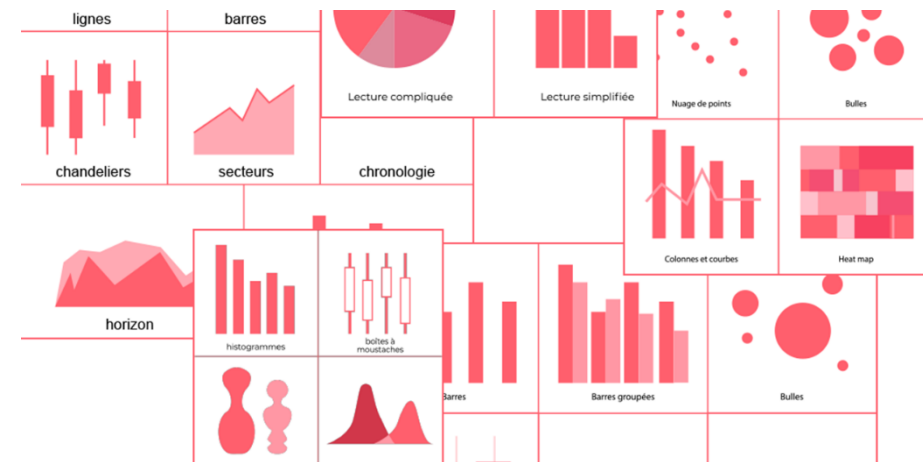


# L'analyse exploratoire avec Matplotlib

Les différent types de graphiques.



# Types de graphiques selon les types de variables

Les types de variables :

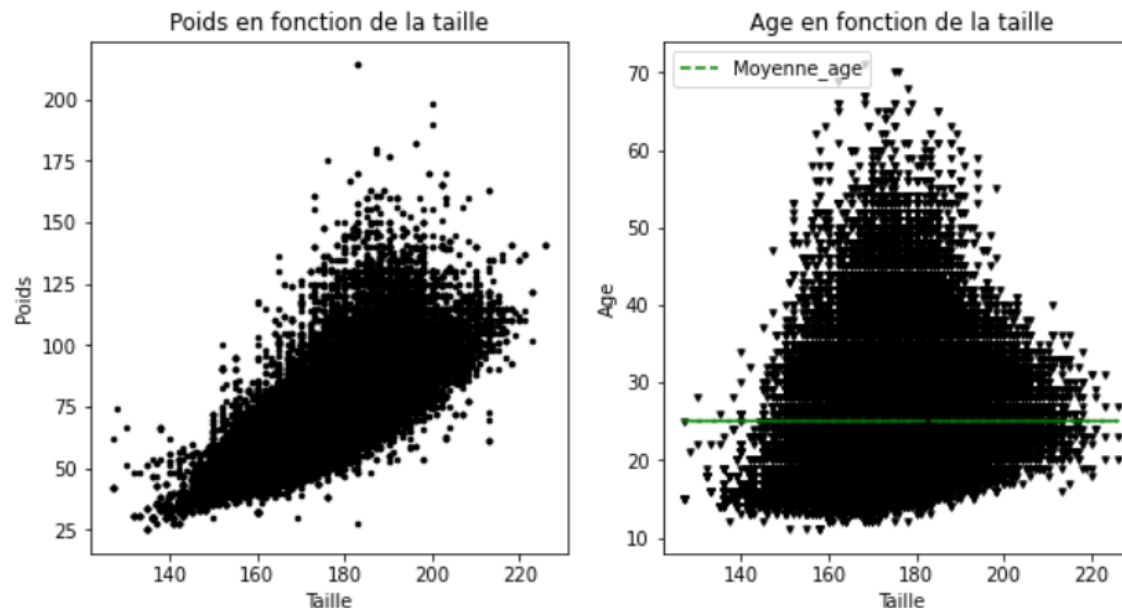
- Variables quantitatives : valeurs quantifiables à valeur continue ou discrète -> l'Age, la taille, le poids des athlètes.
- Variables catégorielles ou qualitatives : des colonnes de dataframe contenant des modalités -> sexe, équipe, saison, etc...

Différents types de variables implique différents types de graphiques.

# Types de graphiques selon les types de variables

## Exemple : les nuages de points :

Par exemple, les nuages de points que nous tracions précédemment se faisaient sur les variables âge, poids et taille, car un nuage de points ne peut se faire que sur des valeurs quantitatives.



## Scatterplot :

Le scatterplot (ou nuage de points en français) est un type de graphique très utilisé pour représenter deux variables quantitatives (alternative à la fonction `plot()`).

- Le but de ce type de graphique est d'avoir deux axes x et y quantitatifs, avec des nombres, et de placer des points aux coordonnées x et y.
- Il existe une fonction dédiée pour créer des scatterplots, qui s'appelle `scatter()`.
- Si vous souhaitez essayer de trouver une relation entre deux variables qualitatifs, le nuage de points est le graphique à utiliser.

## Graphique à barres (bargraph)

### a. Graphique à barres simple

- Le bargraph (ou graphique à barres, en français) est le graphique le plus utilisé pour représenter des variables qualitatives.
- Le principe du bargraph est que chaque valeur  $x$  correspond aux modalités d'une catégorie et on va représenter le nombre de fois que cette modalités apparait.

Exemple : pour la variable qualitative Sport, l'axe  $x$  comprendra les valeurs Athletics, Swimming, Shooting, etc., et la valeur  $y$  correspondant à chaque  $x$  sera le nombre de fois où cette modalité  $x$  est présente dans la variable qualitative choisie.

## b. Graphique à barres groupées

Il peut être intéressant de représenter une variable qualitative sur un graphique à barres et de séparer les barres selon une deuxième variable qualitative, pour créer des sous-groupes.

On parle aussi de barres groupées.

Exemple : On souhaite séparer le nombre d'athlètes dans les dix disciplines les plus populaires par leur genre. Pour cela, nous utiliserons toujours la fonction `bar()` de Pyplot, mais il faudra coder quelques lignes plus complexes pour arriver à ce résultat de barres groupées.

# Autre graphes

- L'histogramme est une représentation graphique permettant de représenter la répartition d'une variable continue en la représentant avec des colonnes verticales.
- Les diagramme circulaire d'usage courant en économie et le monde des affaires (pour leur présentation « élégante ») ou dans les médias (presse/magazines, télévision, Web, résultats de sondages...), ils le sont moins dans le milieu scientifique car ils représentent moins bien les données et facilitent moins leurs comparaison que les autres types de diagrammes. En informatique, ils sont notamment utilisés pour représenter l'espace disque occupé/disponible. La représentation de nombres négatifs est impossible avec ce type de diagramme.

## Boxplots :

Les boxplots (ou boîte à moustaches, en français) permettent de visualiser la répartition des valeurs d'une variable quantitative, en permettant de connaître :

- la valeur médiane de cette variable
- la valeur minimale
- la valeur maximale
- les quartiles
- les outliers

Les boîtes à moustaches sont très pratiques pour avoir une première idée de la distribution des valeurs dans une variable quantitative.

Ce graphique ne se fait uniquement que sur des variables quantitatives, sur lesquelles on peut faire des opérations numériques, comme l'âge.

