Aprendiendo Metodología x Vale

Taller de Visualización de Datos Longitudinales y Panel con RStudio

Cristóbal Ortiz Viches
Asistente de Datos, OLES

24 de enero del 2022

Contenidos del curso

Bloque teórico

1. Lógica de la visualización de datos longitudinales
2. Flujo de trabajo para visualización
    - flujo clásico
    - flujo tidyverse

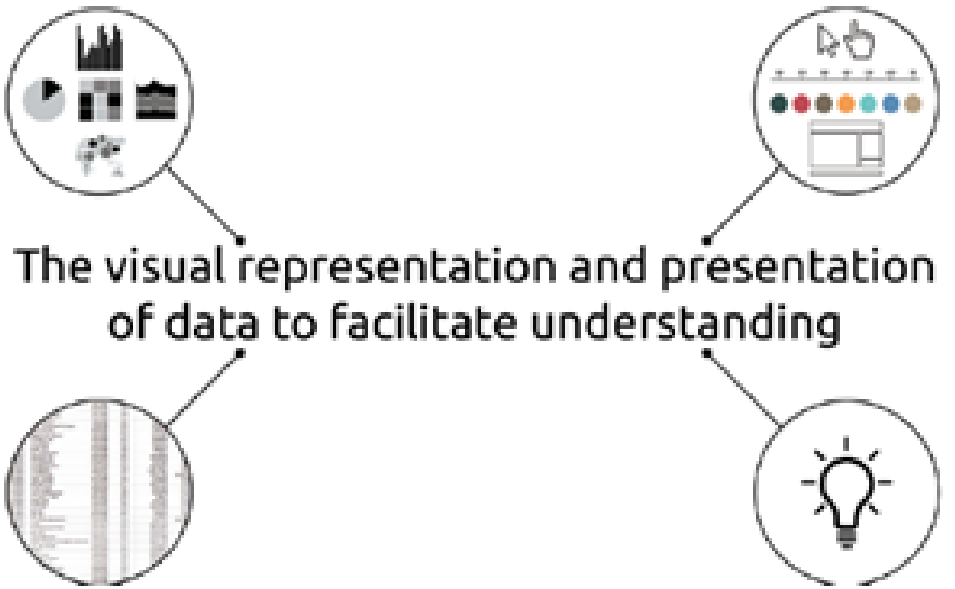
Bloque práctico

1. Visualización de datos con ggplot2
    - Gráfico de barra
    - Gráfico de barra apilada
    - Gráfico de puntos
    - Gráfico alluvial

I. Lógica de la visualización de datos longitudinales

1.1. ¿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**.



1.2. 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.

1.3. Ventajas y desventajas de los gráficos

Ventajas 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.

Desventajas de los gráficos

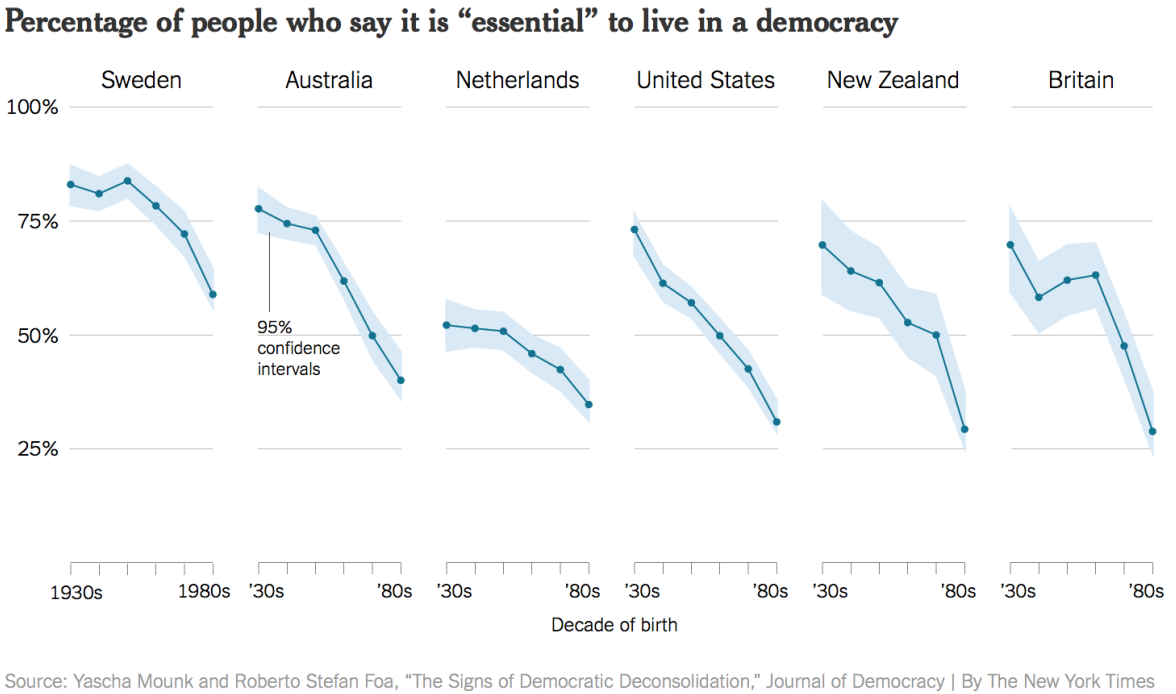
- Limitaciones visuales a medida que crece la cantidad de variables que quiero representar

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

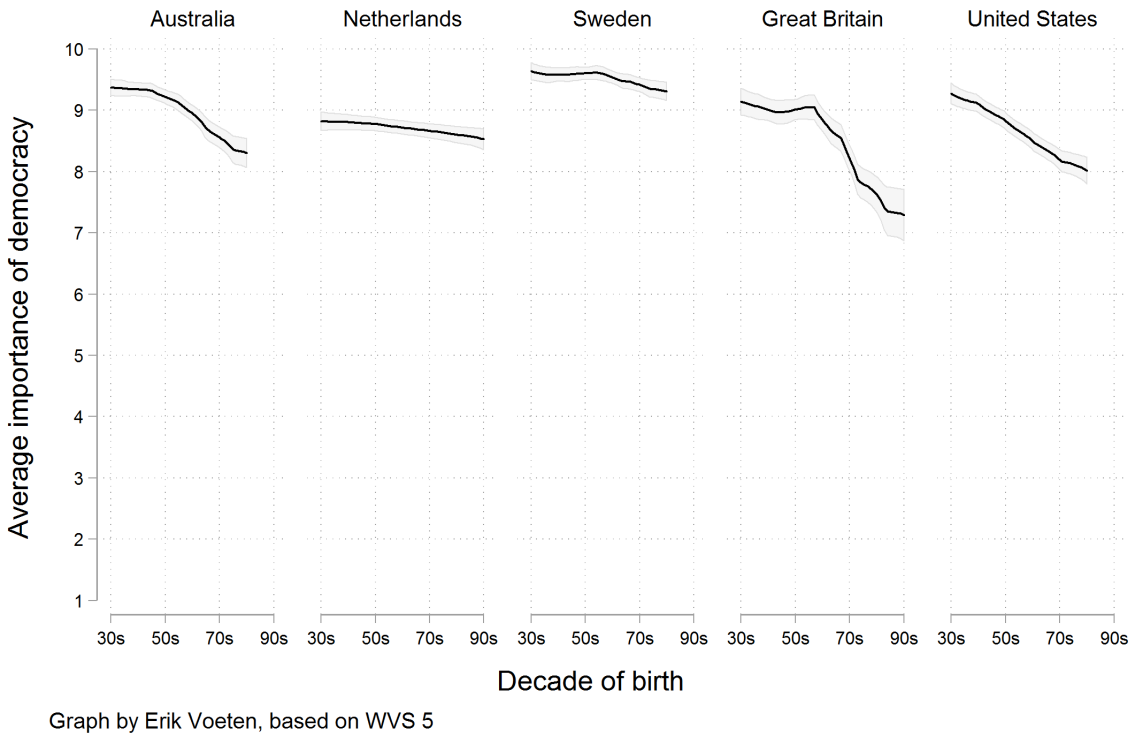
- Un ejemplo de esto es el gráfico ¿Crisis de fe en la democracia? (New York Times), que veremos a continuación.

¿Crisis de la fe en la democracia?



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

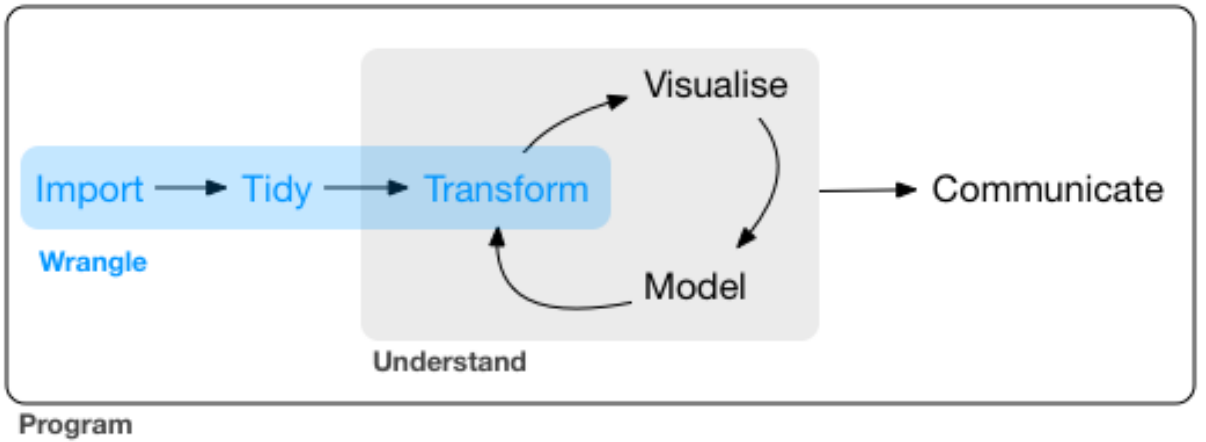
Quizás no tanto...



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

II. Flujo de trabajo para visualizar datos

2.1. Flujo tidyverse

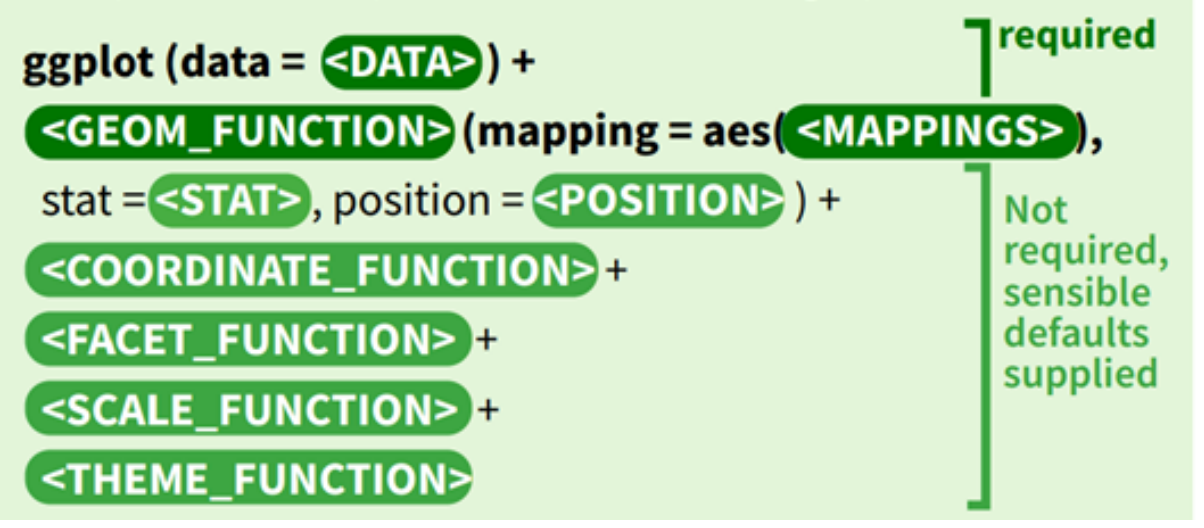


Flujo de trabajo de datos Tidyverse. Fuente: Wickham & Grolemund (2016)

- **Flujo clásico:** separación entre *wrangle* y *visualise*.
  - Un código para preparar los datos y otro para analizarlos.
- **Flujo tidyverse:** orientado a la comunicación de datos
  - herramientas tidyverse permiten juntar en un mismo código *wrangle* y *visualise* (o preparación y análisis).

. 2.2. Visualización con ggplot

- 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**.



Fuente: Cheat Sheet ggplot2

2.3. 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.

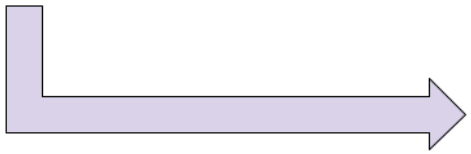
2.4. Preámbulo del bloque práctico

Pasos:

1. cargar librerías y dataset
2. limpieza dataset
3. transformar datos de wide a long (¿por qué?)
4. generar tabla con datos a visualizar
5. encadenar tabla con funciones de `ggplot2`.

Paso de wide a long

ID	Edad 2016 (m0_edad_w01)	Edad 2017 (m0_edad_w02)	Edad 2018 (m0_edad_w03)
1 (Pedro)	15	16	17
2 (Juan)	67	68	69
3 (Diego)	44	45	46



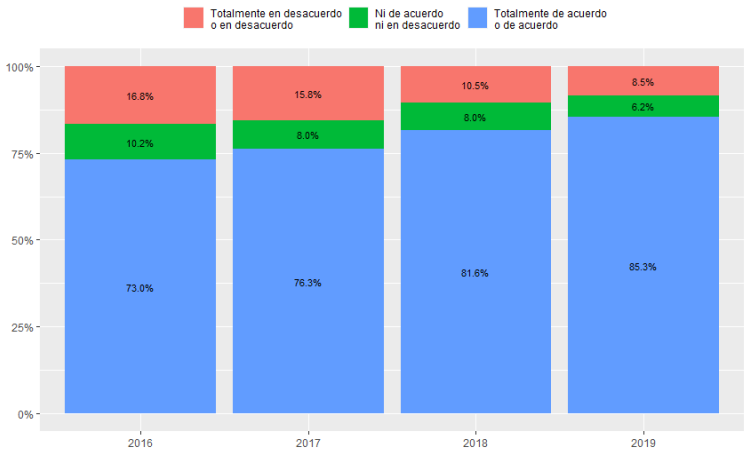
ID	Ola (w0_)	Edad (m0_edad)
1 (Pedro)	2016	15
1 (Pedro)	2017	16
1 (Pedro)	2018	17
2 (Juan)	2016	67
2 (Juan)	2017	68
2 (Juan)	2018	69
3 (Diego)	2016	44
3 (Diego)	2017	45
3 (Diego)	2018	46

Fuente: Elaboración propia.

Encadenar datos y la visualización con ggplot

ola	c10_03	n	freq
2016	Totalmente en desacuerdo o en desacuerdo	316	0,16808511
2016	Ni de acuerdo ni en desacuerdo	192	0,10212766
2016	Totalmente de acuerdo o de acuerdo	1372	0,72978723
2017	Totalmente en desacuerdo o en desacuerdo	296	0,15761448
2017	Ni de acuerdo ni en desacuerdo	150	0,0798722
2017	Totalmente de acuerdo o de acuerdo	1432	0,76251331
2018	Totalmente en desacuerdo o en desacuerdo	197	0,10473153
2018	Ni de acuerdo ni en desacuerdo	150	0,07974482
2018	Totalmente de acuerdo o de acuerdo	1534	0,81552366
2019	Totalmente en desacuerdo o en desacuerdo	159	0,08470964
2019	Ni de acuerdo ni en desacuerdo	117	0,06233351
2019	Totalmente de acuerdo o de acuerdo	1601	0,85295685

%>%



Fuente: Elaboración propia en base a datos ELSOC (2021).

III. Referencias

1. Healy, K. (2018). Data visualization: a practical introduction. Princeton University Press. https://socviz.co/index.html/

2. Kirk, A. (2016). Data visualisation: A handbook for data driven design. Sage. https://book.visualisingdata.com/

3. Wickham, H., & Grolemund, G. (2016). R for data science: import, tidy, transform, visualize, and model data. O'Reilly Media, Inc. https://es.r4ds.hadley.nz/

4. [ELSOC] Reproducible Research, Centre for Social Conflict and Cohesion Studies COES. (2021). Estudio Longitudinal Social de Chile 2016-2019 [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/SOQJ0N

Gracias por su atención!

cristobal.ortiz.v@ug.uchile.cl