



INSTITUTO PROVINCIAL DE LA ADMINISTRACIÓN PÚBLICA (IPAP)

2022



USO DE NUEVAS TECNOLOGÍAS PARA EL ANÁLISIS DE DATOS EN EL ESTADO

Tomás Barbieri
2021

Clase 4

Objetivos de la clase 4

Identificar los diferentes gráficos y las herramientas que nos permiten visualizarlos.

Reconocer las ventajas de las diferentes formas de visualización de datos (Mapas, gráficos, estadísticas).

Comprender cómo exportar la información nueva generada en los diferentes formatos y diferentes criterios para la exportación.

Clase 4

Habiendo atravesado ya varias etapas del análisis de los datos: la exploración, la segmentación, necesaria para poder filtrar los datos, llega la última parte que es la visualización de la información.

Esta parte es tal vez una de las más importante ya que buscamos la mejor forma de mostrar nuestras conclusiones, nuestros análisis y nuestros subconjuntos de datos.

Veremos diferentes formas de visualización, desde gráficos, mapas, visualización de números, fechas, categorías, etc.

Cuadros y tablas

- Son una forma sencilla de mostrar los datos.
- Es una primera aproximación de cómo mostrar la información.
- Puede servir para mostrar una muestra de datos.

Año	2003	2007	2004	2008	2009	2010	2006
Cantidad obras	6	5	5	4	3	2	2

Gráficos

Los gráficos tal vez sean una de las herramientas más utilizadas para la visualización de los datos.

Dependiendo los datos que se quieran mostrar, cambiará el gráfico y la forma en visualizarlos. Veremos varios ejemplos de gráficos posibles y cuáles nos convienen para los diferentes tipos de datos.

Los tipos de gráficos son, entre otros: gráficos de torta, gráficos de barras, gráficos de líneas, histogramas y los gráficos de dispersión.

Gráficos de Torta

Los gráficos de torta son utilizados para representar diferentes categorías y una cantidad por ejemplo de datos de esa categoría. En este ejemplo por ejemplo el tipo de obras y el porcentaje de cada una.

```
import matplotlib.pyplot as plt

labels = dataframe_obras_viales['tipo_obra'].value_counts().index
sections = dataframe_obras_viales['tipo_obra'].value_counts()

plt.pie(sections, labels=labels,
        startangle=90,
        explode = (0.1, 0.1, 0.1, 0.1),
        autopct = '%1.2f%')

#plt.axis('equal') # Try commenting this out.
plt.title('Tipos de obras')
plt.show()
```



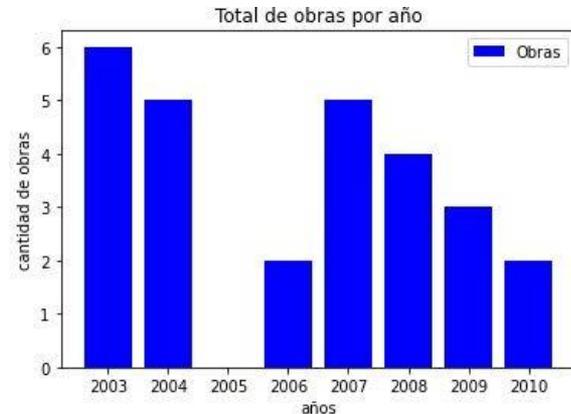
Gráficos de Barras

Los gráficos de barras son utilizados para mostrar los valores de una variable categórica en diferentes columnas, es decir, la cantidad de valores que tendrían dichos. Podemos mostrar el acumulado de obras por año.

```
▶ x1 = dataframe_obras_viales['año'].value_counts().index
  y1 = dataframe_obras_viales['año'].value_counts()

plt.bar(x1, y1, label="Obras", color='b')
plt.plot()

plt.xlabel("años")
plt.ylabel("cantidad de obras")
plt.title("Total de obras por año")
plt.legend()
plt.show()
```



Gráficos de Línea

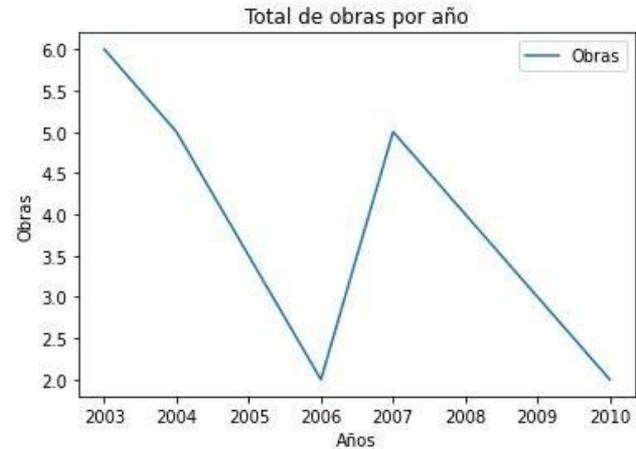
Los gráficos de línea son utilizados para mostrar una secuencia de valores representados a lo ancho del gráfico, conectando los diferentes puntos con una línea (recta).

```
serie = dataframe_obras_viales['año'].value_counts().sort_index()

x1 = serie.index
y1 = serie

plt.plot(x1, y1, label="Obras")
plt.plot()

plt.xlabel("Años")
plt.ylabel("Obras")
plt.title("Total de obras por año")
plt.legend()
plt.show()
```

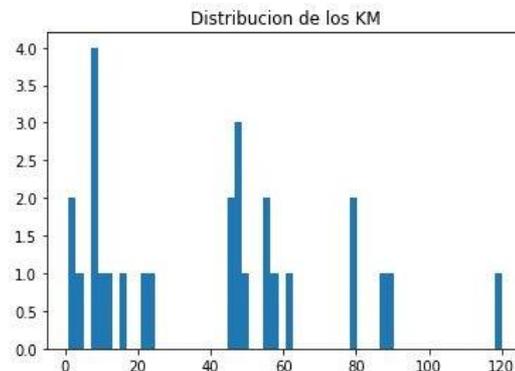


Histogramas

Un histograma se utiliza para representar los valores de una variable en barras. La cantidad de barras o segmentos es configurable y depende de la frecuencia de los valores que se tienen.

Por ejemplo, si quisiéramos saber cómo es la distribución de kilómetros entre todas las obras.

```
km = dataframe_obras_viales['cant_km']  
  
plt.hist(km, bins=60)  
plt.title("Distribucion de los KM")  
plt.show()
```

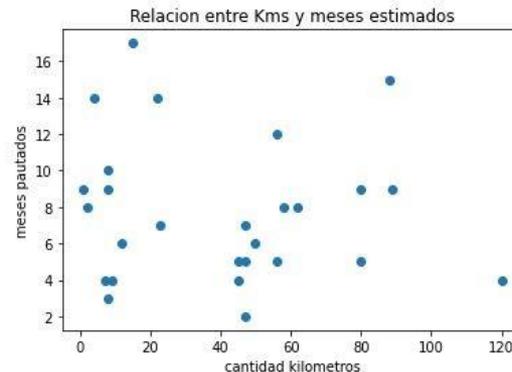


Gráficos de Dispersión

Se usa para hacer una correlación entre dos variables de nuestro conjunto de datos. Genera una coordenada cartesiana con ambos valores y luego se visualiza en un gráfico.

Por ejemplo, si quisiéramos ver la relación entre la Cantidad de kilómetros y cantidad de meses estimados de una obra.

```
x1 = dataframe_obras_viales['cant_km']  
y1 = dataframe_obras_viales['meses_pautados']  
  
plt.scatter(x1, y1)  
plt.title('Relacion entre Kms y meses estimados')  
plt.xlabel("cantidad kilometros")  
plt.ylabel("meses pautados")  
plt.show()
```



Mapas de calor

El mapa de calor es una herramienta muy interesante. En el ejemplo que estamos viendo es difícil de aplicar, pero podemos ver otro ejemplo. Pensemos en un equipo de fútbol, si quisiéramos mostrar en una cancha por donde hubo más recorridos del jugador...

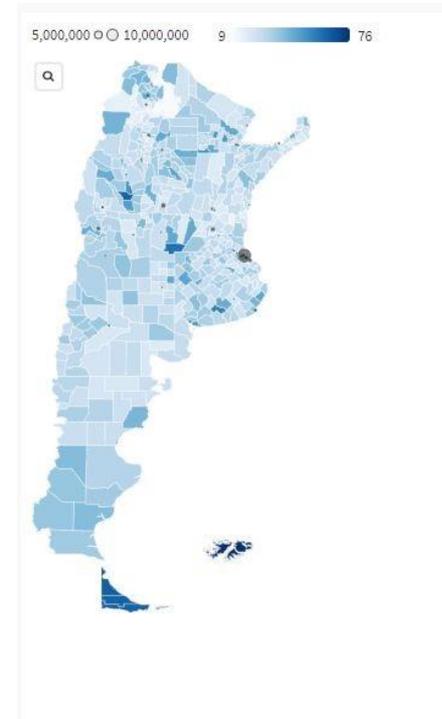
tendríamos algo así
zonas rojas, más transitado
zonas azules, menos transitado



Mapas

Luego tenemos los mapas tradicionales
En este caso tenemos que tener un tipo de dato especial que es la georeferenciación, es muy importante y también tener los mapas a medida de lo que queremos visualizar.

En este caso cada departamento de Argentina tiene un color asignado dependiendo de un valor interno, cuanto más grande el valor, más azul.



Exportar la información

Tal vez nos interese guardar la información procesada para posteriores análisis o para añadir a informes que tengamos que confeccionar. En este caso también podemos conservar los conjuntos de datos de forma separada.

Pandas nos permite exportar la información en formato excel o en formato CSV.

Esto nos permite también hacer visualizaciones en otras herramientas.

Links de interés

- [Visualización de datos - tableau](#)
- [Mapas de calor](#)
- [Ejemplo de gráficos con matplotlib](#)
- [Colores en gráficos](#)
- [Flourish](#)

Pregunta/Debate

Para comentar en el Foro:

¿Cuál gráfico les parece más interesante? ¿Por que?

¿Qué otros gráficos conocen?

ipap.gba.gov.ar