

Práctica Visualización de Datos I

Ainoa Iglesias Dasilva, 2024/25

alu0101164403@ull.edu.es

Máster, Análisis de Datos Masivos

[GitHub de esta práctica](#)

1. Resumen: Why scientists need to be better at data visualization

La visualización de datos en ciencia es fundamental para facilitar la comprensión de la información, sobre todo en la actualidad donde se acumulan cantidades enormes de datos. Es un medio muy útil para el análisis, interpretación y comunicación de los datos, ya que las imágenes permiten identificar patrones y relaciones que serían difíciles de ver en tablas o textos. Esto se debe a que los humanos entendemos y aprendemos mejor la información cuando se presenta de forma visual.

Sin embargo, los científicos no suelen dedicar mucho tiempo a esta parte, lo que lleva a visualizaciones erróneas y engañosas.

Hay estudios enfocados en este tema, que investigan cómo percibimos las distintas formas gráficas y cuál es la mejor manera de representar los datos. Por ejemplo, está demostrado que interpretamos mejor las longitudes, como las de los gráficos de barras o líneas, pero no tanto los colores o áreas.

Pero no todo se basa en cómo es más fácil interpretar la información, hay que tener en cuenta también lo que se quiere analizar o comunicar y elegir en consecuencia, ya que no cualquier gráfico sirve para representar todo.

La conclusión obtenida con este texto, es que una buena visualización no se basa solo en la parte estética. Es importante saber elegir el gráfico correcto según lo que se quiere mostrar para lograr una interpretación sencilla y útil.

2. Otros portales visualización

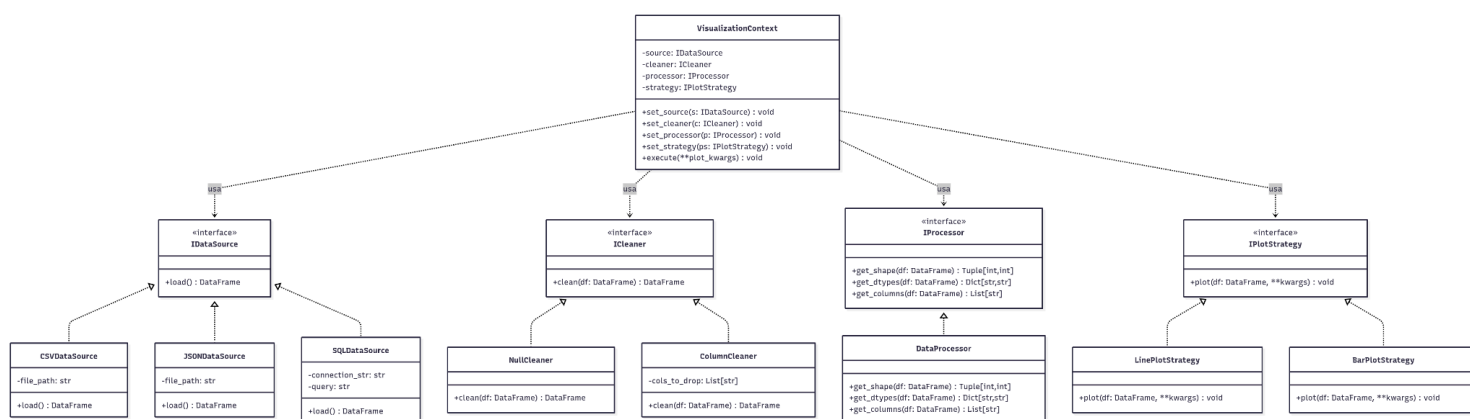
Gapminder es una fundación sueca que promueve el desarrollo global sostenible. Lo hacen ayudando a las personas a entender mejor el mundo y una de sus maneras es usando la visualización de datos. Gapminder crea gráficos e imágenes fáciles de entender para comunicar información compleja sobre temas globales.

Algunos [portales](#) con funciones similares son:

- **[Tableau Public](#)**: Muestra visualizaciones de datos curiosos de cualquier tipo y de maneras tanto muy creativas como más gráficas.
- **[World Bank Open Data](#)**: Tiene una amplia colección de datos sobre desarrollo global, incluyendo indicadores económicos, sociales y ambientales.
- **[Our World in Data](#)**: Presenta visualizaciones interactivas sobre una variedad de temas relacionados con el bienestar global, como salud, educación, pobreza e igualdad.
- **[Information is beautiful](#)**: presenta proyectos personalizados con un diseño brillante y minimalista.

3. Diagrama y tecnologías usadas

Crear un pequeño framework de visualización genérico que pueda adaptarse fácilmente a nuevas fuentes de datos.



- **Carga de datos**

1. **IDataSource** (interfaz)

- Método: load() -> pd.DataFrame (abstracto)

2. **CSVDataSource** (implementa IDataSource)

- Carga dataset en formato CSV y devuelve el dataframe

3. **JSONDataSource** (implementa IDataSource)

- Carga dataset en formato json

- **Limpieza datos dataset**

3. **ICleaner** (interfaz)

- Método: clean(dataframe: pd.DataFrame) → pd.DataFrame (abstracto)

4. **Cleaner** (implementa ICleaner)

- Contiene la lógica de limpieza e imputación de datos. Tiene varios modos, donde varía las condiciones para eliminar/imputar datos.

5. **TypeOnlyCleaner**(implementa ICleaner)

- Para cambiar el tipo de datos de una columna

- **Procesador de datos**

6. **IProcessor** (interfaz)

7. **DataProcessor** (implementa IProcessor)

- Contiene varias funciones que devuelven datos útiles del dataset, como la lista de columnas, valores únicos, tipos de datos, porcentajes de nulos, datos descriptivos...

- **Visualización de datos**

8. **IPlotStrategy** (interfaz)

- Método: plot() (abstracto)

9. **LineChartStrategy** (implementa IPlotStrategy)

10. **BarChartStrategy** (implementa IPlotStrategy)

Tecnologías usadas

- Python 3

- Pandas: para manipulación de datos.

- Matplotlib / Seaborn / Plotly: para visualizaciones.

- Jupyter Notebook: entorno presentación.

- Patrón de diseño: Strategy Pattern, para desacoplar visualización y fuentes de datos.

- Mermaid para diagrama de clases