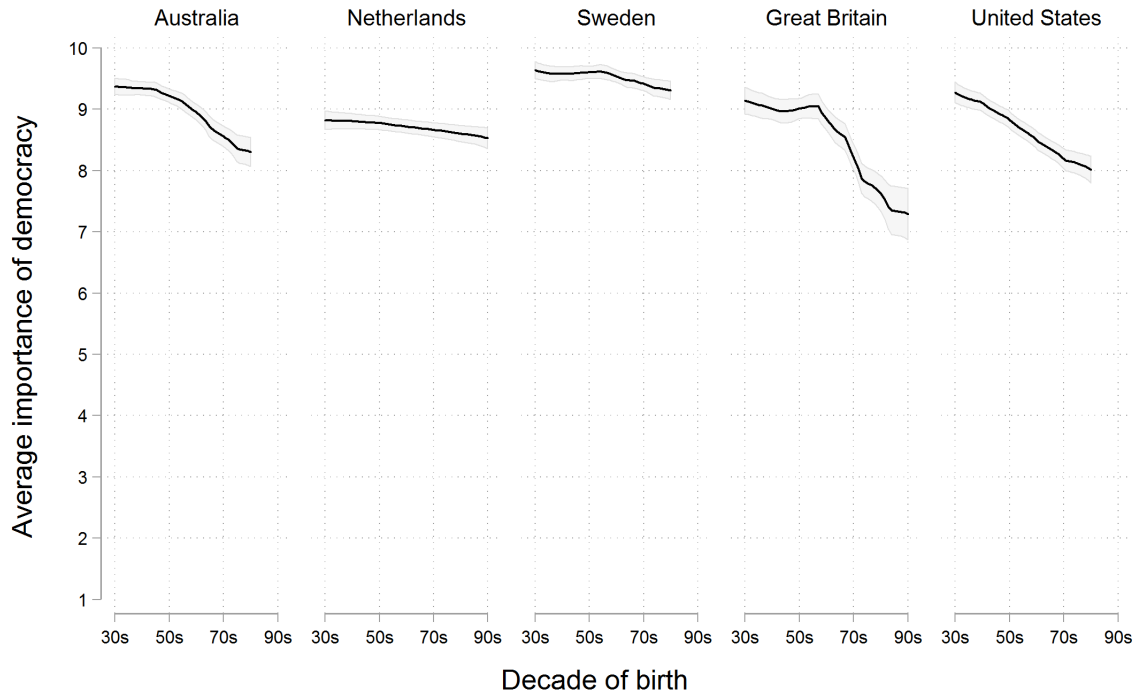
Percentage of people who say it is “essential” to live in a democracy



Source: Yascha Mounk and Roberto Stefan Foa, “The Signs of Democratic Deconsolidation,” *Journal of Democracy* | By The New York Times

A crisis of faith in democracy? (New York Times). Fuente: Healy (2018)

Quizás no tanto...



Graph by Erik Voeten, based on WVS 5

Perhaps the crisis has been overblown. (Erik Voeten). Fuente: Healy (2018)

2. Uso de ggplot2 y tidyverse

- Para la visualización de datos se utiliza el paquete `ggplot2`, el cual pertenece a `tidyverse`. Al igual que otros paquetes de R, presenta distintas funciones que van desde el manejo de los datos hasta la estética en los gráficos.

```
ggplot (data = <DATA> ) +  
  <GEOM_FUNCTION> (mapping = aes( <MAPPINGS> ),  
    stat = <STAT>, position = <POSITION> ) +  
  <COORDINATE_FUNCTION> +  
  <FACET_FUNCTION> +  
  <SCALE_FUNCTION> +  
  <THEME_FUNCTION>
```

required

Not required,
sensible
defaults
supplied

Fuente: Cheat Sheet ggplot2

2.1. Componentes ggplot2

- Data [`data`]: Es la base donde se encontrarán los datos para la creación de los gráficos
- Geometries [`geoms`]: Configura los elementos visuales de los gráficos. Puede modificar datos estadísticos y estética.
- Aesthetics [`aes`]: Se encarga de la estética del gráfico. Se puede cambiar lo colores, tamaños y formas. También, es posible hacer agrupaciones y editar la posición (x, y).
- Stats [`stat`]: Se utiliza para hacer transformaciones estadísticas que nos permite comprender los datos.

- Position [**Position**]: Los ajustes de posición determinan cómo organizar [**geoms**].
- Coordinate systems [**coord**]: Modifica los ejes x e y . Si es que este no es modificado, por defecto se genera el plano cartesiano.
- Facetting [**facet**]: Sive para realizar conjuntos o sub - conjuntos de datos.
- Scale [**scale**]: Transforma valores de la base de datos a valores visuales con su respectiva estética.
- Themes [**theme**]: Controla la visualización de todos los elementos gráficos, a excepción de los datos.

¡Gracias!

1. Healy, K. (2018). *Data Visualization. A practical introduction.*
<https://socviz.co/index.html>
2. Kirk, A. (2018). *Data Visualisation. A Handbook for Data Driven Design.*