

Data Science & Machine Learning



GOBIERNO
DE ESPAÑA

VICEPRESIDENCIA
PRIMERA DEL GOBIERNO
MINISTERIO
DE ASUNTOS ECONÓMICOS
Y TRANSFORMACIÓN DIGITAL

SECRETARÍA DE ESTADO
DE DIGITALIZACIÓN
E INTELIGENCIA ARTIFICIAL

red.es

Centro de
Bases de Datos
en Ciencia, Tecnología
y Planificación

CRN
Digital

GARANTÍA
JUVENIL



UNIÓN EUROPEA

Barrabés

The Valley

"El FSE invierte en tu futuro"
Fondo Social Europeo

Data Science & Machine Learning



Data Science & Machine Learning



GOBIERNO
DE ESPAÑA

VICEPRESIDENCIA
PRIMERA DEL GOBIERNO
MINISTERIO
DE ASUNTOS ECONÓMICOS
Y TRANSFORMACIÓN DIGITAL

SECRETARÍA DE ESTADO
DE DIGITALIZACIÓN
E INTELIGENCIA ARTIFICIAL

red.es

Centro de
Búsqueda y
Análisis de
Datos en
Ciencia, Tecnología
y Políticas

CRN
Digital

GARANTÍA
JUVENIL



UNIÓN EUROPEA

Barrabés

The Valley

"El FSE invierte en tu futuro"
Fondo Social Europeo

Índice.

1. Visualización de datos
2. Paquetes para la visualización de datos
 - 2.1 Matplot Lib
 - 2.2 Seaborn
 - 2.3 Plotly
 - 2.4 Folium
3. Tipos de gráficas
 - 3.1 Gráfico de barras
 - 3.2 Gráfico de líneas
 - 3.3 Gráfico circular
 - 3.4 Gráfico de puntos



¿Qué es la visualización de datos?

¿Qué es la visualización de datos?

dataset_part_2.csv

```

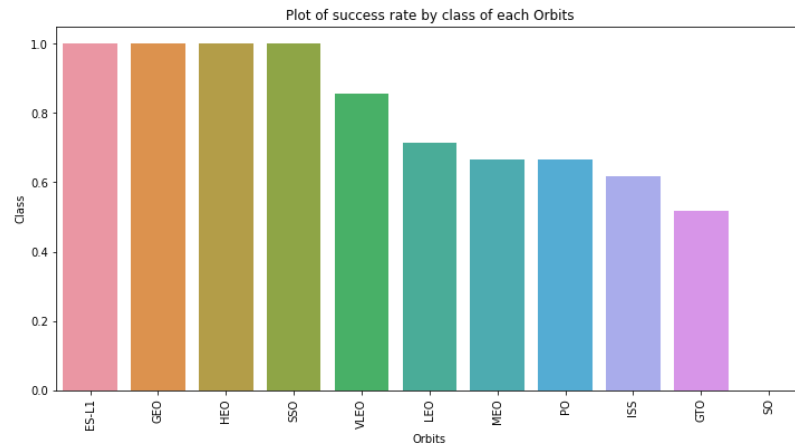
1 FlightNumber,Date,BoosterVersion,PayloadMass,Payload,Orbit,LaunchSite,LandingOutcome,Flights,GridFins,Reused,Legs,LandingPad,Block,Customers,Mission_Outcome,ReusedCount,Serial,Longitude,Latitude,Classification
2 1,2010-06-04,Falcon 9,6123.547647058824,Dragon Qualification Unit,LEO,CCSFS SLC 40,None None,1,False,False,False,,1.0,SpaceX,True,0,B0003,-80.577366,28.5618571,0
3 2,2012-05-22,Falcon 9,525.0,COTS Demo Flight 2,LEO,CCSFS SLC 40,None None,1,False,False,False,,1.0,NASA(COTS),True,0,B0005,-80.577366,28.5618571,0
4 3,2013-03-01,Falcon 9,677.0,CRS-2,ISS,CCSFS SLC 40,None None,1,False,False,False,,1.0,NASA (CRS),True,0,B0007,-80.577366,28.5618571,0
5 4,2013-09-29,Falcon 9,500.0,CASSIOPE,PO,VAFB SLC 4E,False Ocean,1,False,False,False,,1.0,MDA,True,0,B1003,-120.610829,34.632093,0
6 5,2013-12-03,Falcon 9,3170.0,SES-8,GTO,CCSFS SLC 40,None None,1,False,False,False,,1.0,SES,True,0,B1004,-80.577366,28.5618571,0
7 6,2014-01-06,Falcon 9,3325.0,Thaicom 6,GTO,CCSFS SLC 40,None None,1,False,False,False,,1.0,Thaicom,True,0,B1005,-80.577366,28.5618571,0
8 7,2014-04-18,Falcon 9,2296.0,CRS-3,ISS,CCSFS SLC 40,True Ocean,1,False,False,True,,1.0,NASA (CRS),True,0,B1006,-80.577366,28.5618571,1
9 8,2014-07-14,Falcon 9,1316.0,Orbcomm-0G2-M1,LEO,CCSFS SLC 40,True Ocean,1,False,False,True,,1.0,Orbcomm,True,0,B1007,-80.577366,28.5618571,1
10 9,2014-08-05,Falcon 9,4535.0,AsiaSat 8,GTO,CCSFS SLC 40,None None,1,False,False,False,,1.0,AsiaSat,True,0,B1008,-80.577366,28.5618571,0
11 10,2014-09-07,Falcon 9,4428.0,AsiaSat 6,GTO,CCSFS SLC 40,None None,1,False,False,False,,1.0,AsiaSat,True,0,B1011,-80.577366,28.5618571,0
12 11,2014-09-21,Falcon 9,2216.0,CRS-4,ISS,CCSFS SLC 40,False Ocean,1,False,False,False,,1.0,NASA (CRS),True,0,B1010,-80.577366,28.5618571,0
13 12,2015-01-10,Falcon 9,2395.0,CRS-5,ISS,CCSFS SLC 40,False ASDS,1,True,False,True,5e9e3032383ecb761634e7cb,1.0,NASA (CRS),True,0,B1012,-80.577366,28.5618571,0
14 13,2015-02-11,Falcon 9,570.0,DISCOVER,ES-L1,CCSFS SLC 40,True Ocean,1,True,False,True,,1.0,"NASA,NOAA,USAF",True,0,B1013,-80.577366,28.5618571,1
15 14,2015-04-14,Falcon 9,1898.0,CRS-6,ISS,CCSFS SLC 40,False ASDS,1,True,False,True,5e9e3032383ecb761634e7cb,1.0,NASA (CRS),True,0,B1015,-80.577366,28.5618571,0
16 15,2015-04-27,Falcon 9,4707.0,TürkmenÄlem 52°E / MonacoSAT,GTO,CCSFS SLC 40,None None,1,False,False,False,,1.0,Turkmenistan National Space Agency,True,0,B1016,-80.577366,28.5618571,0
17 16,2015-06-28,Falcon 9,2477.0,CRS-7,ISS,CCSFS SLC 40,None ASDS,1,True,False,True,5e9e3032383ecb6bb234e7ca,1.0,NASA (CRS),False,0,B1018,-80.577366,28.5618571,0
18 17,2015-12-22,Falcon 9,2034.0,Orbcomm-0G2-M2,LEO,CCSFS SLC 40,True RTLS,1,True,False,True,5e9e3032383ecb267a34e7c7,1.0,Orbcomm,True,0,B1019,-80.577366,28.5618571,1
19 18,2016-01-17,Falcon 9,553.0,Jason-3,PO,VAFB SLC 4E,False ASDS,1,True,False,True,5e9e3032383ecb9e534e7cc,1.0,"NASA,NOAA",True,0,B1017,-120.610829,34.632093,0
20 19,2016-03-04,Falcon 9,5271.0,SES-9,GTO,CCSFS SLC 40,False ASDS,1,True,False,True,5e9e3032383ecb6bb234e7ca,1.0,SES,True,0,B1020,-80.577366,28.5618571,0
21 20,2016-04-08,Falcon 9,3136.0,CRS-8,ISS,CCSFS SLC 40,True ASDS,1,True,False,True,5e9e3032383ecb6bb234e7ca,2.0,NASA (CRS),True,1,B1021,-80.577366,28.5618571,1
22 21,2016-05-06,Falcon 9,4696.0,JCSAT-28,GTO,CCSFS SLC 40,True ASDS,1,True,False,True,5e9e3032383ecb6bb234e7ca,2.0,SKY Perfect JSAT Group,True,0,B1022,-80.577366,28.5618571,1
23 22,2016-05-27,Falcon 9,3100.0,Thaicom 8,GTO,CCSFS SLC 40,True ASDS,1,True,False,True,5e9e3032383ecb6bb234e7ca,2.0,Thaicom,True,1,B1023,-80.577366,28.5618571,1
24 23,2016-07-18,Falcon 9,2257.0,CRS-9,ISS,CCSFS SLC 40,True RTLS,1,True,False,True,5e9e3032383ecb267a34e7c7,2.0,NASA (CRS),True,1,B1025,-80.577366,28.5618571,1
25 24,2016-08-14,Falcon 9,4600.0,JCSAT-16,GTO,CCSFS SLC 40,True ASDS,1,True,False,True,5e9e3032383ecb6bb234e7ca,2.0,SKY Perfect JSAT Group,True,0,B1026,-80.577366,28.5618571,1
26 25,2016-09-01,Falcon 9,5500.0,Amos-6,GTO,CCSFS SLC 40,None ASDS,1,True,False,True,5e9e3032383ecb6bb234e7ca,3.0,Spacecom,False,0,B1028,-80.577366,28.5618571,0
27 26,2017-01-14,Falcon 9,9600.0,Iridium NEXT 1,PO,VAFB SLC 4E,True ASDS,1,True,False,True,5e9e3032383ecb9e534e7cc,3.0,Iridium Communications,True,1,B1029,-120.610829,34.632093,1
28 27,2017-02-19,Falcon 9,2490.0,CRS-10,ISS,KSC LC 39A,True RTLS,1,True,False,True,5e9e3032383ecb267a34e7c7,3.0,NASA (CRS),True,1,B1031,-80.6039558,28.6080585,1
29 28,2017-03-16,Falcon 9,5600.0,EchoStar 23,GTO,KSC LC 39A,None None,1,False,False,False,,3.0,EchoStar,True,0,B1030,-80.6039558,28.6080585,0
30 29,2017-03-30,Falcon 9,5300.0,SES-10,GTO,KSC LC 39A,True ASDS,2,True,False,True,5e9e3032383ecb6bb234e7ca,2.0,SES,True,1,B1021,-80.6039558,28.6080585,1
31 30,2017-05-01,Falcon 9,6123.547647058824,NROL-76,LEO,KSC LC 39A,True RTLS,1,True,False,True,5e9e3032383ecb267a34e7c7,3.0,NRO,True,1,B1032,-80.6039558,28.6080585,1
32 31,2017-05-15,Falcon 9,6070.0,Inmarsat-5 F4,GTO,KSC LC 39A,None None,1,False,False,False,,3.0,Inmarsat,True,0,B1034,-80.6039558,28.6080585,0
33 32,2017-06-03,Falcon 9,2708.0,CRS-11,ISS,KSC LC 39A,True RTLS,1,True,False,True,5e9e3032383ecb267a34e7c7,3.0,NASA (CRS),True,1,B1035,-80.6039558,28.6080585,1
34 33,2017-06-23,Falcon 9,3669.0,BulgariaSat-1,GTO,KSC LC 39A,True ASDS,2,True,True,True,5e9e3032383ecb6bb234e7ca,3.0,Bulgaria Sat,True,1,B1029,-80.6039558,28.6080585,1
35 34,2017-06-25,Falcon 9,9600.0,Iridium NEXT 2,PO,VAFB SLC 4E,True ASDS,1,True,False,True,5e9e3032383ecb9e534e7cc,3.0,Iridium Communications,True,1,B1036,-120.610829,34.632093,1
36 35,2017-07-05,Falcon 9,6761.0,Intelsat 35e,GTO,KSC LC 39A,None None,1,False,False,False,,3.0,Intelsat,True,0,B1037,-80.6039558,28.6080585,0
37 36,2017-08-14,Falcon 9,2910.0,CRS-12,ISS,KSC LC 39A,True RTLS,1,True,False,True,5e9e3032383ecb267a34e7c7,4.0,NASA (CRS),True,1,B1039,-80.6039558,28.6080585,1
38 37,2017-08-24,Falcon 9,475.0,Formosat-5,SSO,VAFB SLC 4E,True ASDS,1,True,False,True,5e9e3032383ecb9e534e7cc,3.0,NSPO (Taiwan),True,1,B1038,-120.610829,34.632093,1
39 38,2017-09-07,Falcon 9,4990.0,X-37B OTV-5,LEO,KSC LC 39A,True RTLS,1,True,False,True,5e9e3032383ecb267a34e7c7,4.0,USAF,True,1,B1040,-80.6039558,28.6080585,1
40 39,2017-10-09,Falcon 9,9600.0,Iridium NEXT 3,PO,VAFB SLC 4E,True ASDS,1,True,False,True,5e9e3032383ecb9e534e7cc,4.0,Iridium Communications,True,1,B1041,-120.610829,34.632093,1
41 40,2017-10-11,Falcon 9,5200.0,SES-11 / Echostar 105,GTO,KSC LC 39A,True ASDS,2,True,True,True,5e9e3032383ecb6bb234e7ca,3.0,"SES,Echostar",True,1,B1031,-80.6039558,28.6080585,1
42 41,2017-10-30,Falcon 9,3700.0,KoreaSat 5A,GTO,KSC LC 39A,True ASDS,1,True,False,True,5e9e3032383ecb6bb234e7ca,4.0,KT Corporation,True,0,B1042,-80.6039558,28.6080585,1
43 42,2017-12-15,Falcon 9,2205.0,CRS-13,ISS,CCSFS SLC 40,True RTLS,2,True,True,True,5e9e3032383ecb267a34e7c7,3.0,NASA (CRS),True,1,B1035,-80.577366,28.5618571,1
44 43,2017-12-23,Falcon 9,9600.0,Iridium NEXT 4,PO,VAFB SLC 4E,True Ocean,2,True,True,True,5e9e3032383ecb9e534e7cc,3.0,Iridium Communications,True,1,B1036,-120.610829,34.632093,1
45 44,2018-01-08,Falcon 9,6123.547647058824,ZUMA,LEO,CCSFS SLC 40,True RTLS,1,True,False,True,5e9e3032383ecb267a34e7c7,4.0,Northrop Grumman,True,1,B1043,-80.577366,28.5618571,1
46 45,2018-01-31,Falcon 9,4230.0,GovSat-1,GTO,CCSFS SLC 40,True Ocean,2,True,True,True,,3.0,GovSat,True,1,B1032,-80.577366,28.5618571,1
47 46,2018-03-06,Falcon 9,6092.0,Hispasat 30W-6,GTO,CCSFS SLC 40,None None,1,True,False,True,,4.0,Hispasat,True,0,B1044,-80.577366,28.5618571,0
48 47,2018-03-30,Falcon 9,9600.0,Iridium NEXT 5,PO,VAFB SLC 4E,None None,2,True,True,True,,4.0,Iridium Communications,True,1,B1041,-120.610829,34.632093,0
49 48,2018-04-02,Falcon 9,2760.0,CRS-14,ISS,CCSFS SLC 40,None None,2,True,True,True,,4.0,NASA (CRS),True,1,B1039,-80.577366,28.5618571,0
50 49,2018-04-18,Falcon 9,350.0,TESS,HEO,CCSFS SLC 40,True ASDS,1,True,False,True,5e9e3032383ecb6bb234e7ca,4.0,NASA,True,1,B1045,-80.577366,28.5618571,1
51 50,2018-05-11,Falcon 9,3750.0,Bangabandhu-1,GTO,KSC LC 39A,True ASDS,1,True,False,True,5e9e3032383ecb6bb234e7ca,5.0,Bangladesh Gov,True,3,B1046,-80.6039558,28.6080585,1
52 51,2018-06-04,Falcon 9,5383.85,SES-12,GTO,CCSFS SLC 40,None None,2,False,True,False,,4.0,SES,True,1,B1040,-80.577366,28.5618571,0
53 52,2018-06-29,Falcon 9,2410.0,CRS-15,ISS,CCSFS SLC 40,None None,2,False,True,False,,4.0,NASA (CRS),True,1,B1045,-80.577366,28.5618571,0
54 53,2018-07-22,Falcon 9,7076.0,Telstar 19V,GTO,CCSFS SLC 40,True ASDS,1,True,False,True,5e9e3032383ecb6bb234e7ca,5.0,Telesat,True,2,B1047,-80.577366,28.5618571,1

```

Col 1: FlightNumber Lin 1, col. 1 Espacios: 4 UTF-8 LF CSV

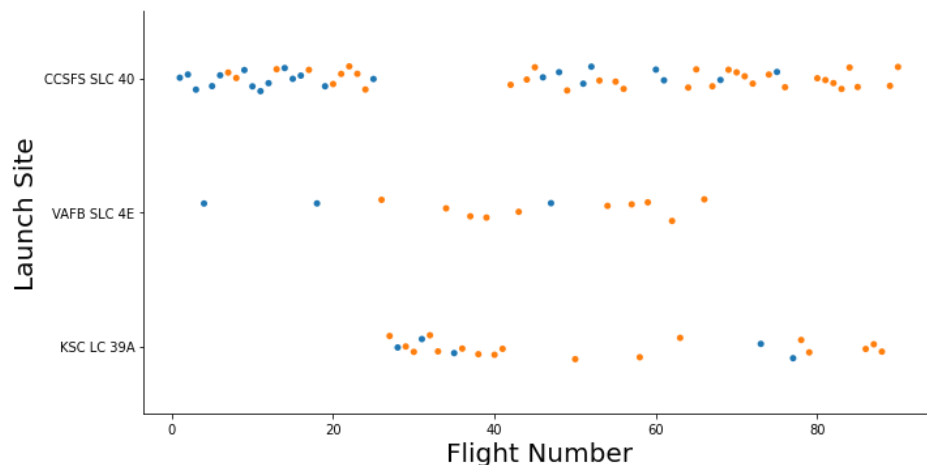
Si abrimos un .csv podemos ver los datos, pero
¿Los estamos visualizando?

¿Qué es la visualización de datos?



La visualización de datos es el proceso de utilizar elementos visuales para representar datos.

Este proceso facilita el análisis de los datos y nos permite extraer conclusiones.



Debe ajustarse al formato del dato y al objetivo, ya que hay numerosos tipos de gráficos, mapas, esquemas, etc.

Es una forma efectiva de comunicación.

Paquetes para la visualización de datos

Pandas & numpy

NO son paquetes destinados a la visualización pero sí necesarios para llevarla a cabo correctamente.

Seaborn

Es un paquete escrito encima de Matplotlib. Añade facilidades a la hora de realizar ciertas visualizaciones y los gráficos son considerados más estéticos.

La sintaxis también es considerada más natural.

Matplotlib

La librería más popular para la visualización de datos.

Destaca a la hora de realizar gráficos estáticos y sencillos. También es posible realizar gráficos animados, interactivos y en varias dimensiones pero no es considerada la mejor opción.

La sintaxis está basada en MATLAB.

Plotly

Es una herramienta de visualización y análisis de datos. Permite la creación de gráficos complejos y variados. Es bastante nueva con nuevas formas de ver los datos pero a su vez es más costosa computacionalmente.

Paquetes para la visualización de datos

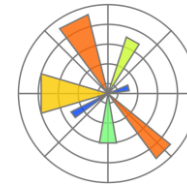
Pandas & numpy



Seaborn



Matplotlib



Plotly



Folium

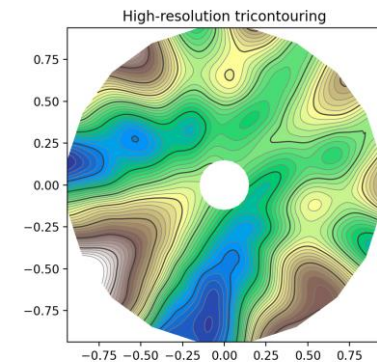
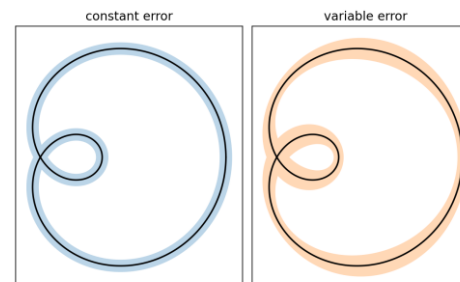
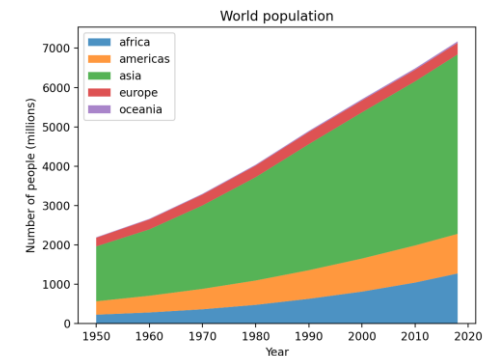
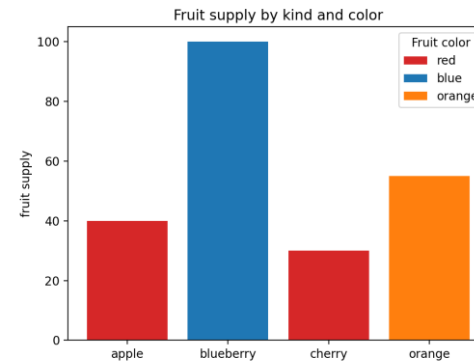


Folium

Paquetes para la visualización de datos

Matplotlib

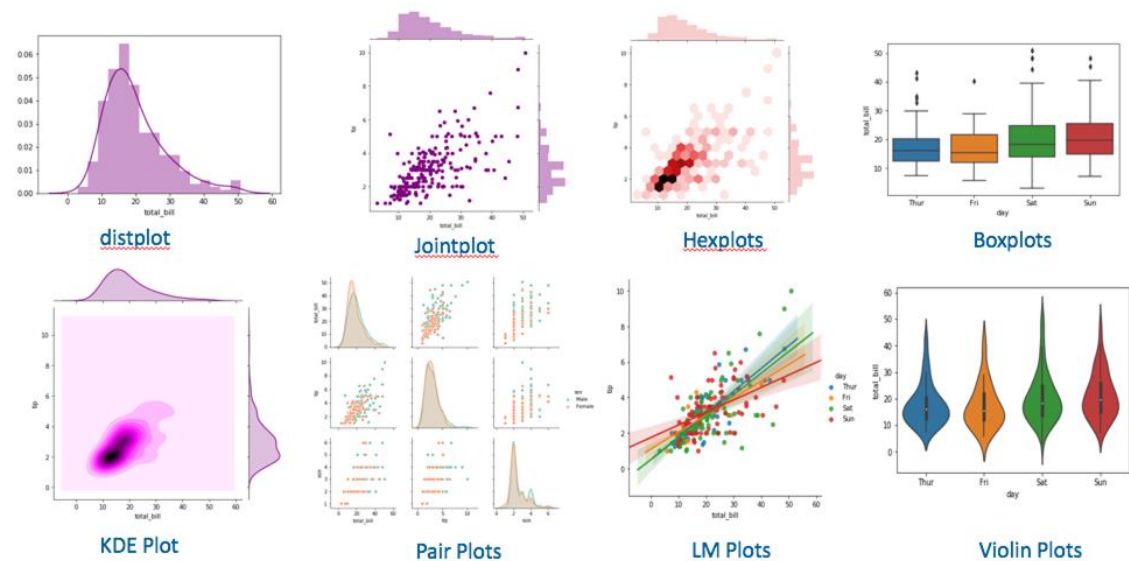
- La más popular.
- Efectiva para gráficos sencillos.
- Es posible realizar gráficos animados, interactivos y en varias dimensiones pero no TOP.
- Basada en MATLAB.



Paquetes para la visualización de datos

Seaborn

- Es un paquete escrito encima de Matplotlib.
- Añade facilidades a la hora de realizar ciertas visualizaciones y los gráficos son considerados más estéticos.
- La sintaxis también es considerada más natural.
- Puede resultar dependiente de Matplotlib en ciertas ocasiones

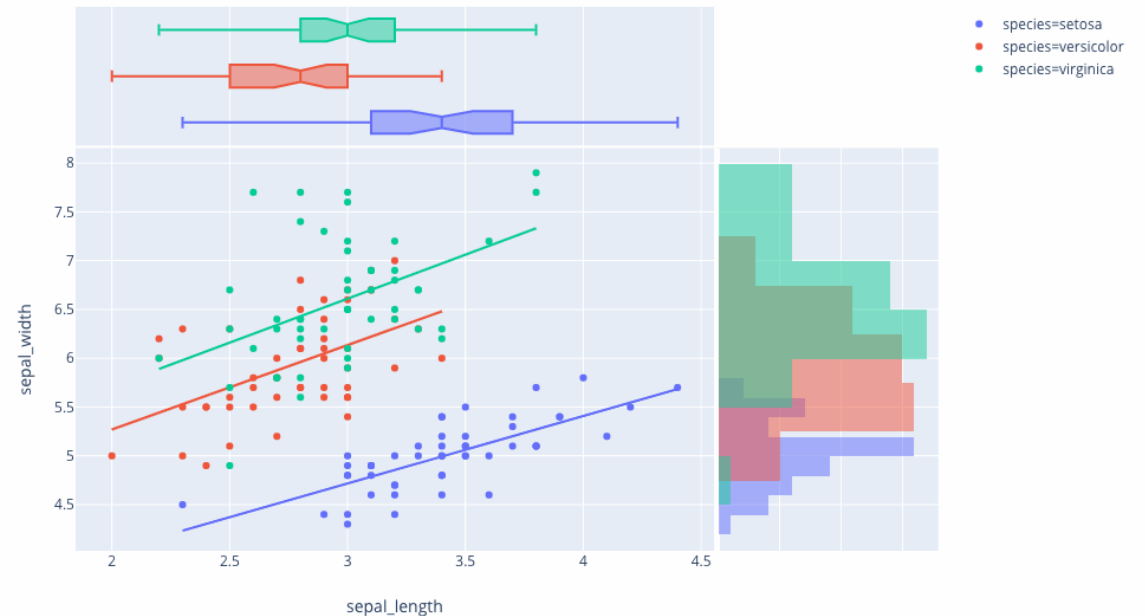


Paquetes para la visualización de datos

Plotly

- Funcionamiento diferente a Matplotlib.
- Paquete estético muy visual.
- Bueno para gráficos interactivos.
- Posibilidad de realizar dashboards (Plotly dash)
- Más pesado y costoso computacionalmente que Matplotlib

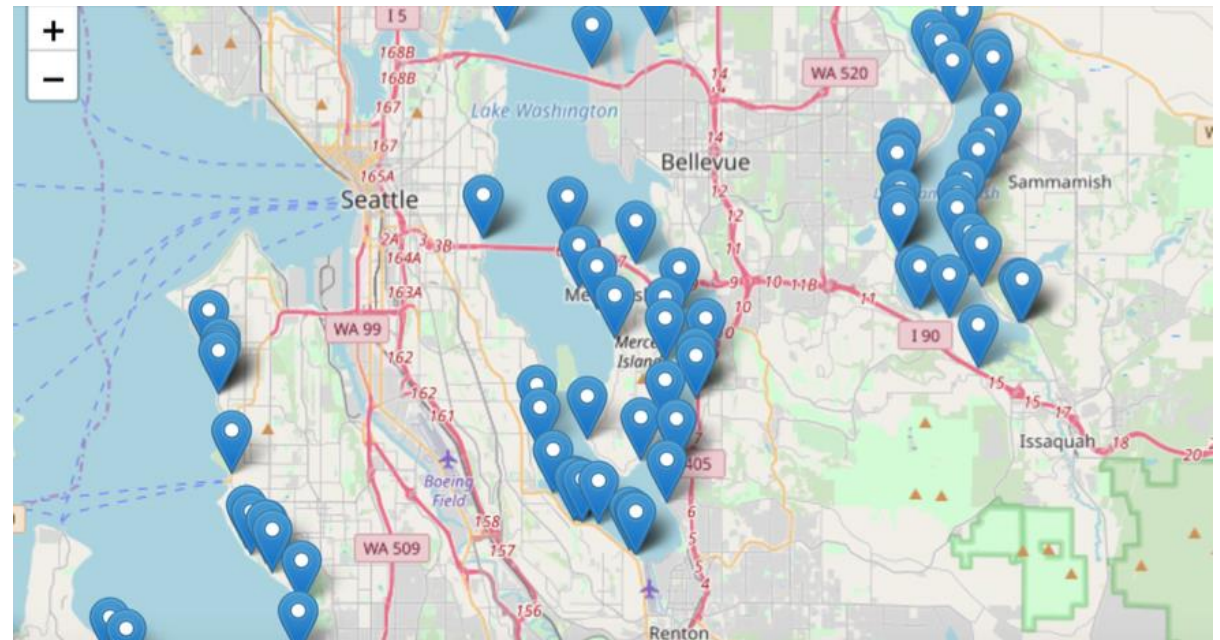
```
px.scatter(iris, x="sepal_width", y="sepal_length", color="species", marginal_y="histogram",  
           marginal_x="box", trendline="ols")
```



Paquetes para la visualización de datos

Folium

- Ideado para la visualización de datos geospaciales.
- Diferentes tipos de mapas.
- Gráficos interactivos.
- Basado el Leaflet.



Un universo infinito.

- Herramientas de visualización externas como Power BI, Tableau, etc.
- Muchas otras librerías de python como Altair, Geoplotlib, Bokeh, Pygal y un larguísimo etc.



Pygal

bokeh

Tipos de gráficos.



Tipos de gráficos.

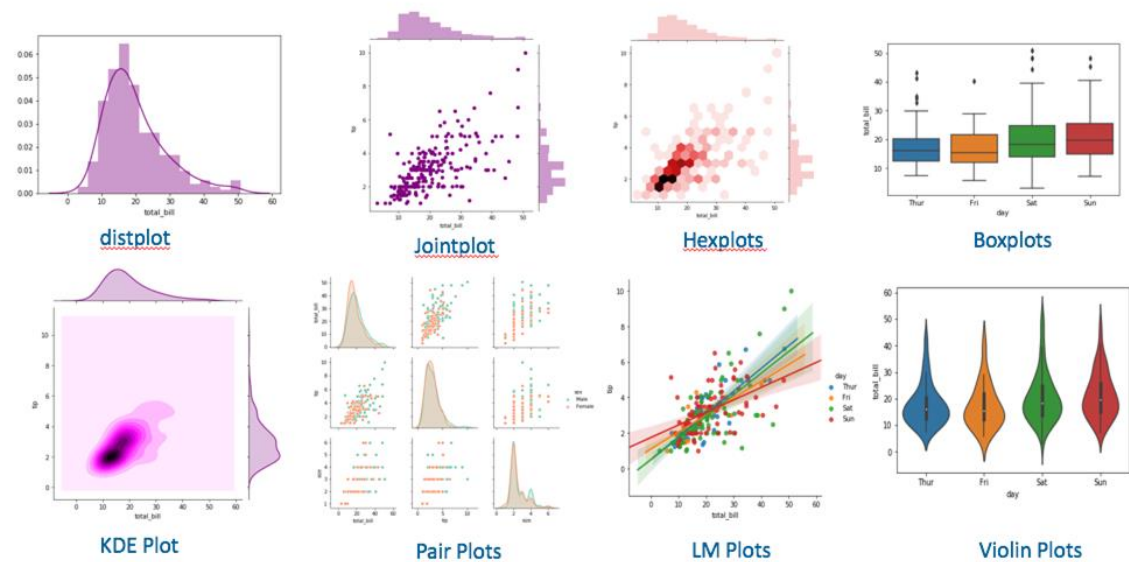
Elegir el gráfico adecuado

Gran variedad de gráficos.

Importante elegir el gráfico adecuado.

Atención a los datos y al objetivo.

Presentamos los más sencillos:



Tipos de gráficos.

Tipos de datos.

Cualitativos: se refieren a cualidades o modalidades que no pueden expresarse numéricamente.

Pueden ser:

- **ordinales:** si siguen un orden o secuencia (ej. el abecedario, los meses del año).
- **categóricos:** si no siguen ningún orden (ej. el estado civil de las personas: solteros, casados, viudos, divorciados y separados).

Cuantitativos: se refieren a cantidades o valores numéricos. Pueden ser:

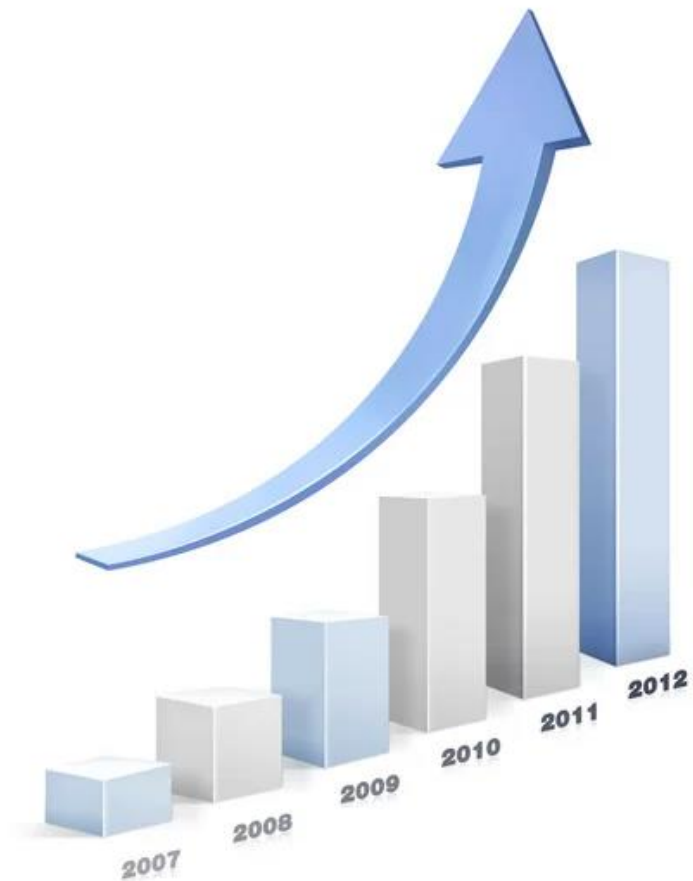
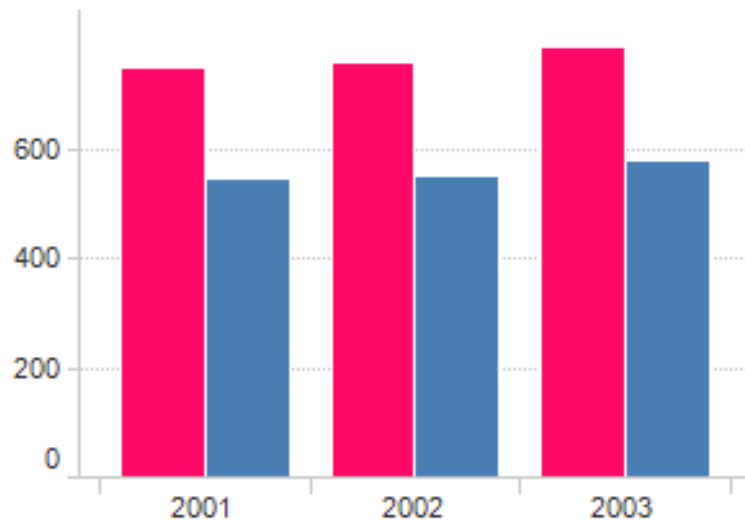
- **discretos:** si toman valores enteros (0, 1, 2, 3...). Ej. el número de hijos, el número de alumnos de una clase...

- **continuos:** si pueden tomar cualquier valor dentro de un intervalo (ej. la estatura o el peso de las personas).

Tipos de gráficos.

Gráfico de barras

Un gráfico de barras es una representación gráfica en un eje cartesiano de las **frecuencias de una variable cualitativa o discreta**.

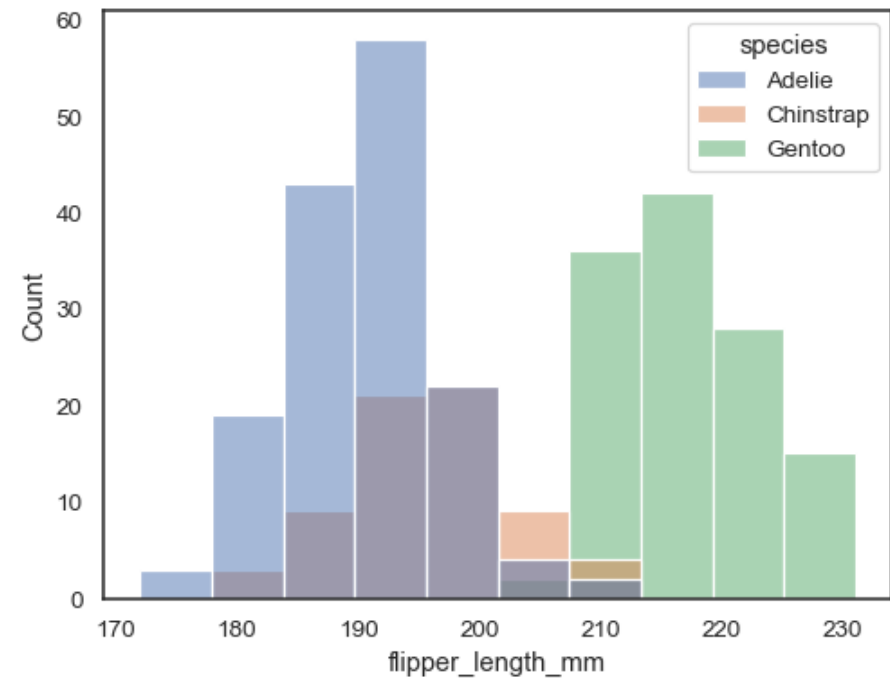
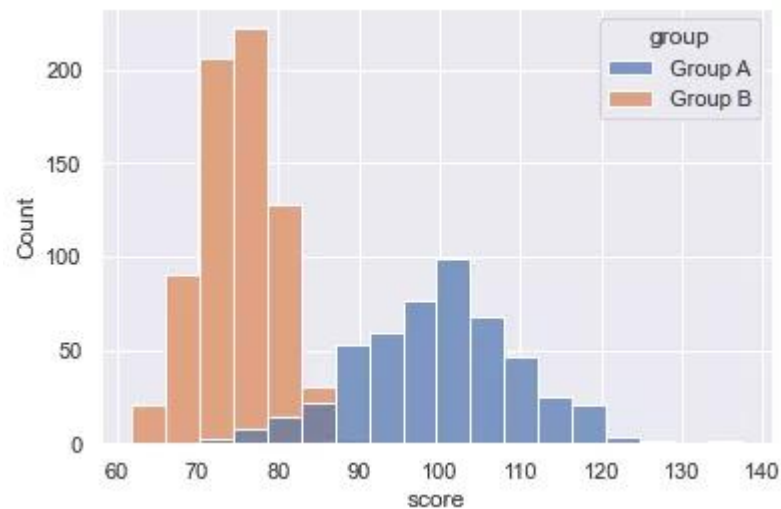


Tipos de gráficos.

Histograma

Semejante al gráfico de barras.

Se utilizan barras para indicar la frecuencia de una variable que **puede ser continua**.

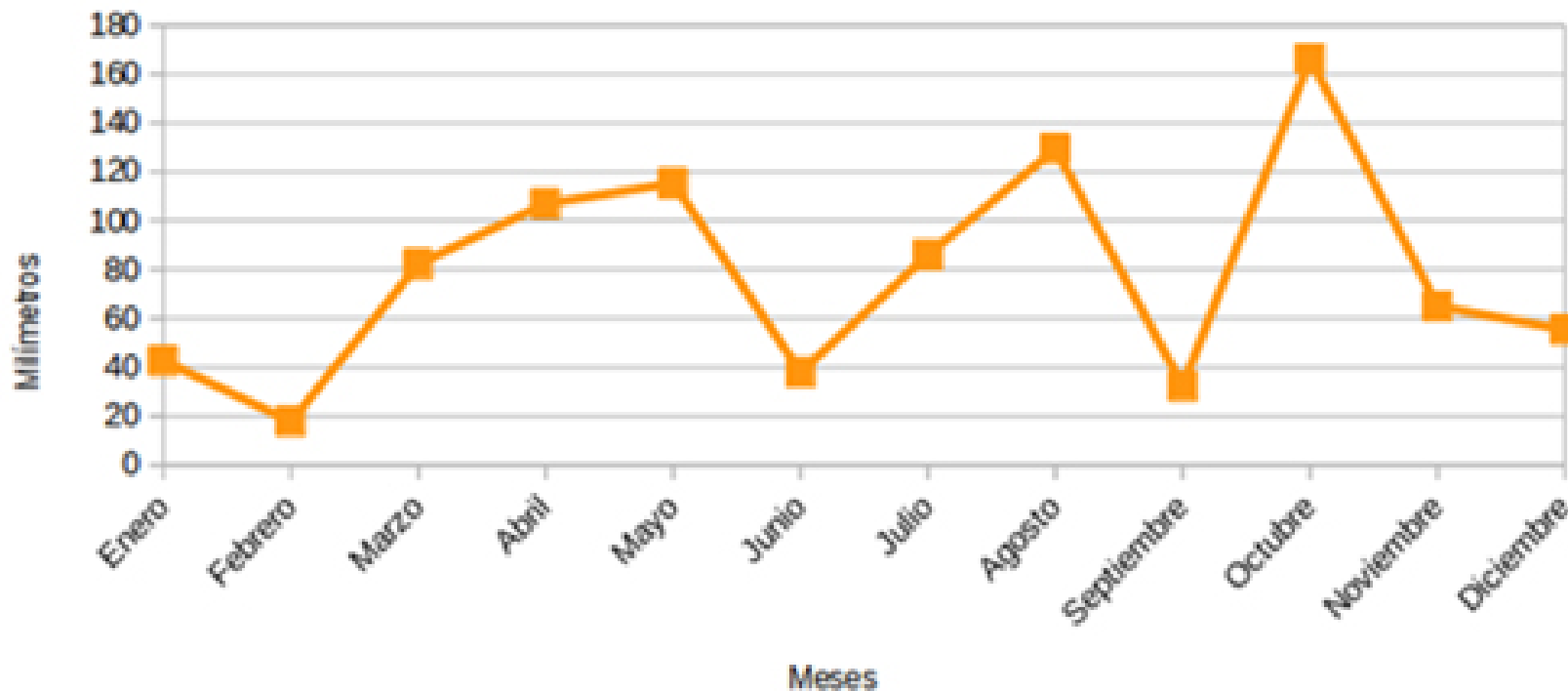


Tipos de gráficos.

Gráfico de líneas

En este tipo de gráfico se emplean líneas para **delimitar el valor de una variable dependiente respecto a otra independiente**. Gráficos de funciones.

Muy utilizado en series temporales.

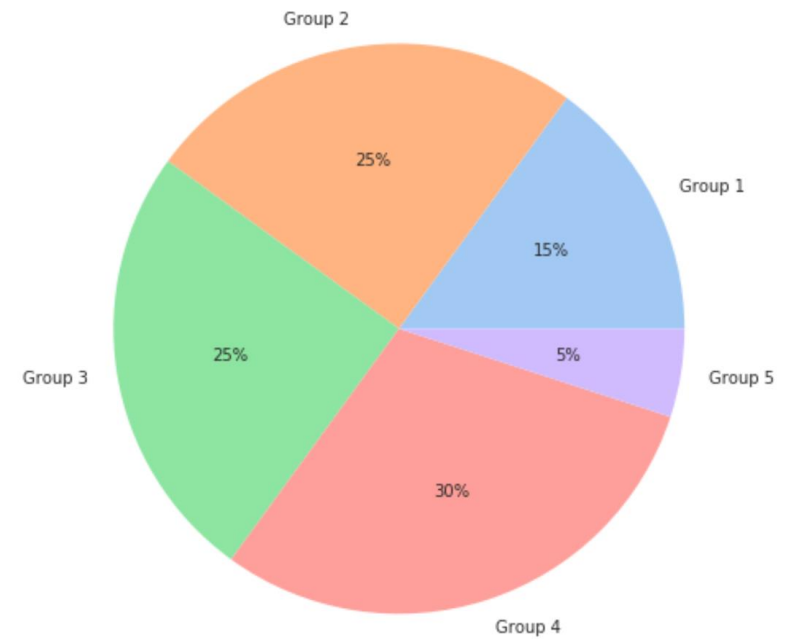
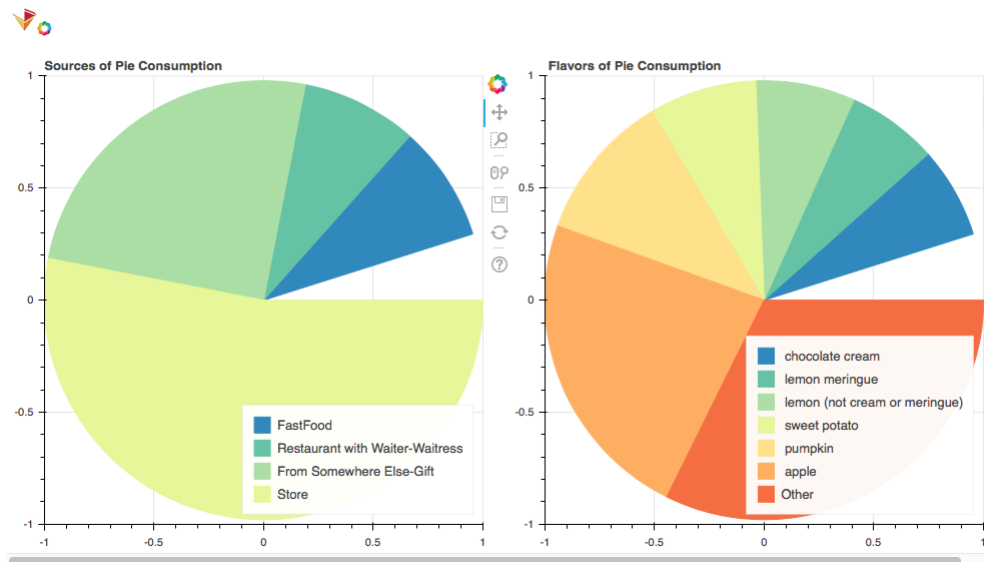


Tipos de gráficos.

Gráfico circular

División de un círculo en partes según una variable categórica.

Cada parte tiene **un tamaño proporcional a su frecuencia dentro del total de los datos.**

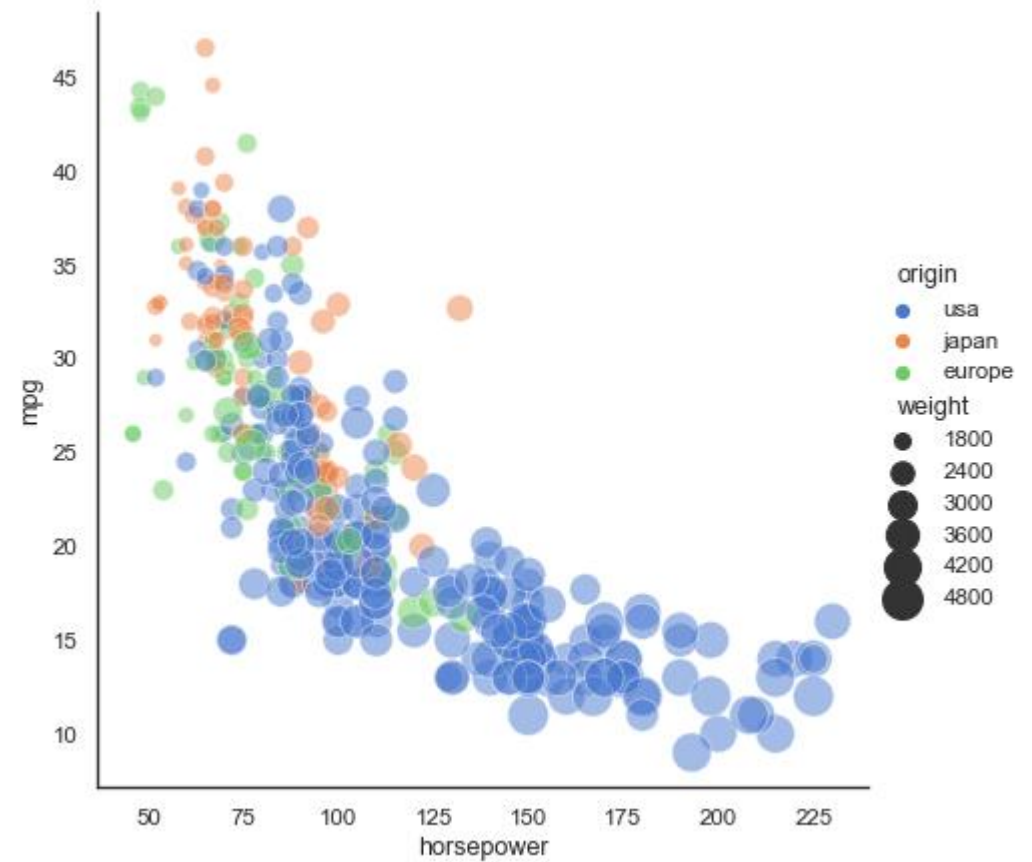
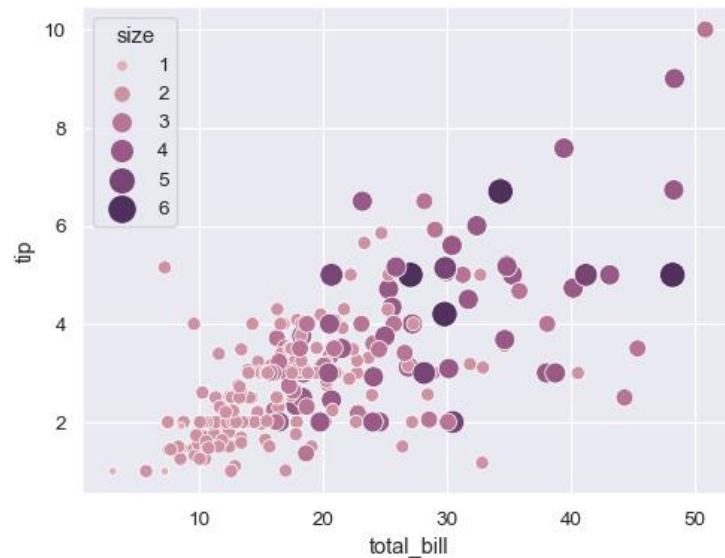


Tipos de gráficos.

Gráfico de puntos

Los ejes x e y muestran cada uno los valores de una variable dependiente y otra independiente o dos variables de la que se esté observando si presentan algún tipo de relación.

Puede no ser una función.



GitHub

Crea un repositorio en GitHub llamado **curso Machine Learning** y sube todos los ejercicios resueltos separados por carpetas con un nombre descriptivo del ejercicio.

Además, crea un archivo **README.md** donde deberás escribir:

- Lenguaje utilizado.
- Lista de ejercicios con una descripción acotada.





Contacto:

luis@lubay.es





red.es



UNIÓN EUROPEA

"El FSE invierte en tu futuro"

Fondo Social Europeo

