Ciencia de datos en R

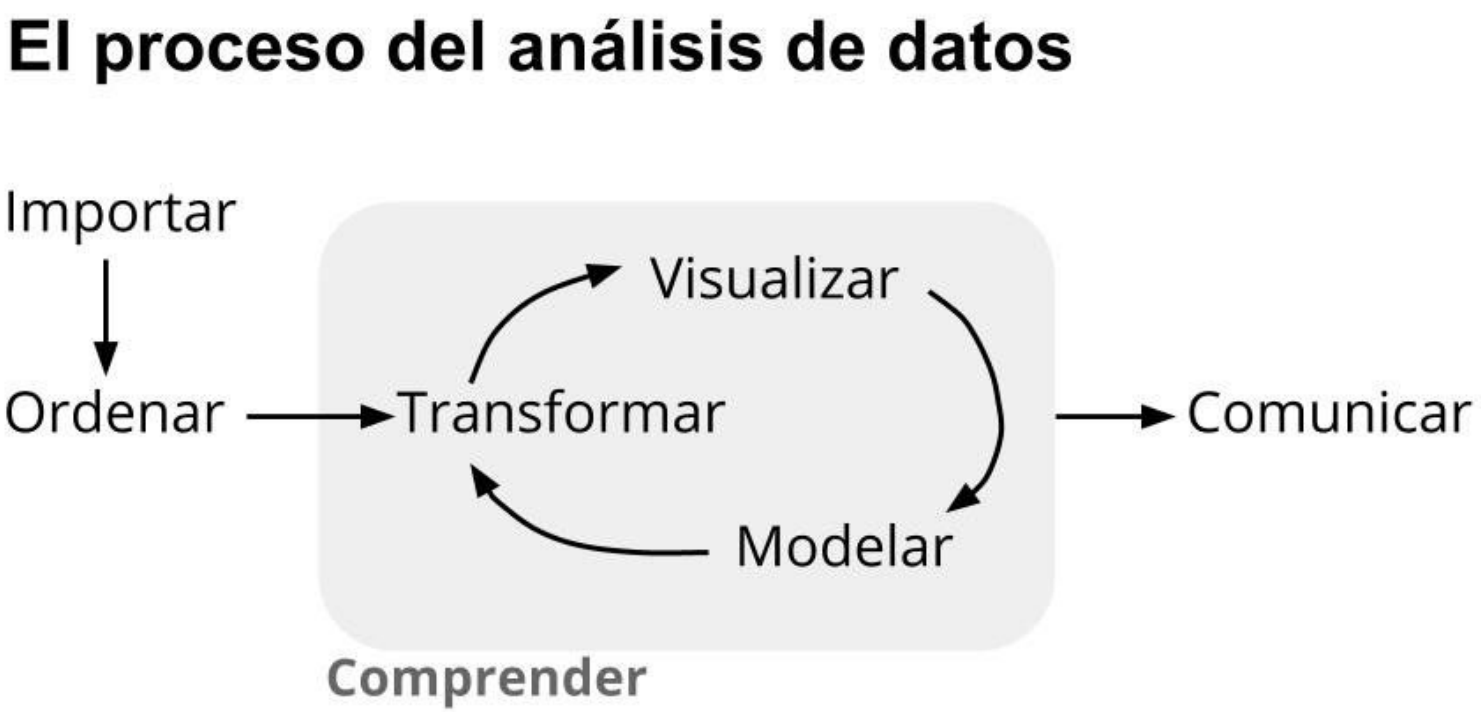
Big Data: Marco conceptual, técnicas y aplicaciones

Clase 3

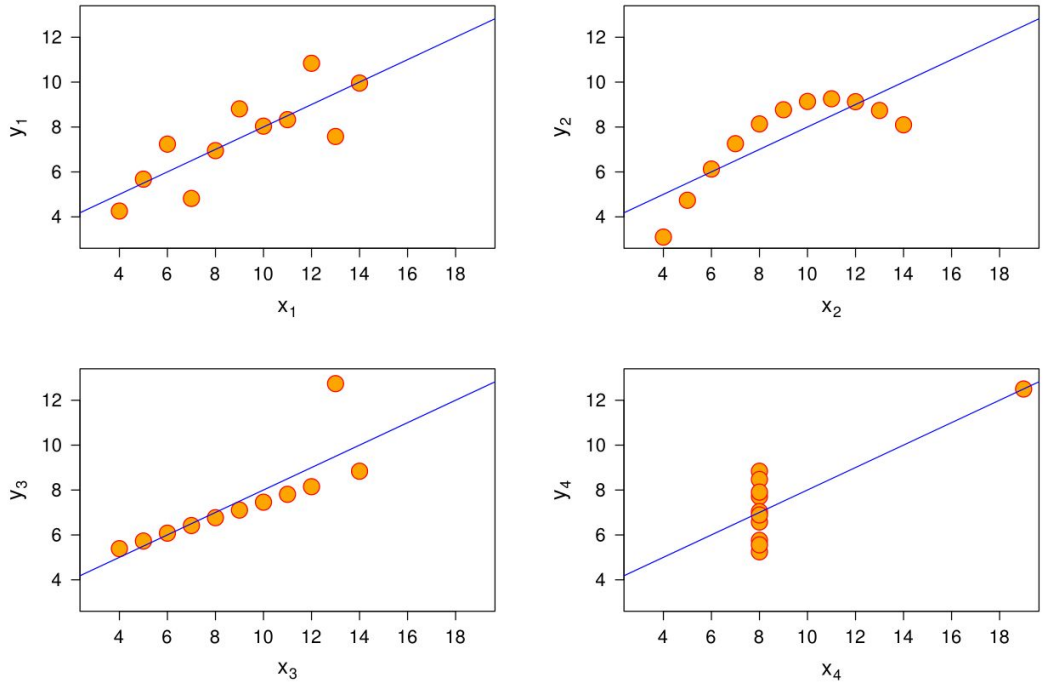
Temario de hoy

- Introducción a las visualizaciones
- ggplot2
- Otras visualizaciones
- BBC
- Recursos

¿Por dónde vamos?

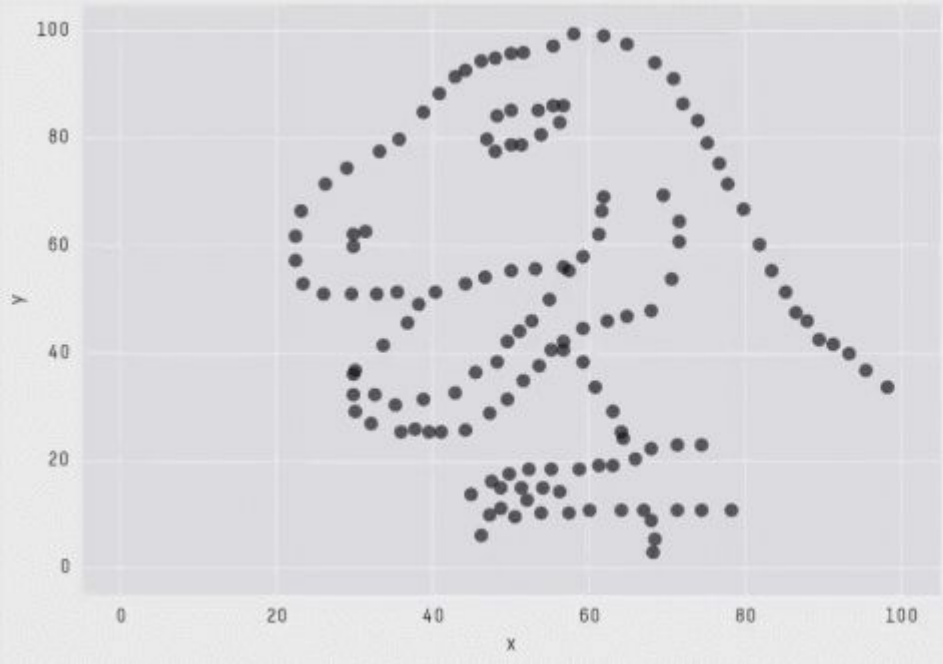


¿Por qué visualizar datos?



Más información: [Wikipedia]

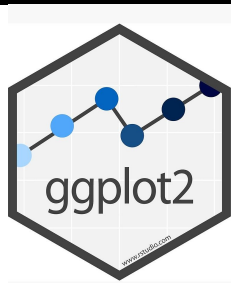
¿Por qué visualizar datos?



X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

Más información: [Wikipedia]

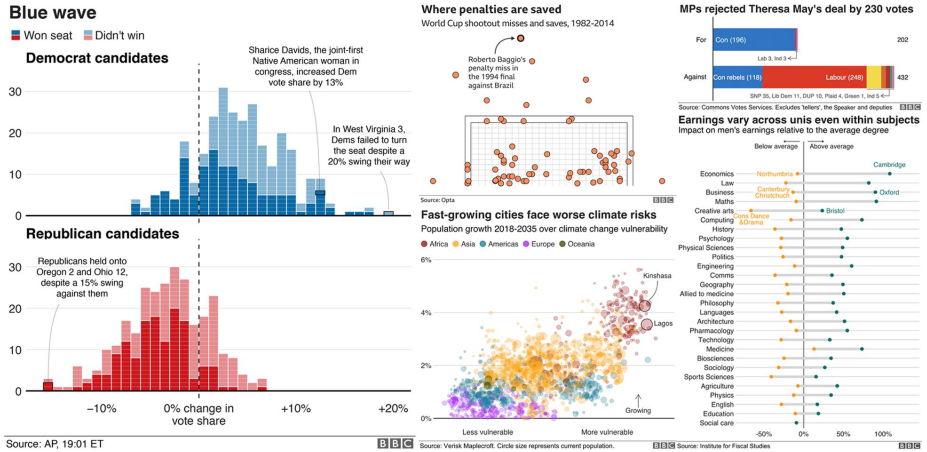
ggplot2



- ➔ Se basa en una teoría de [The Grammar of Graphics]
- ➔ Todo gráfico se puede compone de:
  - ◆ Elementos esenciales:
    - Datos: la vedette de los gráficos
    - Estética: en qué eje va cada elemento y con qué atributos
    - Geometrías: barras, líneas, puntos, etc?
  - ◆ Elementos opcionales:
    - Facetado: pequeños subsets particulares
    - Estadisticas: media, cuartile, mediana, etc
    - Coordenadas: transformar los ejes, etc
    - Temas: hacerlo pituco

Caso de uso: BBC

- ➔ BBC usa la base de “ggplot” para sus gráficos más rápidos.
- ➔ Ventajas: Libertad para graficar + reproductibilidad
- ➔ Links:
  - Recetario “*bbplot*”
  - Nota periodística al respecto
  - Presentación periodistas BBC (video - inglés)



Un mundo de posibilidades...

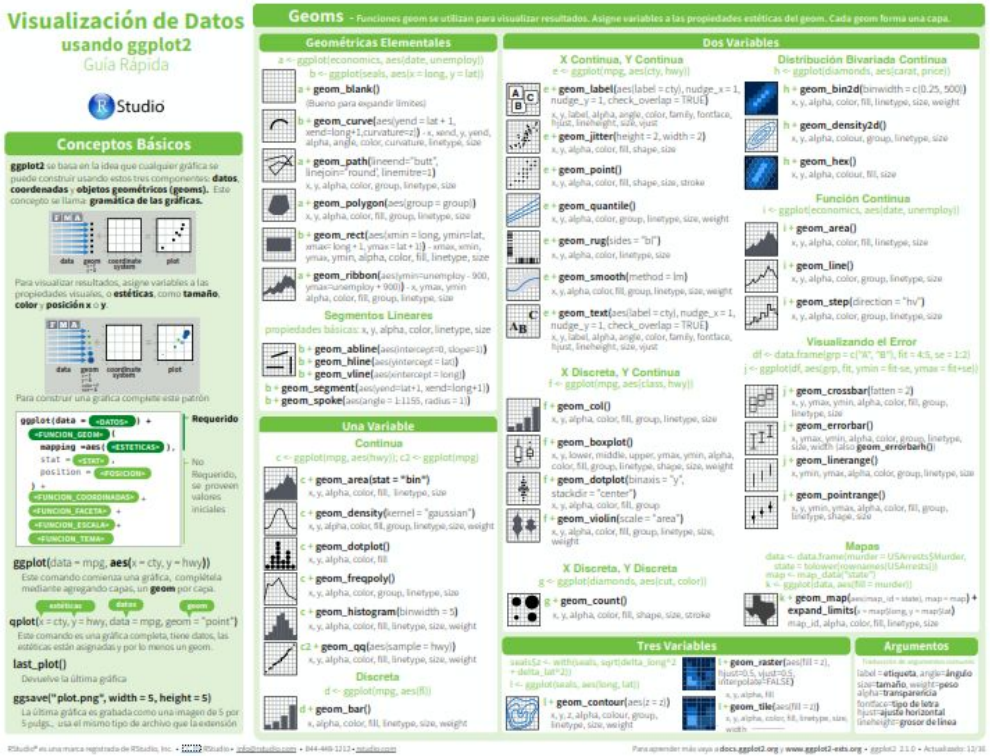
Más información:
[from Data to Viz]

Recursos: ggplot2

- Google
- R for Data Science
  - Cap 3: Data Visualization
  - Cap 28: Graphics for communication
- Libros
  - R Graphics Cookbook (Winston Chang, 2018)
  - ggplot2: Elegant Graphics for Data Analysis (Use R!) (Hadley Wickham, 2016)

Recursos: Cheatsheets

- ➔ RMarkdown
- ➔ Importar datos
- ➔ Transformación de datos
- ➔ ggplot2



Recursos: Repositorios de datos

- ➔ Datos.gob.ar (repositorio datos públicos Nacion)
- ➔ Buenos Aires data (repositorio datos Ciudad)
- ➔ Datos publicos MIOPyV