

Introducción a la Visualización de Datos

**Análisis de Datos con
Python**

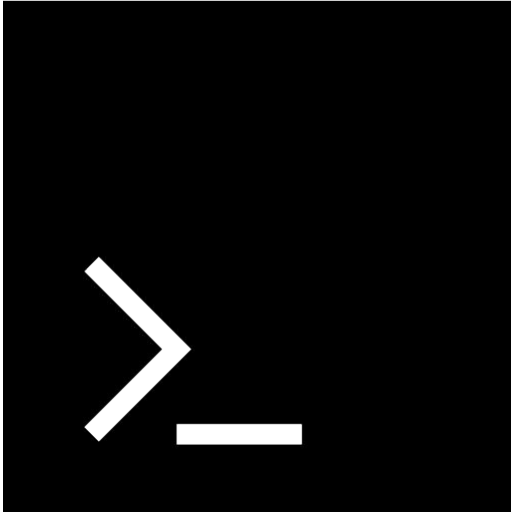
L. en C.C. Manuel Soto Romero



- Comprender el concepto de distribución e identificar la distribución de nuestros datos junto con su importancia.
- Utilizar la biblioteca Seaborn.
- Conocer los boxplots y aprender a generarlos.
- Conocer las tablas de frecuencias y los histogramas como maneras de visualizar distribuciones.
- Clasificar algunas de las formas que generan los histogramas.
- Conocer las gráficas de densidad como una alternativa a los histogramas clásicos.



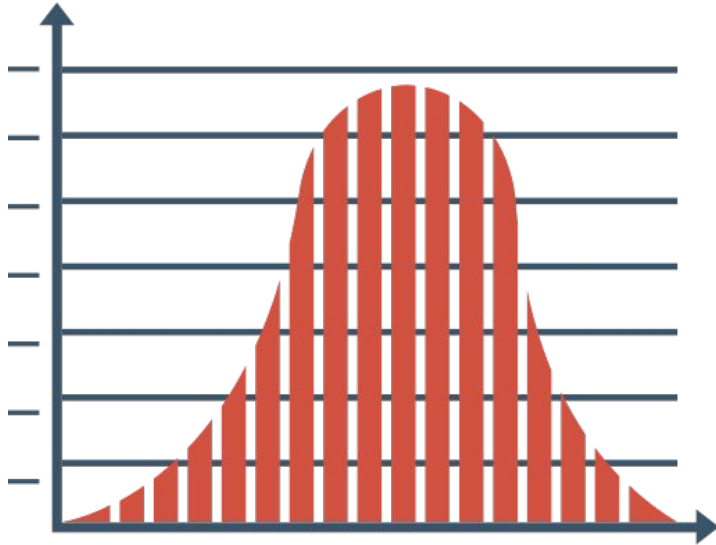
- Tipos de datos estructurados
- Medidas de tendencia central
- Desviación estándar
- Medidas de dispersión
- Medidas de posición



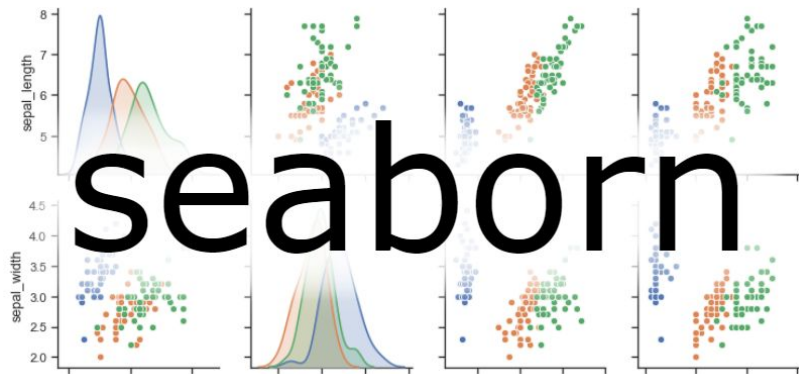
¡No olvides hacer pull del repo!

El material de la sesión se encuentra ahí.

```
git pull origin master
```



- Como vimos en la sesión anterior, los datos pueden adoptar muchas formas:
 - Pueden estar cerca del promedio
 - Cerca del valor mínimo
 - Cerca del valor máximo
 - Completamente dispersos
- Ya los analizamos con métodos estadísticos
- Ahora los analizaremos mediante visualización



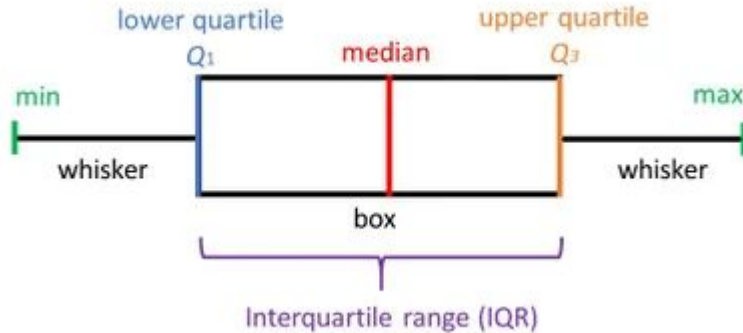
- Es una biblioteca de **Python** para la visualización de datos.
- Está basada en **matplotlib**.

<https://seaborn.pydata.org/>

<https://towardsdatascience.com/data-visualization-using-seaborn-fc24db95a850>

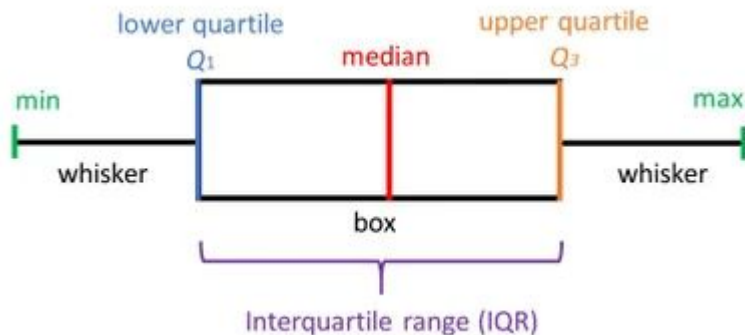
```
pip install seaborn
```

Diagrama de caja



- Es un método de representar gráficamente una serie de datos a través de sus cuartiles.
- Muestra la mediana y los cuartiles asociados.
- También permite revisar de cerca algunos de los valores atípicos a través de los bigotes.

Diagrama de caja



El rango intercuartil es el rango entre el percentil 25 y el percentil 75.

Los bigotes en general se calculan como $1.5 * \text{RIC}$

Se componen de:

- Rango (sin datos atípicos)
- Datos atípicos.
- Rango intercuartil (también conocido como RIC)
- Cuartiles (denotados como Q1, Q2 y Q3)
- Mediana (Q2)
- Mínimo y máximo.

[Ve al Ejemplo 1](#)

[Ve al Reto 1](#)

Tomemos un descanso



EN 10 MINUTOS CONTINUAMOS



Encuentra la figura repetida (forma y número)

64	93	69	89	40
76	59	47	29	72
23	52	70	16	42
65	76	39	30	63
52	90	71	26	22

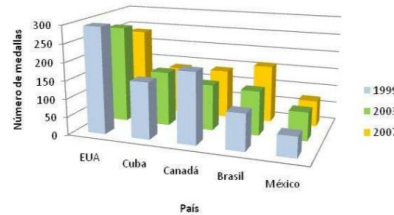
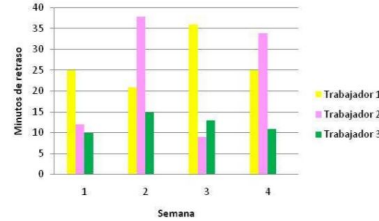
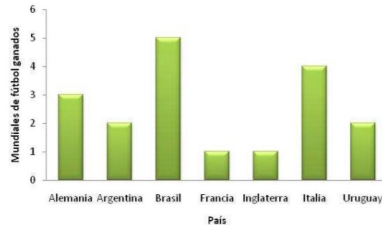
Number of Pets	Number of Students
0	2
1	7
2	3
3	1
4	2

- Permiten seccionar los datos en segmentos.
- Una forma es contabilizar cuántos datos hay por cada posible valor de la columna.
- Otra forma es calcular el porcentaje.
- Lo más idóneo con variables numéricas es seccionar por segmentos de varios valores.

[Ve al Ejemplo 2](#)

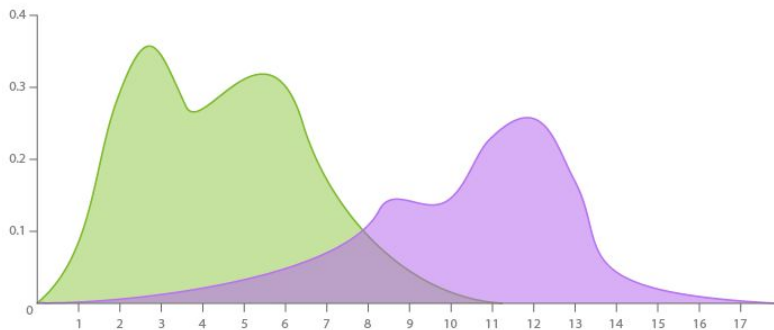
[Ve al Reto 2](#)

Histogramas



- Cada barra sólida, ya sea vertical u horizontal representa un tipo de dato.
- No muestran frecuencias acumuladas, son preferibles para el tratamiento de datos cuantitativos y la barra con mayor altura representa la mayor frecuencia.
- La suma de las alturas de las columnas equivale al 100% de los datos.

Ve al Ejemplo 3



[Ve al Ejemplo 5](#)

- Permiten visualizar la distribución de datos en un intervalo o período de tiempo continuo.
- Este gráfico es una variación de un Histograma que usa el suavizado de cerner para trazar valores, permitiendo distribuciones más suaves al suavizar el ruido.
- Los picos de un gráfico de densidad ayudan a mostrar dónde los valores se concentran en el intervalo.

[Ve al Reto 4](#)



NO OLVIDES REVISAR TU
POSTWORK Y TU PREWORK



Preguntas

