

# Développez votre culture data

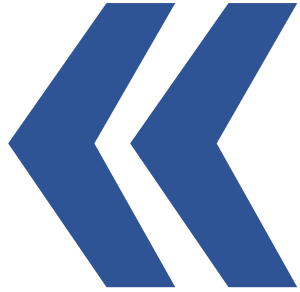
---

Kimba FONDIO & Aurel ATTERE

# Plan

- La data dans notre quotidien
- La collecte et la manipulation des données
- La visualisation des données
- La data storytelling

# La data dans notre quotidien



Au sens large, la data se présente, entre autres, sous forme de données factuelles, de calculs, de texte, de documents, d'images stockés sur un ordinateur.

# Définition et terminologie

---

- Quelle place occupe la données dans notre quotidien?
- Données – Informations – Connaissances – Actions/Décisions
- A quoi va vous servir votre culture data?

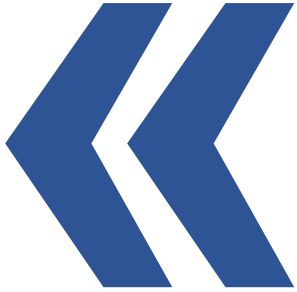


Une donnée est une **collection brute** de faits, de chiffres, de symboles ou d'observations qui n'ont pas encore été interprétées ou analysées. Elle est généralement **brute, non organisée et n'a pas de signification en soi**. Par exemple, un ensemble de nombres sans contexte spécifique, comme "12, 35, 48", est une donnée.



l'information est une donnée **traitée et organisée qui a acquis une signification et une utilité pour un utilisateur.**

Elle résulte du traitement et de l'interprétation des données pour en extraire du sens. Par exemple, si les nombres "12, 35, 48" sont accompagnés de libellés indiquant qu'il s'agit des températures minimales enregistrées dans trois villes différentes, alors cela devient de l'information.



La connaissance est un niveau supérieur d'information qui englobe la **compréhension, l'expérience et l'expertise d'une personne.**

Elle est le résultat de l'assimilation et de la compréhension approfondie de l'information par un individu, qui peut **l'utiliser pour résoudre des problèmes, prendre des décisions et générer de nouvelles idées.**

La connaissance implique également la capacité de contextualiser l'information dans un cadre plus large et de la relier à d'autres connaissances existantes.

Par exemple, si quelqu'un utilise les informations sur les températures minimales dans différentes villes pour comprendre les tendances météorologiques régionales et en tirer des conclusions sur le changement climatique, alors cela relève de la connaissance.

# Pipeline data

---

- De la data à la connaissance puis prise de décisions : comment cela se fait ?
  - **Source de données et données brutes**
  - **Opérations de transformations et acquisition de l'information**
  - **Connaissances et prises de décisions en fonction**

# Pipeline data : Travaux préparatoire

---

Définir les objectifs et avoir une vision globale de la tâche à accomplir :

- De quelles données aurais je besoin ?
- Où trouvé cette donnée?
- Sous quelle forme se présente t-elle?
- ...
- Posez les questions

# La collecte et la manipulation des données

# Collecte de données

---

- Données internes
  - Disponible dans les bases de données de l'entreprise
  - Fournit par le commanditaire
- Données externes
  - Disponible hors de l'écosystème décisionnel
  - Faire recours à une source externe
- Propriétés des données :
  - A qui appartient les droits d'utilisation des données;
  - Données libre,
  - Données publiques
  - Données privées

# Collecte de données

---

- Sensibilité des données :
  - peut on les utiliser sans qu'elles ne portent préjudice,
  - confidentialité des données
- Utilité des données:
  - Valeur informative des données,
  - Variabilités de la données
  - Anticiper le traitement des données pour qu'elles soient exploitable par l'ordinateur
  - Objectif avoir un format tabulaire
  - Format des données csv, Excel, parquet, avro
  - Types et structures de la données

# Manipulation des données

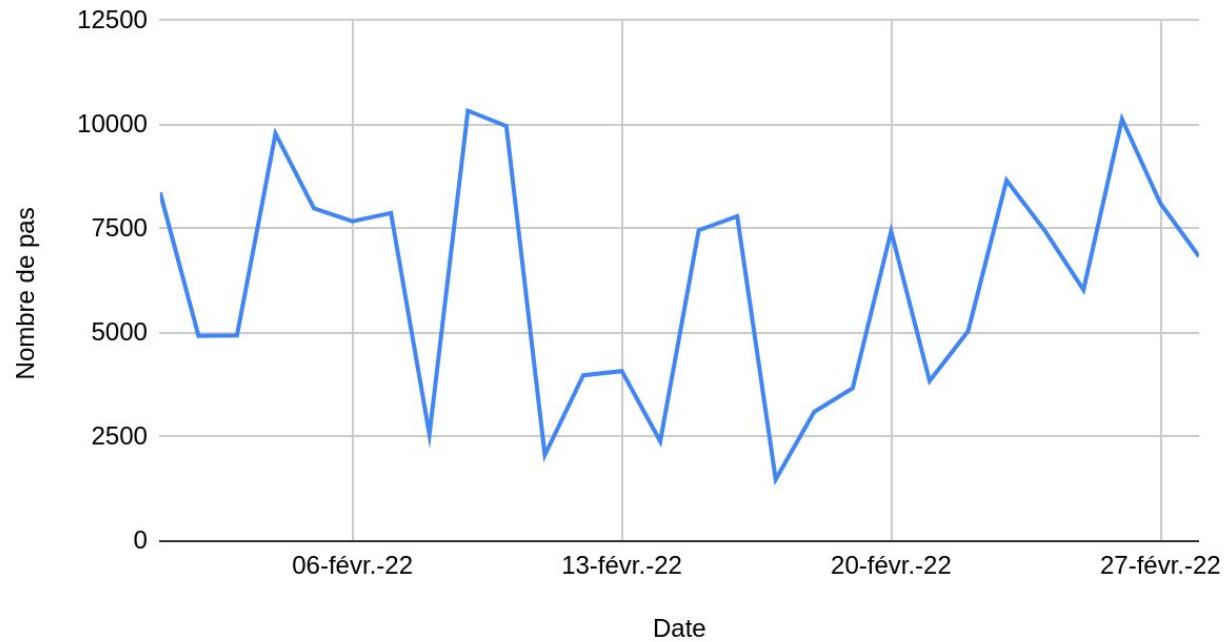
- Reliez
- Agrégez
- Filtrez/ Sélectionnez
- Nettoyez

# Analyse de données

- Types de la données
  - Quantitative (numérique)
  - Qualitative (descriptive)
  - Dates et heures ni quali ni quanti, apporte une information sur une dimension à part (granularité, mais attention elles apportent également leurs lots de problèmes, saisonnalité, évènements calendaires, moyenne mobile...)
- Exploration des données
  - Indicateurs statistique
  - Représentation graphiques

# Date et heure

Nombre de pas par rapport à Date

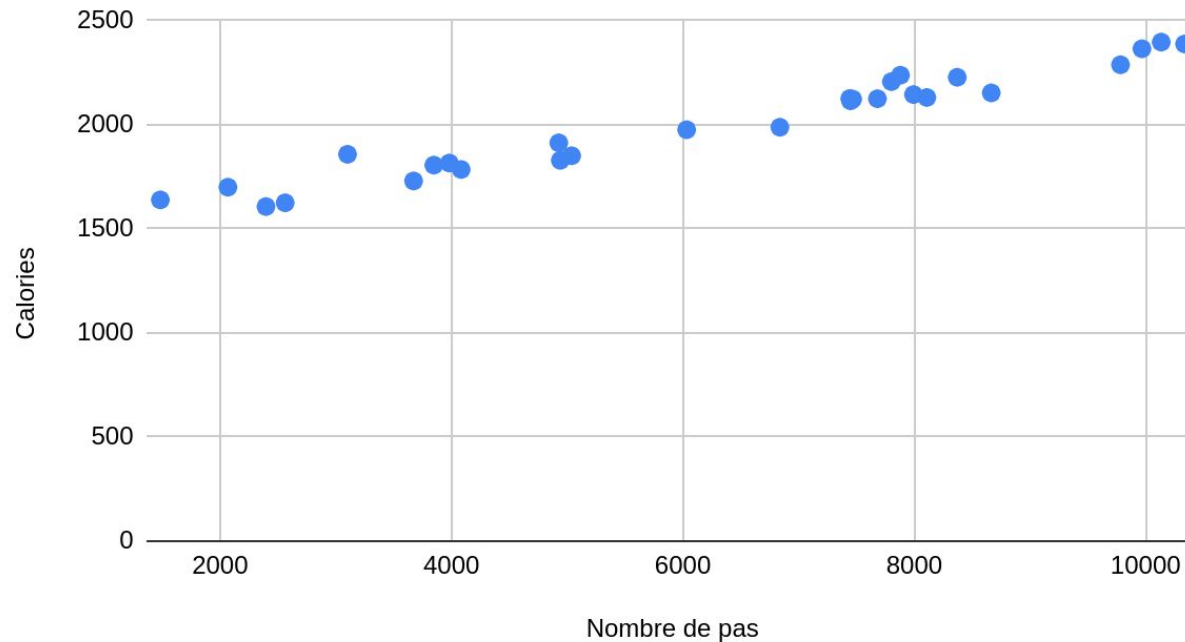


# La visualisation des données

---

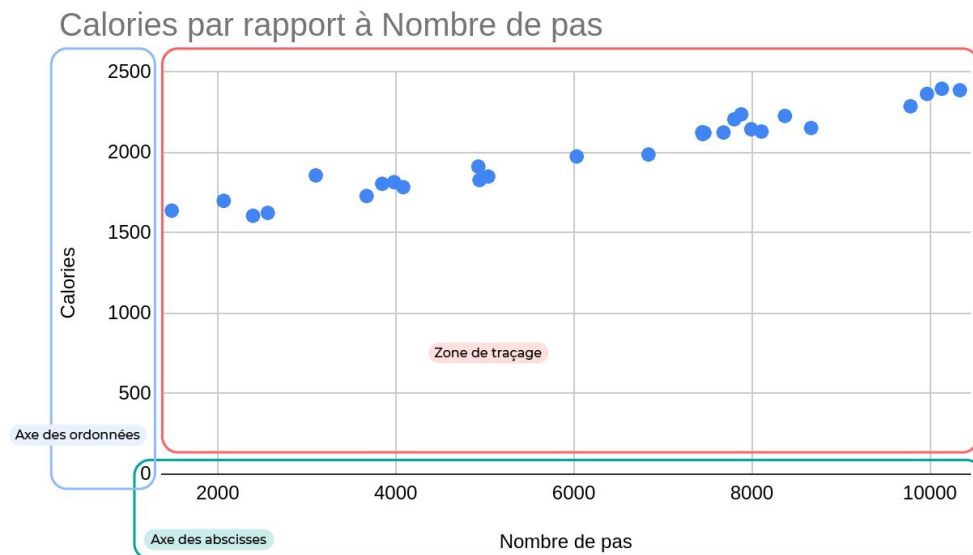
# Visualisation des données (1)

Calories par rapport à Nombre de pas

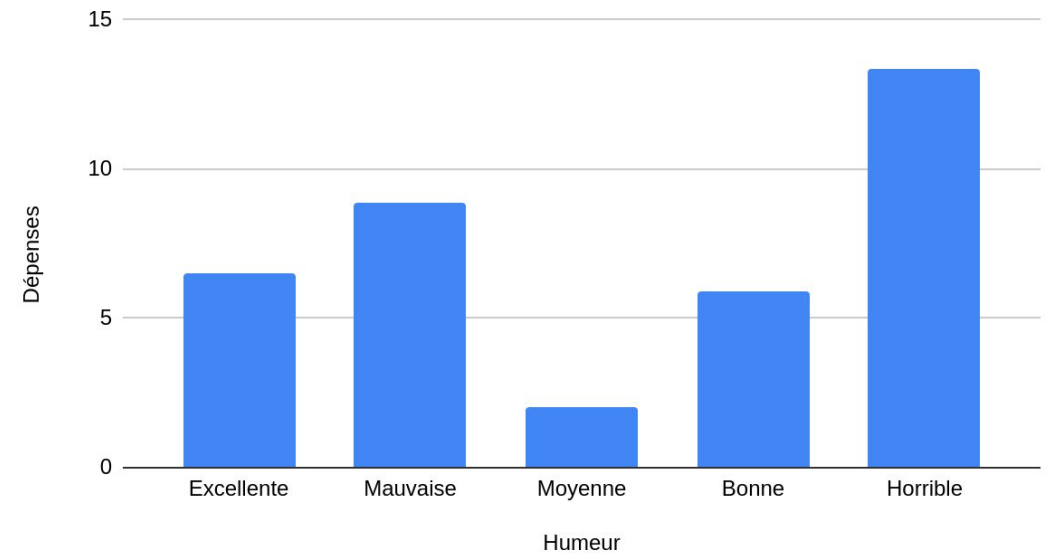


- Comparaison
- Tendence
- Corrélation
- Proportion
- Distribution
- Quantitative
- Qualitative
- Date et heure
- Représentation conjointe
- ...

# Visualisation des données (2)



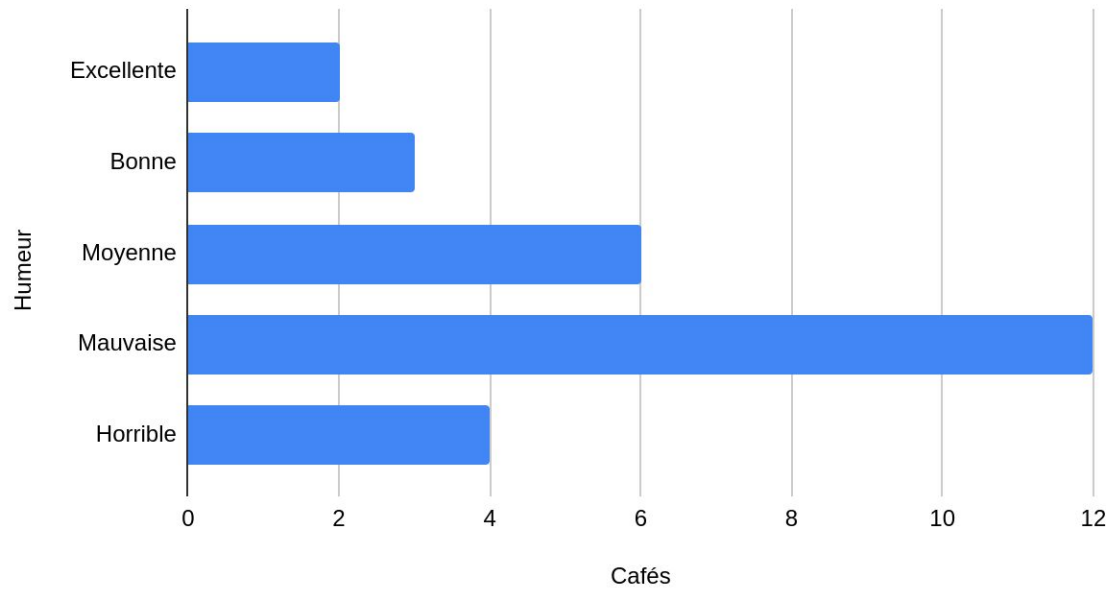
Dépenses en fonction de l'humeur



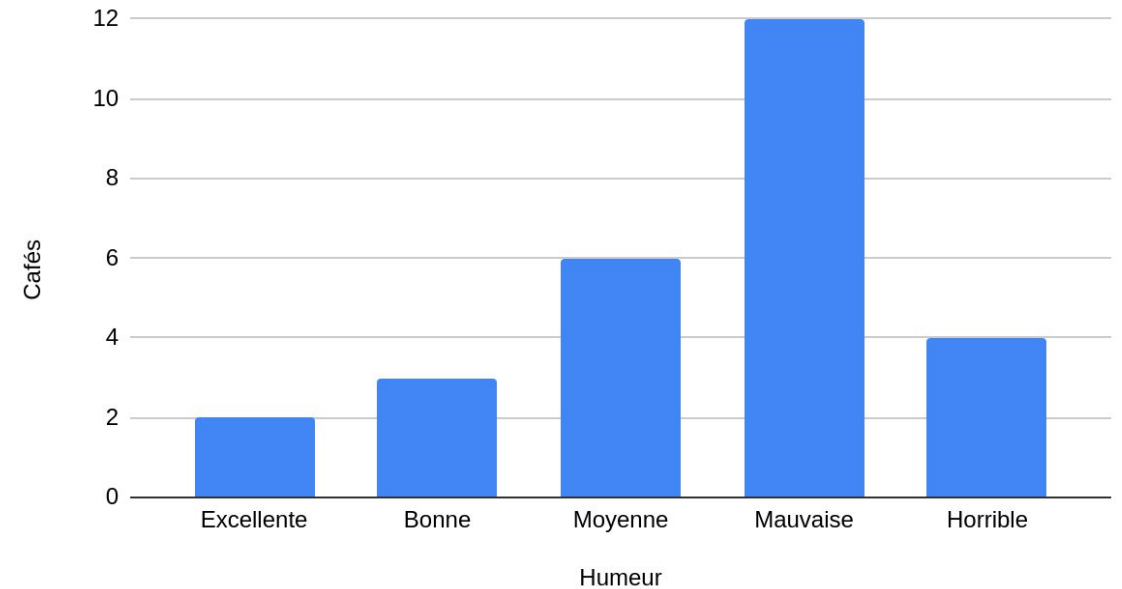
## Visualisation des données (3)

- Diagramme en barre

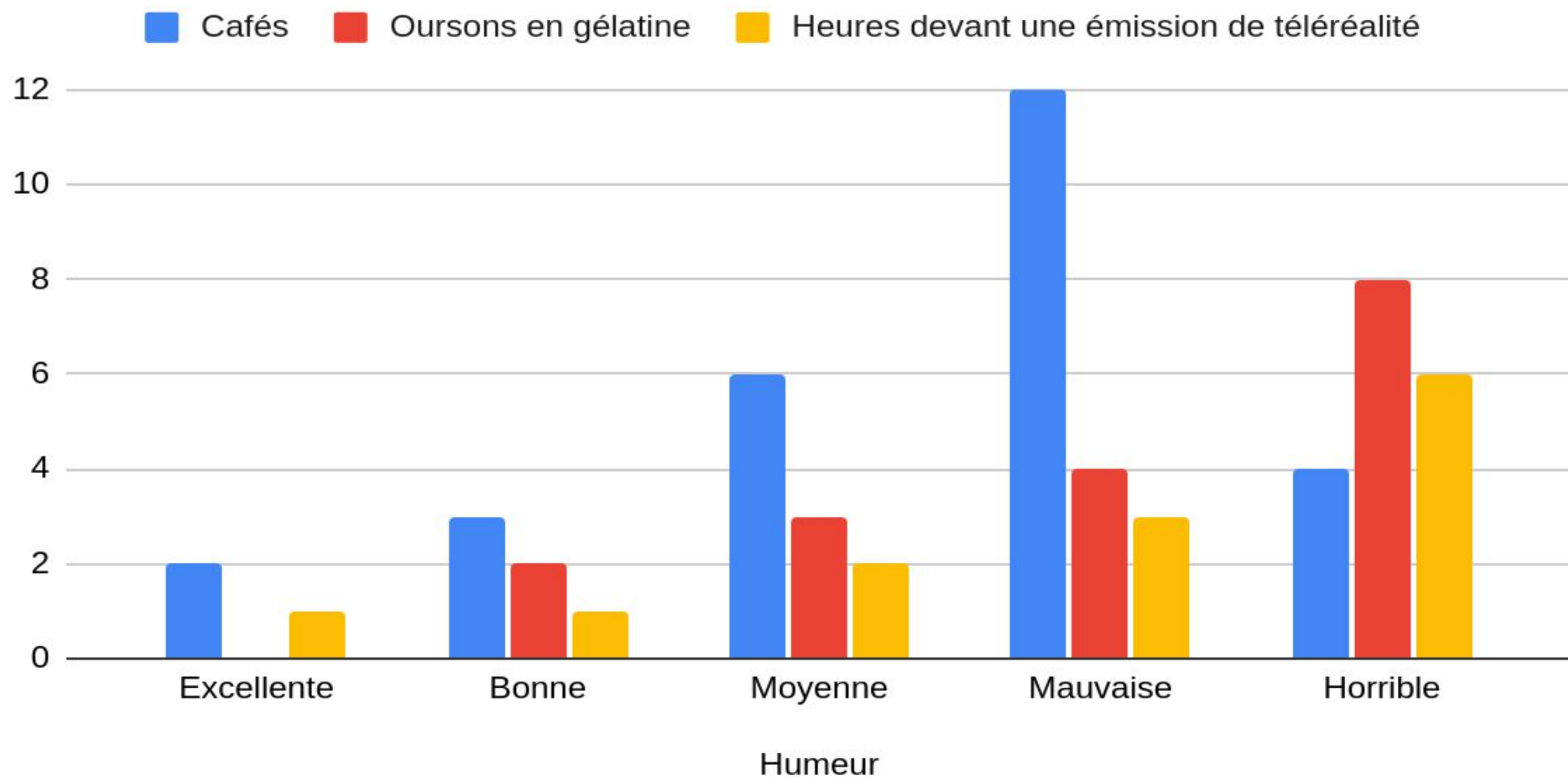
Cafés par rapport à Humeur



Cafés par rapport à l'humeur du jour



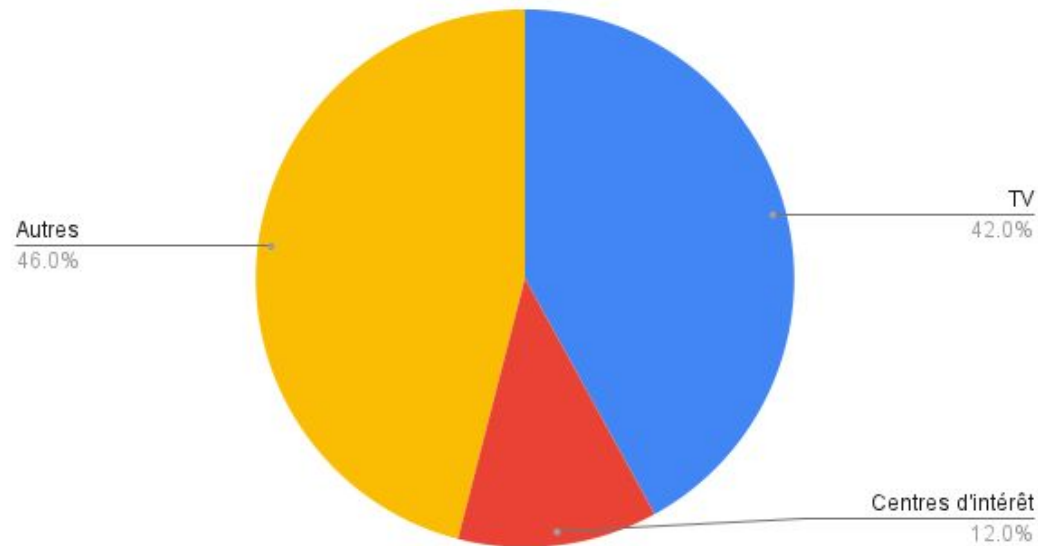
# Les péchés mignons de Sarah en fonction de son humeur



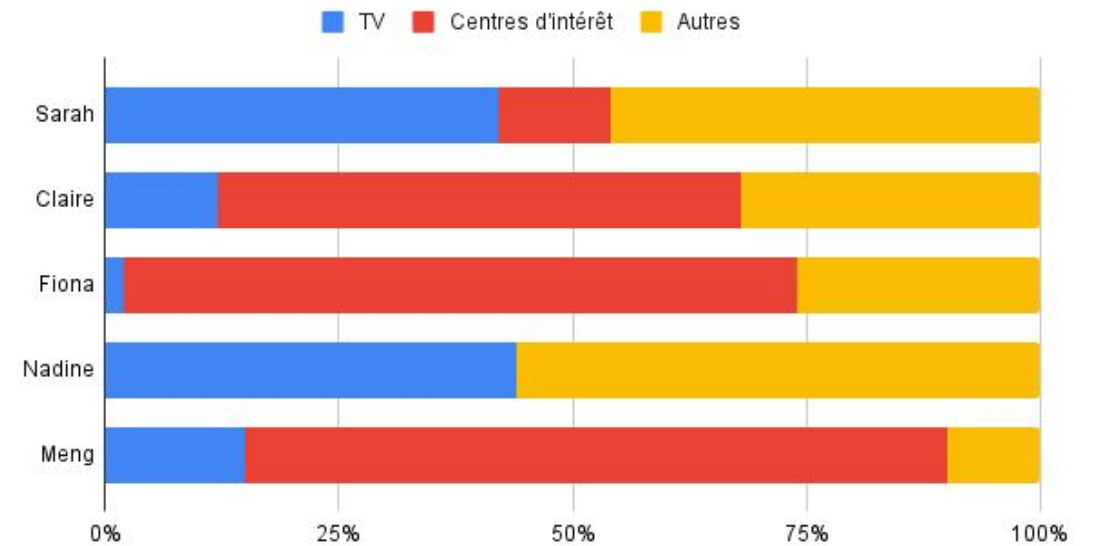
# Visualisation des données (5)

- Proportion

Activités réalisées à la maison par Sarah



Activités réalisées à la maison pour chacun des amis



# Quelques exigences d'une bonne représentation graphique

---

- Claire
- Pas extravagant
- Bien détaillé et légendé
- Pas d'information redondante
- ...

# La data storytelling



Il était une fois...

# Préparez une histoire basée sur vos données





Merci

