

Guía para elegir visualizaciones

Elegir las visualizaciones adecuadas para la exploración de datos depende de varios factores, incluyendo el tipo de datos, las relaciones que deseas explorar y el mensaje que quieres comunicar.

1. Entender los Tipos de Datos

- **Datos Categóricos:** Nombres o etiquetas que no tienen un orden intrínseco (por ejemplo, colores, géneros).
- **Datos Ordinales:** Etiquetas con un orden natural (por ejemplo, clasificaciones de satisfacción).
- **Datos Numéricos:** Valores que son cuantitativos (por ejemplo, edad, ingresos).

2. Objetivo de la Visualización

- **Distribución:** Ver cómo se distribuyen los valores.
- **Comparación:** Comparar diferentes grupos o categorías.
- **Relación:** Examinar relaciones entre variables.
- **Composición:** Mostrar partes de un todo.
- **Tendencia:** Ver cambios a lo largo del tiempo.

3. Visualizaciones Comunes y sus Usos

Distribución

- **Histograma:** Útil para mostrar la distribución de una variable numérica.
- **Box Plot:** Ideal para mostrar la distribución, mediana, y posibles valores atípicos.

Comparación

- **Bar Plot:** Útil para comparar diferentes categorías.
- **Grouped Bar Plot:** Comparación entre subgrupos dentro de categorías.

Relación

- **Scatter Plot:** Ideal para mostrar la relación entre dos variables numéricas.

Composición

- **Pie Chart:** Útil para mostrar la composición de una categoría (con precaución).
- **Stacked Bar Plot:** Muestra la composición de categorías a través de diferentes subgrupos.

Tendencia

- **Line Plot:** Ideal para mostrar cambios a lo largo del tiempo.
- **Area Plot:** Similar al line plot, pero muestra la magnitud del cambio.
- **Time Series Plot:** Especializado en datos de series temporales.

4. Gráficos en Python

A continuación, podéis ver la sintaxis y utilidad de las gráficas que hemos visto en el módulo:

Análisis variables numéricas

Tipo de Gráfica	Método Seaborn	Método Matplotlib	Explicación
Histograma	<code>sns.histplot()</code>	<code>plt.hist()</code>	Representa la distribución de una variable numérica.
Diagrama de caja	<code>sns.boxplot()</code>	<code>plt.boxplot()</code>	Muestra la distribución y valores atípicos de una variable numérica.
Violinplot	<code>sns.violinplot()</code>	-	Combina un diagrama de caja con una estimación de densidad.

Gráfico de dispersión	<code>sns.scatterplot()</code>	<code>plt.scatter()</code>	Muestra la relación entre dos variables numéricas.
Gráfico de regresión	<code>sns.regplot()</code>	-	Muestra una línea de regresión entre dos variables numéricas.
Pairplot	<code>sns.pairplot()</code>	-	Muestra la relación entre múltiples variables numéricas.
Heatmap	<code>sns.heatmap()</code>	<code>plt.imshow()</code>	Visualiza la relación entre variables numéricas mediante colores.

Análisis variables categóricas y relación entre variables categóricas

Aquí tienes la tabla actualizada con los métodos de Matplotlib incluidos:

Tipo de Gráfica	Método Seaborn	Método Matplotlib	Explicación
Countplot	<code>sns.countplot()</code>	-	Muestra la frecuencia de cada categoría en una variable categórica mediante barras.
Pieplot	<code>plt.pie()</code>	<code>plt.pie()</code>	Visualiza la proporción de cada categoría en una variable categórica mediante un gráfico circular.

Relación variables numéricas con categóricas

En las tablas que hemos visto hasta ahora solo tenemos el tipo de gráfica que usaremos para analizar variables numéricas y variables categóricas. ¿Pero qué pasa si quiero relacionar variables numéricas con categóricas? Os dejamos de nuevo una tabla resumen con el tipo de gráficas que podremos usar:

Tipo de Gráfica	Método Seaborn	Método Matplotlib	Explicación
Barplot	<code>sns.barplot()</code>	<code>plt.bar()</code>	Muestra la relación entre una variable categórica y una variable numérica mediante barras.
Violinplot	<code>sns.violinplot()</code>	<code>plt.violinplot()</code>	Muestra la distribución de una variable numérica para cada categoría en una variable categórica.
Boxplot	<code>sns.boxplot()</code>	<code>plt.boxplot()</code>	Muestra la distribución de una variable numérica para cada categoría en una variable categórica.
Pointplot	<code>sns.pointplot()</code>	-	Muestra la relación entre una variable categórica y una variable numérica mediante puntos y líneas.