

Identifier les types de représentations graphiques

Titre du cours : Analyse exploratoire des données

Code officiel : 420-A55-BB

Professeur : Dr Komi SODOKE

Plan

I- Introduction

II- Identifier les types de graphiques

- A. Graphique en barres
- B. Histogramme
- C. Diagramme circulaire (camembert)
- D. Graphique en ligne (courbe)
- E. Nuage de points (scatter plot)
- F. Boîte à moustaches (boxplot)

III- Identifier les types de données dans vos graphiques

IV- Associer le type de représentation graphique aux types de données

V- Conclusion

Introduction

Dans l'analyse et l'interprétation des données, la représentation graphique joue un rôle fondamental. Elle permet de visualiser les tendances, les variations et les relations entre les variables de manière claire et intuitive. Cependant, il est essentiel de choisir le bon type de graphique en fonction du type de données analysé afin d'éviter les erreurs d'interprétation.

Ce chapitre explore les différents types de graphiques, les types de données qu'ils permettent de représenter et les règles de sélection des représentations les plus adaptées.

II-Identifier les types de graphiques

Les graphiques sont des outils de communication visuelle qui permettent de représenter des données sous différentes formes. Voici les types les plus couramment utilisés :

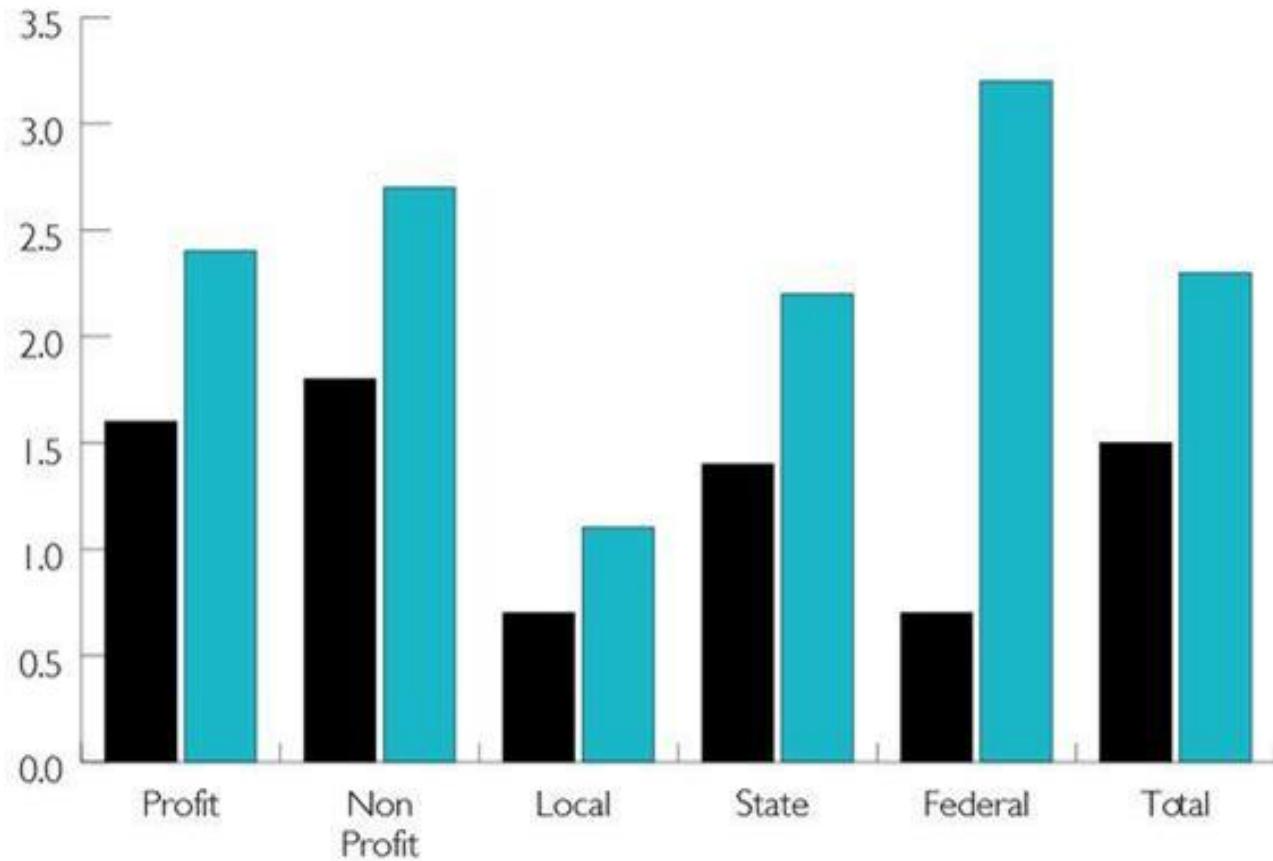
A. Graphique en barres

- Définition** : Un graphique en barres est constitué de barres verticales ou horizontales de hauteur (ou de longueur) proportionnelle à la valeur qu'elles représentent.
- Utilisation** : Il est particulièrement efficace pour comparer des valeurs associées à des catégories distinctes.
- Exemple** : Un graphique en barres peut être utilisé pour comparer le nombre d'étudiants inscrits dans différentes filières universitaires (informatique, droit, médecine, économie).

II-Identifier les types de graphiques

A. Graphique en barres

- **Caractéristiques clés :**
 - **Axe des catégories** : Un axe représente les différentes catégories comparées.
 - **Axe des valeurs** : L'autre axe représente les valeurs associées à chaque catégorie.
 - **Barres** :
 - La longueur ou la hauteur de chaque barre est proportionnelle à la valeur qu'elle représente.
 - Les barres peuvent être verticales ou horizontales.



II-Identifier les types de graphiques

B. Histogramme

Définition : Un histogramme est une représentation graphique qui visualise la distribution d'une variable quantitative continue. Contrairement au diagramme en barres, qui compare des catégories distinctes, l'histogramme regroupe les données en intervalles contigus appelés classes ou bins. La hauteur de chaque rectangle (ou barre) représente la fréquence (ou le nombre) de données qui se trouvent dans chaque classe.

Utilisation :

- **Visualisation de la distribution :**
 - L'histogramme permet de voir rapidement la forme de la distribution, c'est-à-dire comment les données sont réparties.
 - On peut identifier si la distribution est symétrique, asymétrique, unimodale (un seul pic) ou bimodale (deux pics).

II-Identifier les types de graphiques

B. Histogramme

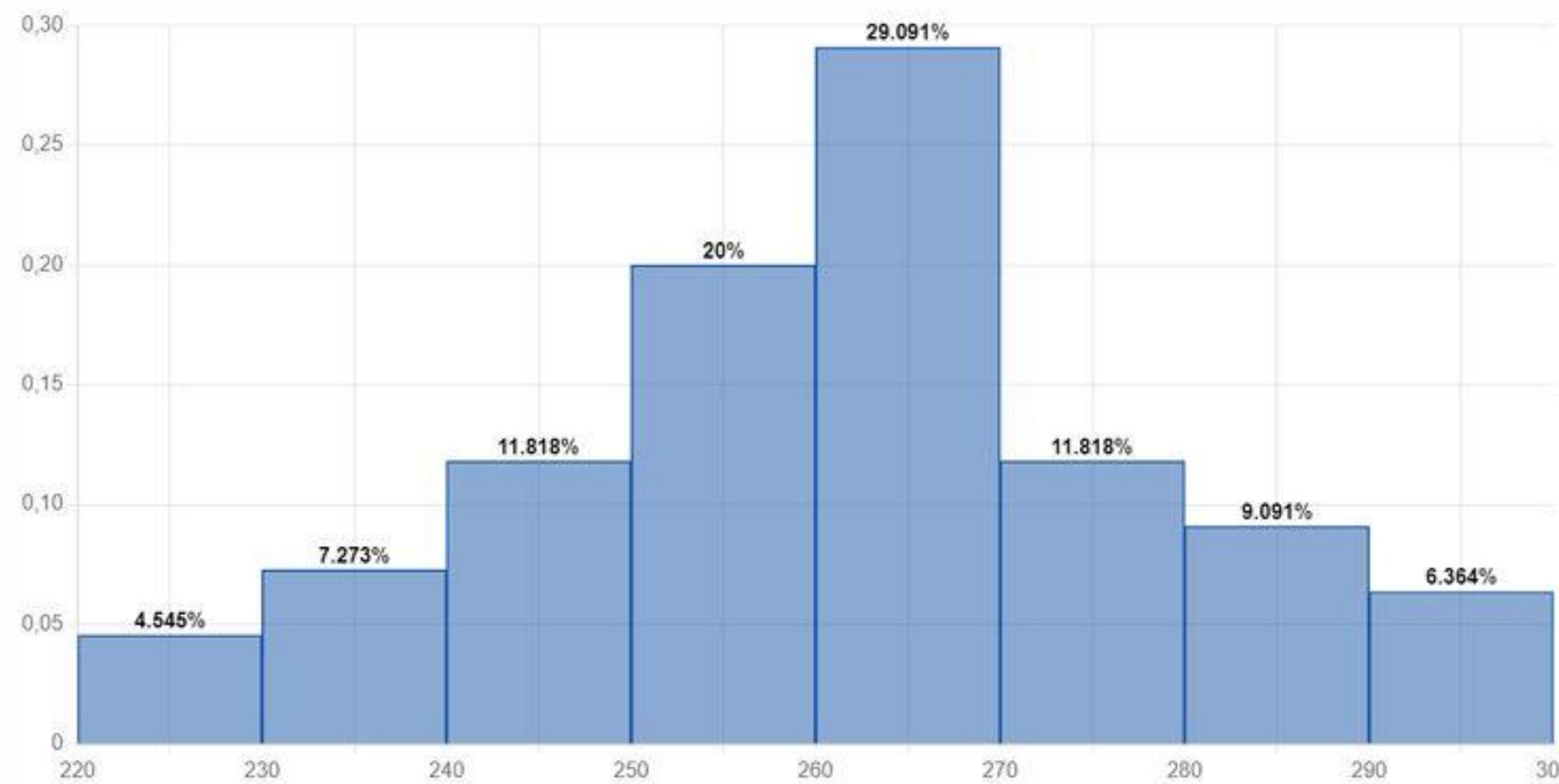
- Identification des valeurs aberrantes
- Analyse de la dispersion

Caractéristiques clés :

- Axe des abscisses (x) : Représente les intervalles de valeurs de la variable continue.
- Axe des ordonnées (y) : Représente la fréquence (ou le nombre) de données dans chaque intervalle.
- Rectangles (ou barres) : La hauteur de chaque rectangle est proportionnelle à la fréquence des données dans l'intervalle correspondant.

Exemples :

- Répartition des tailles des élèves d'une classe.
- Distribution des revenus d'une population.
- ...



II-Identifier les types de graphiques

C. Diagramme circulaire (camembert)- piechart

Définition : Un diagramme circulaire, également appelé camembert, est une représentation graphique qui divise un cercle en secteurs. Chaque secteur représente une partie d'un ensemble, et la taille de chaque secteur est proportionnelle à la valeur qu'il représente. En d'autres termes, plus la valeur d'une catégorie est importante, plus le secteur correspondant sera grand.

Utilisation :

- **Répartition des catégories**
- **Comparaison des proportions**
- **Données catégorielles :** Ils sont particulièrement adaptés aux données catégorielles, c'est-à-dire aux données qui peuvent être classées en catégories distinctes.

II-Identifier les types de graphiques

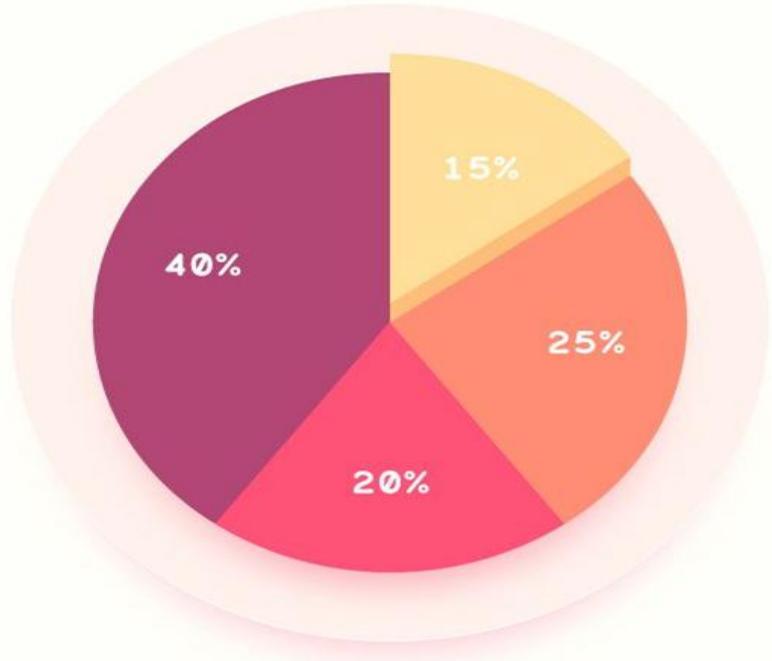
C. Diagramme circulaire (camembert)

Caractéristiques clés :

- **Cercle** : Représente l'ensemble des données.
- **Secteurs** : Représentent les différentes catégories.
- **Angle** : L'angle de chaque secteur est proportionnel à la valeur de la catégorie correspondante.
- **Pourcentages** : Les secteurs sont souvent accompagnés de pourcentages pour indiquer la proportion de chaque catégorie.

Exemples :

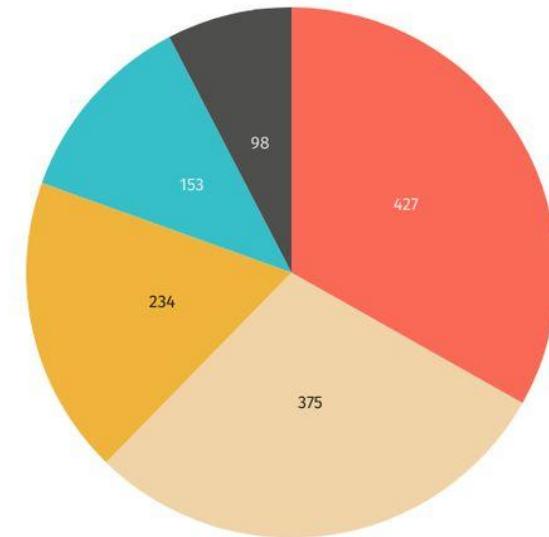
- Répartition des dépenses d'un ménage (logement, alimentation, transport, loisirs).
- Répartition des parts de marché d'une entreprise.
- Répartition des votes lors d'une élection.
- ...



Lemon Orange Raspberry Grape

PRODUCT SALES

iPad iPhone Macbook Mac Peripherals



II-Identifier les types de graphiques

D. Graphique en ligne (courbe)

Définition : Un graphique en ligne, également appelé courbe, est une représentation visuelle qui relie une série de points de données par des segments de ligne. Chaque point de données représente une valeur à un moment précis, et la ligne qui les relie montre l'évolution de cette valeur au fil du temps ou d'une autre variable continue.

Utilisation :

- **Tendances temporelles :** Les graphiques en ligne sont particulièrement efficaces pour visualiser l'évolution d'une variable au fil du temps. Ils permettent d'identifier les tendances, les fluctuations et les cycles.
- **Comparaison de tendances :** Ils peuvent être utilisés pour comparer l'évolution de plusieurs variables sur la même période.

II-Identifier les types de graphiques

D. Graphique en ligne (courbe)

- **Prévision :** Les tendances observées dans un graphique en ligne peuvent être utilisées pour faire des prévisions sur les valeurs futures.

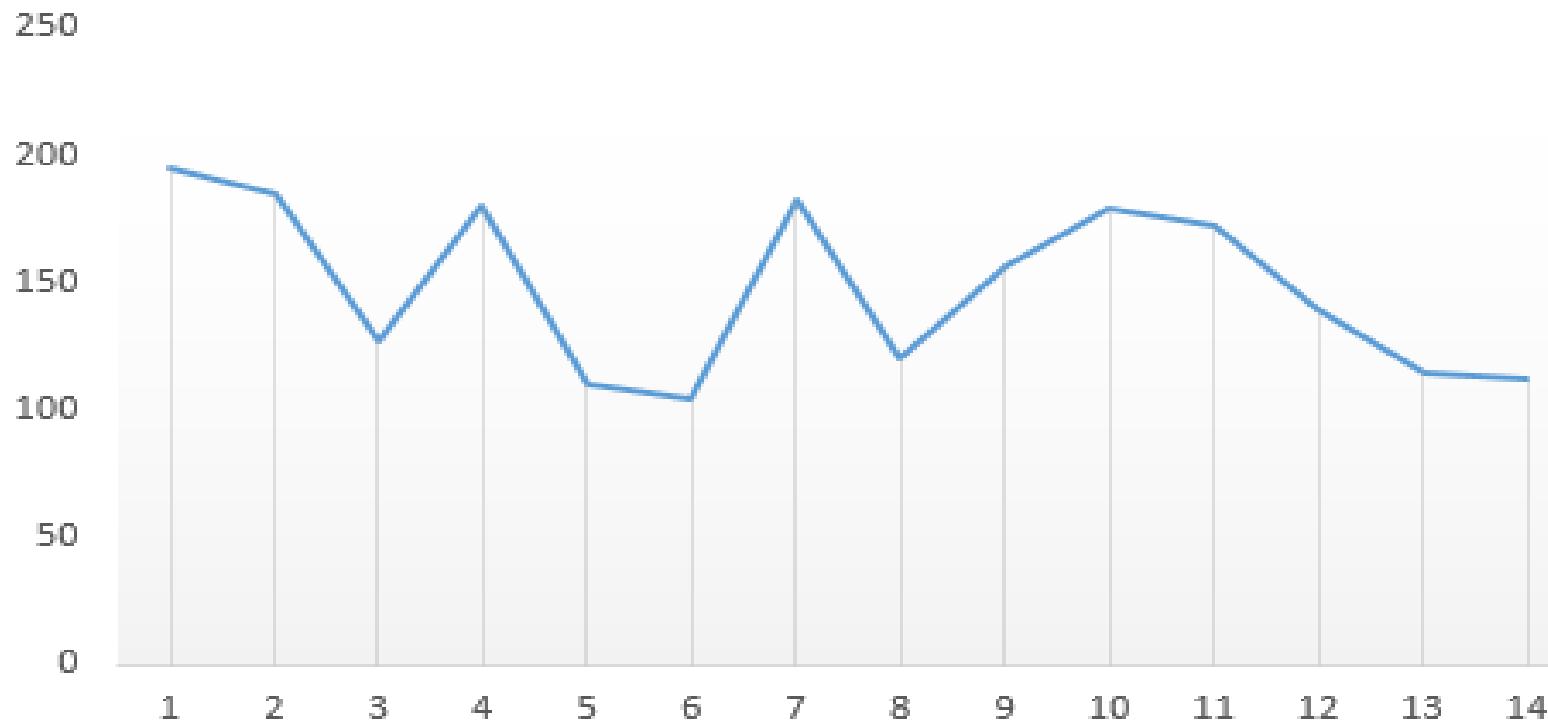
Caractéristiques clés :

- **Axe des abscisses (x) :** Représente généralement le temps ou une autre variable continue.
- **Axe des ordonnées (y) :** Représente la variable dont l'évolution est étudiée.
- **Points de données :** Représentent les valeurs mesurées à différents moments.
- **Ligne :** Relie les points de données et montre l'évolution de la variable.

Exemples :

- Évolution des cours de la bourse au fil des jours.
- Variations de la température au cours d'une journée.
- Croissance de la population d'une ville au fil des années.

Drop lines



II-Identifier les types de graphiques

E. Nuage de points (scatter plot)

Définition :

Un nuage de points, également appelé diagramme de dispersion, est une représentation graphique qui affiche des points sur un plan cartésien. Chaque point représente une paire de valeurs pour deux variables quantitatives. La position du point sur le plan est déterminée par les valeurs des deux variables.

Utilisation :

- **Analyse de la corrélation :**
 - Le nuage de points est principalement utilisé pour visualiser la relation entre deux variables continues.
- **Identification des tendances :**
 - Il permet de repérer des tendances linéaires, non linéaires, ou l'absence de tendance entre les variables.

II-Identifier les types de graphiques

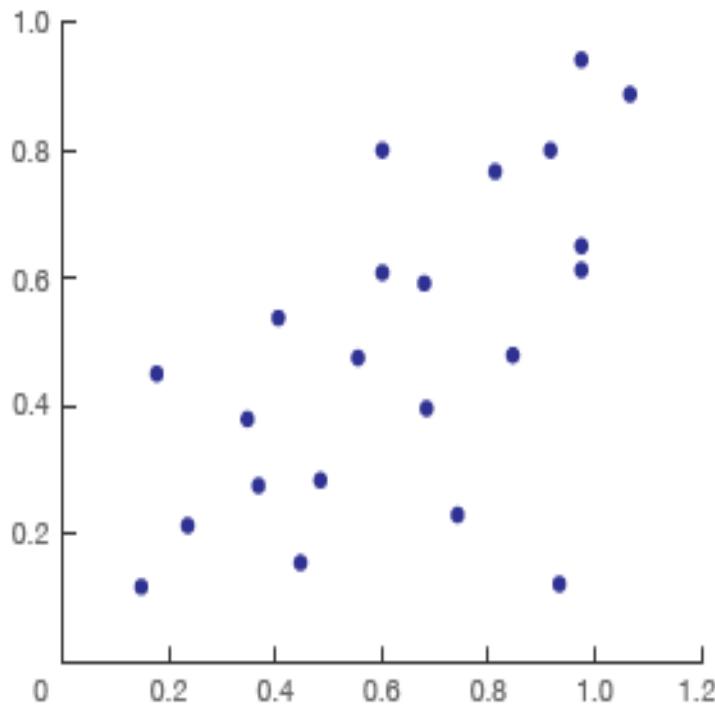
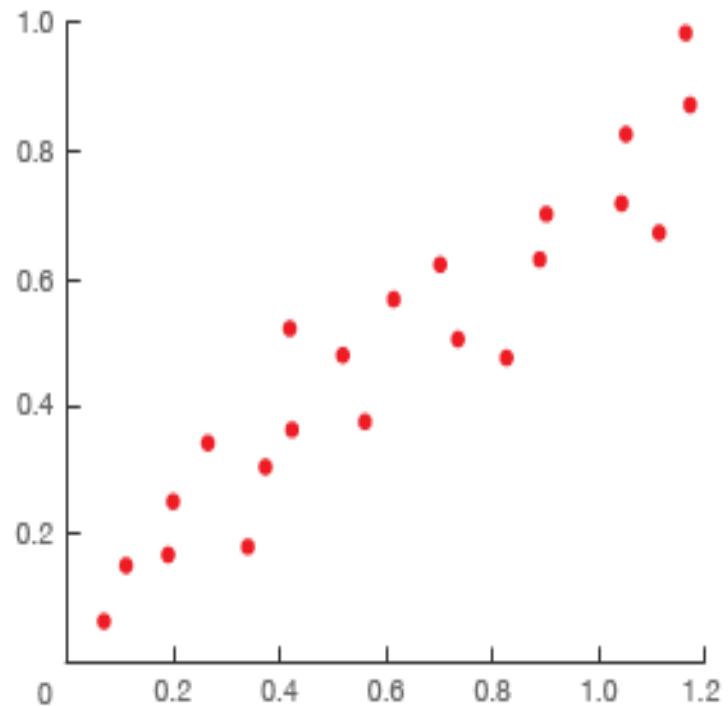
E. Nuage de points (scatter plot)

Caractéristiques clés :

- **Axe des abscisses (x)** :Représente une variable quantitative.
- **Axe des ordonnées (y)** :Représente l'autre variable quantitative.
- **Points** : Chaque point représente une paire de valeurs (x, y).

Exemples :

- Relation entre la taille et le poids d'une population.
- Corrélation entre la température et les ventes de glaces.
- Relation entre le nombre de publicités et les ventes d'un produit.
- ...



II-Identifier les types de graphiques

F. Boîte à moustaches (boxplot)

Définition : La boîte à moustaches, également appelée diagramme en boîte, est une représentation graphique qui résume la distribution d'un ensemble de données quantitatives. Elle affiche la médiane, les quartiles (Q1 et Q3) et les valeurs extrêmes (minimum et maximum), permettant ainsi de visualiser la dispersion et l'asymétrie des données, ainsi que d'identifier les valeurs aberrantes.

Utilisation :

- **Visualisation de la distribution**
- **Identification des valeurs aberrantes**
- **Comparaison de données :** Très pratique pour comparer la répartition de plusieurs séries de données.

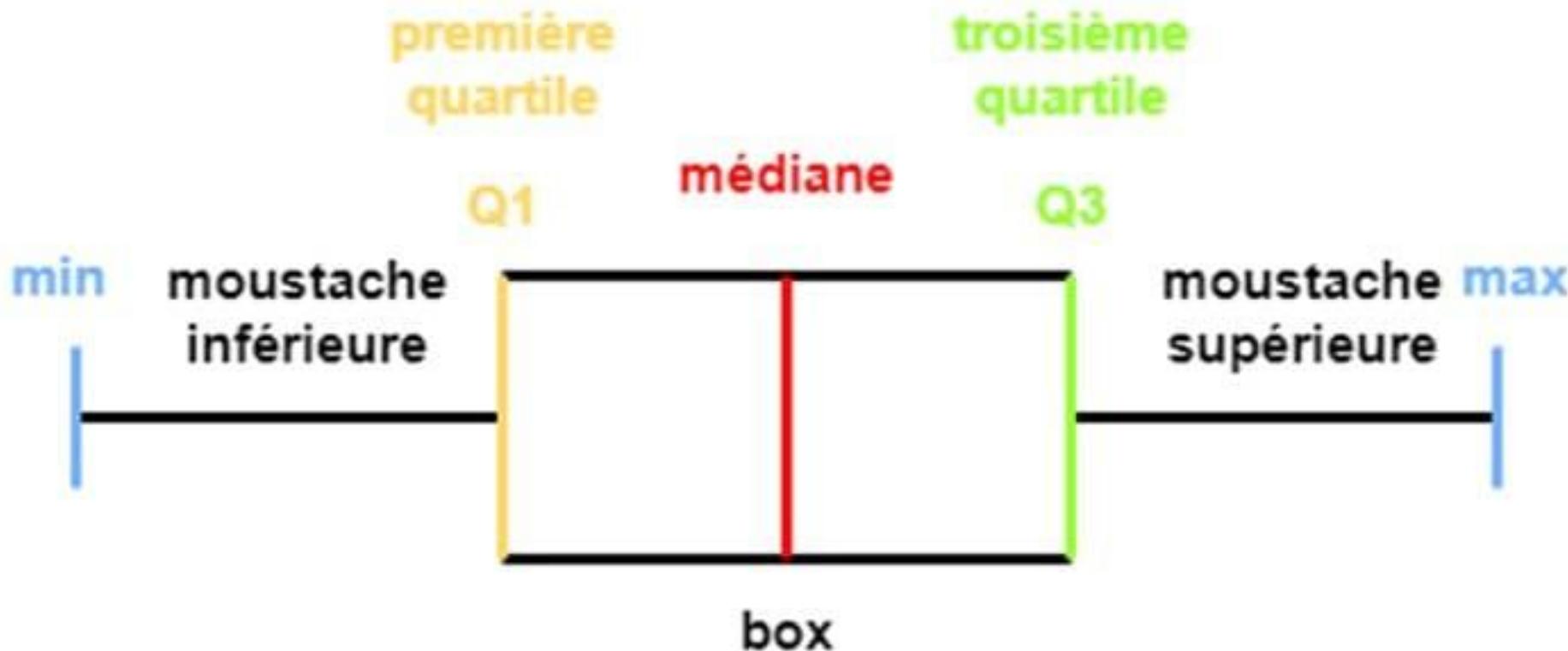
Exemple : Une boîte à moustaches peut être utilisée pour comparer les salaires des employés dans différentes entreprises.

II-Identifier les types de graphiques

F. Boîte à moustaches (boxplot)

Caractéristiques clés :

- **Médiane** : La ligne au milieu de la boîte représente la médiane, qui est la valeur centrale de l'ensemble de données.
- **Quartiles (Q1 et Q3)** :
 - La boîte est délimitée par le premier quartile (Q1) et le troisième quartile (Q3).
 - Q1 représente le 25e percentile et Q3 le 75e percentile.
 - La boîte représente donc 50% de la population étudiée.
- **Moustaches** : Les moustaches s'étendent à partir de la boîte jusqu'aux valeurs extrêmes.
- **Valeurs aberrantes** : Les valeurs aberrantes sont affichées comme des points individuels au-delà des moustaches.



III- Identifier les types de données dans vos graphiques

Avant de choisir une représentation graphique, il est crucial d'identifier le type de données à visualiser. Les données se divisent en deux grandes catégories :

➤ **Données qualitatives (catégoriques)**

Ces données expriment des catégories ou des caractéristiques qui ne sont pas mesurables numériquement.

- **Nominales** : Il s'agit de données sans ordre ou hiérarchie particulière.
 - *Exemple* : Les couleurs des voitures (rouge, bleu, noir), les types de fruits (pomme, banane, orange).
 - *Graphiques adaptés* : Diagramme en barres, camembert.
- **Ordinales** : Ces données présentent un ordre ou une hiérarchie.
 - *Exemple* : Niveaux d'éducation (primaire, secondaire, universitaire), échelle de satisfaction (faible, moyen, élevé).
 - *Graphiques adaptés* : Diagramme en barres ordonnées, histogramme.

III- Identifier les types de données dans vos graphiques

➤ Données quantitatives (numériques)

Ces données expriment des valeurs numériques mesurables.

- **Discrètes** : Elles prennent des valeurs entières et distinctes.
 - *Exemple* : Nombre d'élèves dans une classe, nombre de voitures dans un parking.
 - *Graphiques adaptés* : Diagramme en barres, histogramme.
- **Continues** : Elles peuvent prendre n'importe quelle valeur dans un intervalle donné.
 - *Exemple* : Taille des individus, température, vitesse d'un véhicule.
 - *Graphiques adaptés* : Histogramme, boîte à moustaches, courbe.

IV- Associer le type de représentation graphique aux types de données

Le choix d'une représentation graphique dépend du type de données à illustrer. Voici un tableau récapitulatif :

Type de données	Type de graphique adapté
Qualitatives nominales	Diagramme en barres, camembert
Qualitatives ordinaires	Diagramme en barres ordonnées, histogramme
Quantitatives discrètes	Diagramme en barres, histogramme
Quantitatives continues	Histogramme, boîte à moustaches, courbe
Relation entre deux variables	Nuage de points, courbe, boîte à moustaches

IV- Associer le type de représentation graphique aux types de données

Exemples d'application :

- Pour analyser la répartition des ventes de différents produits dans un magasin, un diagramme en barres ou un camembert est recommandé.
- Pour observer l'évolution du chiffre d'affaires d'une entreprise sur plusieurs années, un graphique en ligne est plus approprié.
- Pour comparer la taille moyenne des élèves dans différentes écoles, un boxplot permet d'analyser la dispersion des tailles.

En résumé, la sélection du graphique approprié garantit une communication claire et efficace des informations clés de vos données.

V- Conclusion

L'utilisation des graphiques permet de transformer des données brutes en informations exploitables. Pour une interprétation correcte, il est primordial de choisir un graphique en adéquation avec la nature des données. Ainsi, une bonne compréhension des types de données et de leurs représentations graphiques assure une communication efficace et pertinente des résultats d'analyse.