

ACÀMICA

# Agenda

---

Repaso: correlación y causalidad, bitácora y challenge.

Explicación: Matplotlib

**Break.**

Hands-on training

Bonus Track: gráficos engañosos.

Cierre.



# Visualizaciones

Visualizar datos es una parte fundamental del trabajo del Data Scientist. Hoy veremos cómo hacer gráficos en Python usando la librería Matplotlib.

# REPASO



# Repaso del encuentro pasado



# Correlación, Causalidad e Independencia estadística

- Tres conceptos que tratan sobre \_\_\_\_\_.
- Es muy fácil \_\_\_\_\_.
- En su uso cotidiano tienen un significado un poco más \_\_\_\_\_ que en su uso estadístico.

**COMPLETA LAS ORACIONES!**

# Correlación, Causalidad e Independencia estadística

- Tres conceptos que tratan sobre la relación entre dos variables aleatorias.
- Muy fácil confundirlos entre ellos
- En su uso cotidiano tienen un significado un poco más “laxo” que en su uso estadístico

Llamamos probabilidad de  $X$ ,  
 **$P(X)$  a un número entre 0 y 1**  
**que le asignamos a cada**  
**posible valor que puede**  
**tomar esa variable aleatoria**



Llamamos probabilidad de  $X$ ,  
 **$P(X)$  a un número entre 0 y 1**  
**que le asignamos a cada**  
**posible valor que puede**  
**tomar esa variable aleatoria**

También dijimos que la **mejor descripción** de  $P(X)$  está dada por la distribución **(densidad de probabilidades)**.

# ¿Pero qué ocurre si tengo más de una variable aleatoria?

## 2 variables

Peso y altura de una persona  
Temperatura y humedad

## 3 variables

Temperatura, humedad y velocidad del viento

## 4 variables

Iris Dataset: ancho y largo del sépalos y pétalo

# Estas variables pueden o no estar correlacionadas.

| Un ejemplo sencillo de entender es la relación el peso y la altura: ¿pensás que hay correlación?

| Las personas más altas suelen ser más pesadas: ¿esto implica que existe una correlación?

# Conclusiones

- Correlación no implica causalidad
- La correlación de Pearson es muy útil para encontrar correlaciones lineales
- Si la relación entre las variables NO es lineal, existen otras correlaciones que pueden ser útiles: Spearman y Kendall

## REPASO

# EJERCICIO DEL ENCUENTRO PASADO

¡Muéstranos qué hiciste!

¿Qué cosas te costaron más del ejercicio? ¿Cómo las resolviste?

¿Cuál el principal aprendizaje que te llevas?

Si tuvieras que hacerle alguna recomendación a alguien que va a hacer el ejercicio por primera vez, ¿qué le dirías?



## EJERCICIO DEL ENCUENTRO PASADO

¿Alguien hizo algo diferente que quiera mostrar?



# Repaso de la bitácora



REPASO

TEMAS BITÁCORA





# Visualización de **DATOS**



# ¿Por qué visualizar?

**Visualizar** los datos es una parte fundamental del análisis en ciencia de datos.



## ¿Por qué visualizar?

No solo sirve para **comunicar** (que es una parte fundamental del trabajo) sino que también es una herramienta esencial para **comprender el dataset** con el que estamos trabajando.

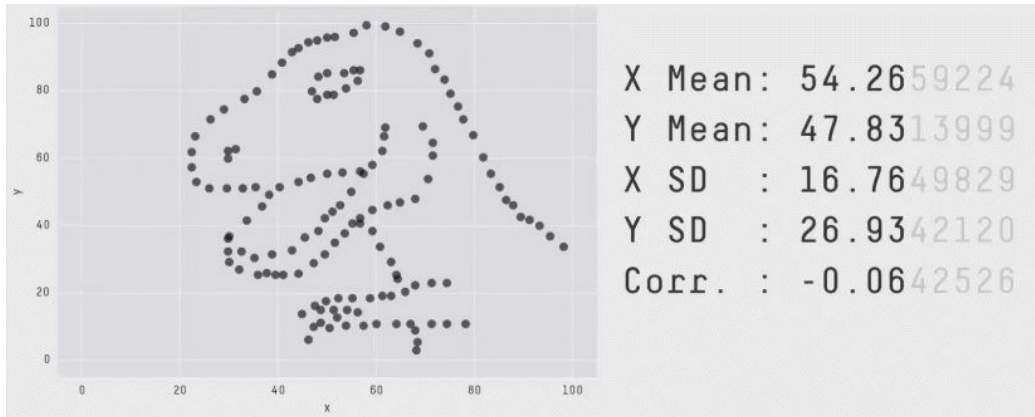
## ¿Por qué visualizar?

Hay veces que sólo indicadores numéricos no alcanzan para describir las características principales de nuestro dataset.

# ¿Por qué visualizar?

Hay veces que sólo indicadores numéricos no alcanzan para describir las características principales de nuestro dataset.

Conjuntos de datos con tres métricas en común: su media, su varianza y correlación.



## CHALLENGE BITÁCORA



1. ¿Buscaste algún tipo de gráfico que **no hayamos visto** en la bitácora?
2. ¿Resolviste el challenge del notebook?



## CHALLENGE BITÁCORA



¿Alguien hizo algo diferente que quiera mostrar?



# Matplotlib





# Herramientas de Visualización



# Herramientas de Visualización



**antes de empezar...**

# Introducción a Matplotlib

Por si no la tienen instalada de clases anteriores:

**conda install matplotlib**

# Introducción a Matplotlib

La documentación de matplotlib es **excelente**, es importantísimo aprovecharla:

<https://matplotlib.org/index.html>

**¡ahora sí!**

**matplotlib**

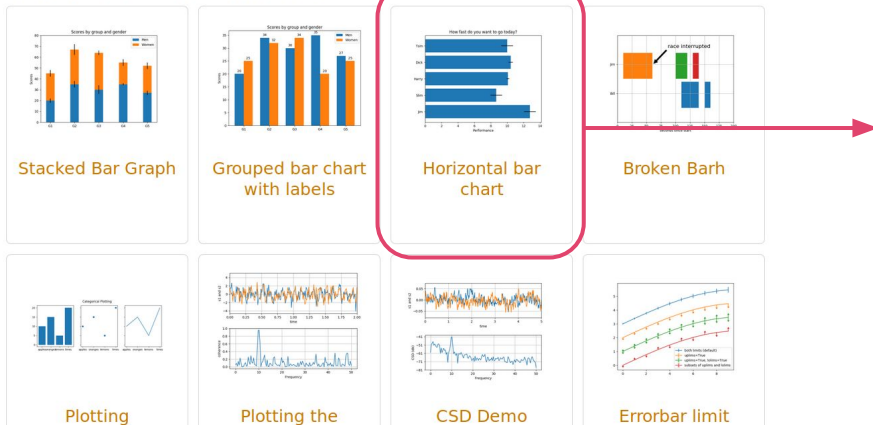
# Documentación

## Gallery

This gallery contains examples of the many things you can do with Matplotlib. Click on any image to see the full code.

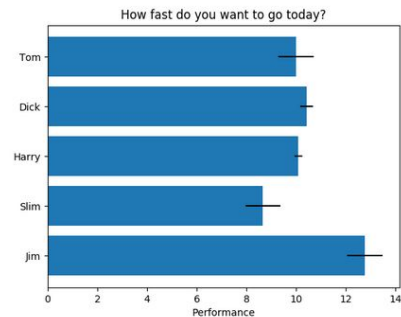
For longer tutorials, see our [tutorials page](#). You can also find [external resources](#) and a [FAQ](#) in our [user guide](#).

## Lines, bars and markers



## Horizontal bar chart

This example showcases a simple horizontal bar chart.



```
import matplotlib.pyplot as plt
import numpy as np

# Fixing random state for reproducibility
np.random.seed(19680801)

plt.rcdefaults()
fig, ax = plt.subplots()

# Example data
people = ('Tom', 'Dick', 'Harry', 'Slim', 'Jim')
y_pos = np.arange(len(people))
performance = 3 + 10 * np.random.rand(len(people))
error = np.random.rand(len(people))

ax.barh(y_pos, performance, xerr=error, align='center')
ax.set_yticks(y_pos)
ax.set_yticklabels(people)
ax.invert_yaxis() # labels read top-to-bottom
ax.set_xlabel('Performance')
ax.set_title('How fast do you want to go today?')

plt.show()
```



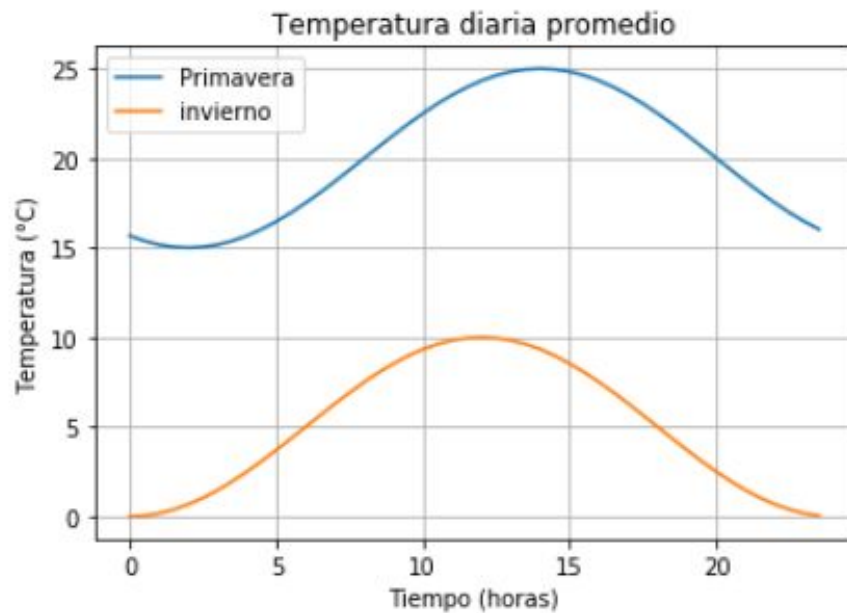
# Graficar en un notebook

La parte de la librería que usaremos para graficar es **matplotlib.pyplot** y se suele importar con el nombre **plt**

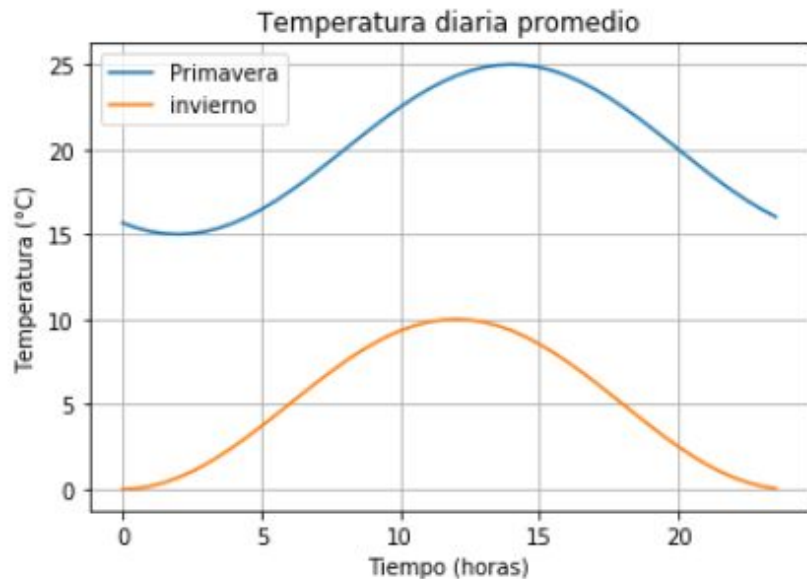
```
[ ]: import matplotlib.pyplot as plt
```



# Veamos cómo graficar este cuadro:



# Veamos cómo graficar este cuadro:

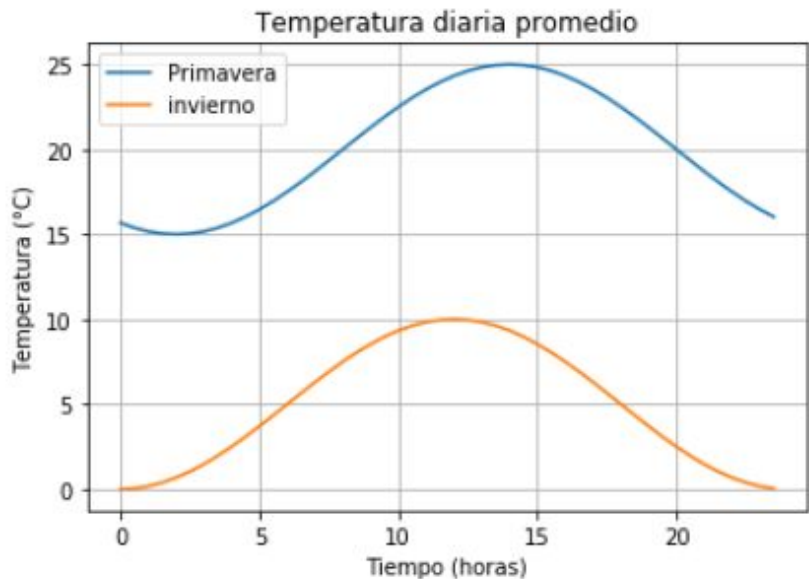


Escribimos los valores a graficar



```
[8]: x = np.arange(0.0, 24, 0.5)
      y1 = 20 + 5*np.sin(2*np.pi*(x - 8)/24)
      y2 = 5 + 5*np.sin(2*np.pi*(x - 6)/24)
```

# Veamos cómo graficar este cuadro:

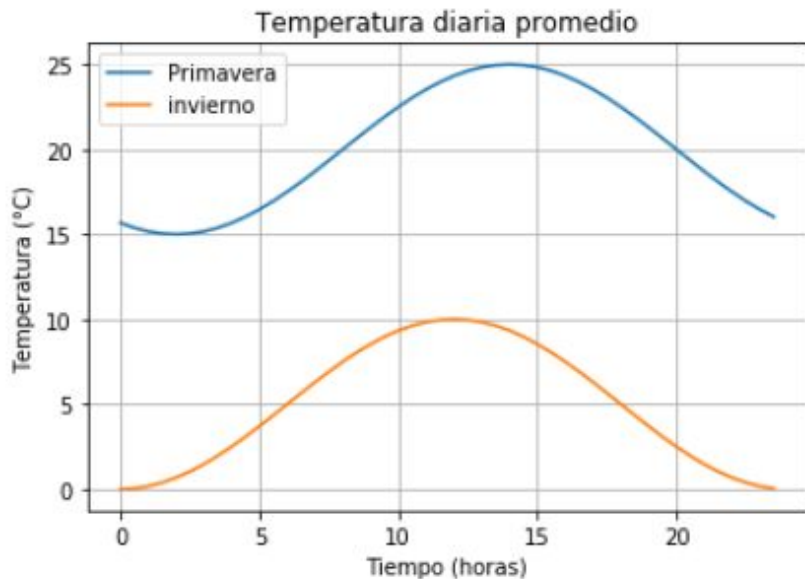


Ploteamos las dos líneas, dándole un nombre a cada una mediante el argumento "label".



```
plt.plot(x, y1, label = 'Primavera')  
plt.plot(x, y2, label = 'invierno')
```

# Veamos cómo graficar este cuadro:

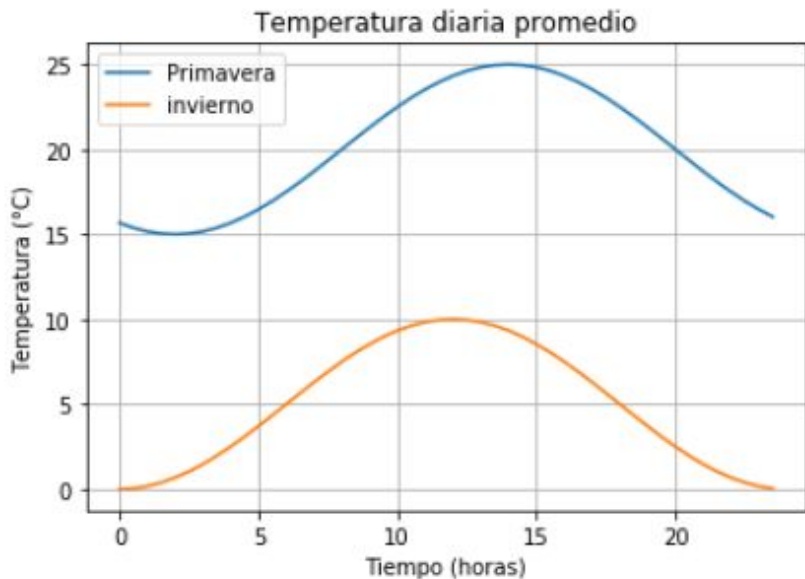


Seteamos los títulos a cada eje.



```
plt.xlabel('Tiempo (horas)')  
plt.ylabel('Temperatura (°C)')  
plt.title('Temperatura diaria promedio')
```

# Veamos cómo graficar este cuadro:



```
plt.grid()
```

Activamos la grilla.

```
plt.legend()
```

Agregamos la leyenda.

```
plt.show()
```

¡Mostramos el gráfico!

# Veamos cómo graficar este cuadro:

```
[8]: x = np.arange(0.0, 24, 0.5)
     y1 = 20 + 5*np.sin(2*np.pi*(x - 8)/24)
     y2 = 5 + 5*np.sin(2*np.pi*(x - 6)/24)

     # Ploteamos las dos lineas, dandole un nombre a cada una mediante el parámetro 'label'.
     # Notar que para agregar una curva, simplemente debemos poner una debajo de la otra
     plt.plot(x, y1, label = 'Primavera')
     plt.plot(x, y2, label = 'invierno')

     plt.xlabel('Tiempo (horas)')
     plt.ylabel('Temperatura (°C)')
     plt.title('Temperatura diaria promedio')

     # Este comando enciende la grilla de referencia
     plt.grid()

     # Agregamos la leyenda al gráfico
     plt.legend() #loc='upper center', shadow=True, fontsize='x-large')
     plt.show()
```

A close-up photograph of a white ceramic cup filled with a latte. The surface of the milk is decorated with intricate latte art, featuring a central heart shape surrounded by concentric, wavy lines. The cup is placed on a matching white saucer. In the background, a white napkin and a silver spoon are visible, though they are out of focus. The overall lighting is soft and even, highlighting the textures of the coffee and the smooth surface of the cup.

**¡BREAK!**

---





# Hands-on training

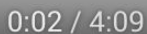


# DS\_Bitácora\_06\_Matplotlib.ipynb



# Visualizar datos también puede ser engñoso...

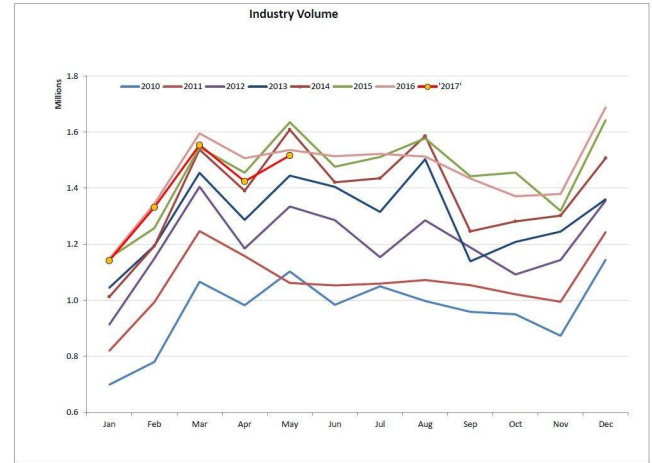
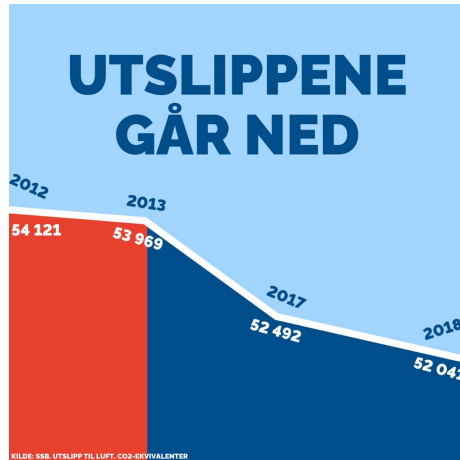




## Cómo detectar un gráfico engañoso - Lea Gaslowitz

# Ejercitación

¿Crees que los siguientes gráficos son engañosos o están mal hechos?  
Explica por qué.



## Ejercitación

En grupos de 3 personas, pensar y hacer una visualización con datos de la coyuntura actual que sea engañoso o esté malinterpretado.

Por ejemplo, pueden usar datos sobre encuestas, intención de votos en elecciones, datos sobre la pandemia de Covid-19, etc.

¡No hace falta programar! Pueden hacerlo en lápiz y papel.



# Recursos



# Probabilidad y Estadística

- [Python Data Science Handbook](#) - Capítulo 4, “Visualization with Matplotlib”.



# Para la próxima

---

- Termina el notebook de hoy.
- Lee la bitácora 07 y carga las dudas que tengas al Trello.
- Resuelve el Challenge.

En el encuentro que viene uno/a de ustedes será seleccionado/a para mostrar cómo resolvió el challenge de la bitácora. De esta manera, ¡aprendemos todos/as de (y con) todas/as, así que vengan preparados/as.

ACÀMICA