

# Visualisation des données

Atelier théorique #5





# Pourquoi visualiser ses données?

- Les statistiques descriptives permettent de faire des calculs de base sur les données brutes d'un jeu de données.
- Elles ne sont pas toujours suffisantes pour se bâtir une idée concrète.
- La visualisation permet de voir des détails qu'il serait impossible de cerner avec des tableaux et des statistiques.
- Les statistiques descriptives et les visualisations sont complémentaires.

Statistiques  
descriptives

+

Visualisation

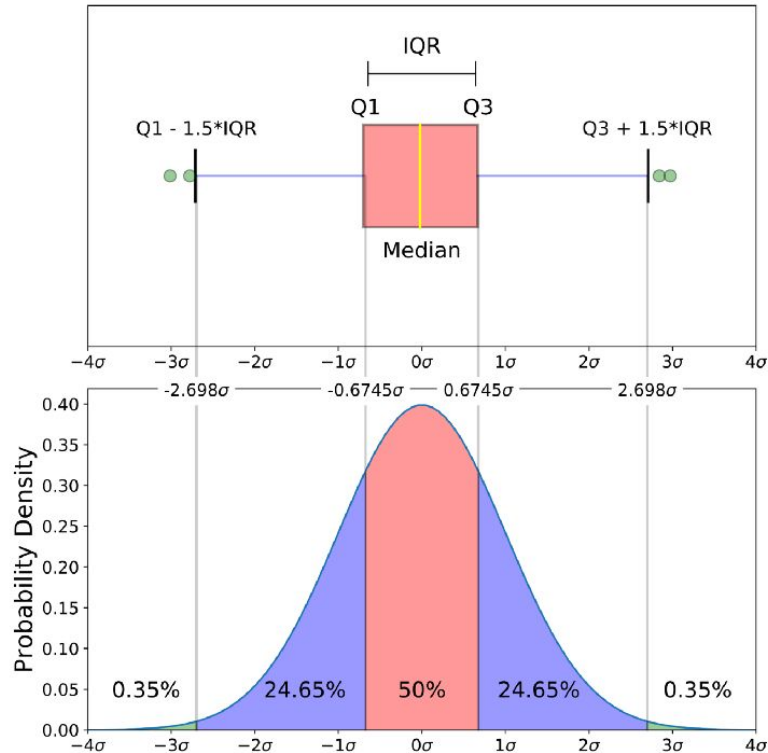
=

Analyses  
exploratoires des  
données (EDA)



# Distribution d'une boîte à moustache

- Importante relation à faire entre une boîte à moustache et le diagramme de distribution.
- Différentes façons de visualiser les mêmes données tout en obtenant des informations similaires.





# Librairies Python pour la visualisation

Trio pour les visualisations : Pour optimiser les options de visualisation dans Python, 3 librairies peuvent être utilisées ensemble.

- pandas : Visualisations de bases, rapide à produire et simple à programmer
- matplotlib : Librairie centrale, très flexible, plus complexe à programmer
- seaborn : Développer à partir de matplotlib, améliore l'aspect esthétique et visuel

Librairies essentielles





# Liens vers les librairies de visualisation

## Essentiel d'utiliser la documentation web pour utiliser ces librairies

- pandas : <https://pandas.pydata.org/>
- matplotlib : <https://matplotlib.org/>
- seaborn : <https://seaborn.pydata.org/>

## Librairies essentielles

