

Est-ce que les chiffres peuvent
mentir?

Est-ce qu'on peut leur faire dire
ce que l'on veut?

Option Math-Philo

Collège Rosa Parks

Séance du 25 mars 2024

Deux opinions communes

« Les faits sont têtus. »

Attribué à Lénine

Deux opinions communes

« Les faits sont têtus. »

Attribué à Lénine

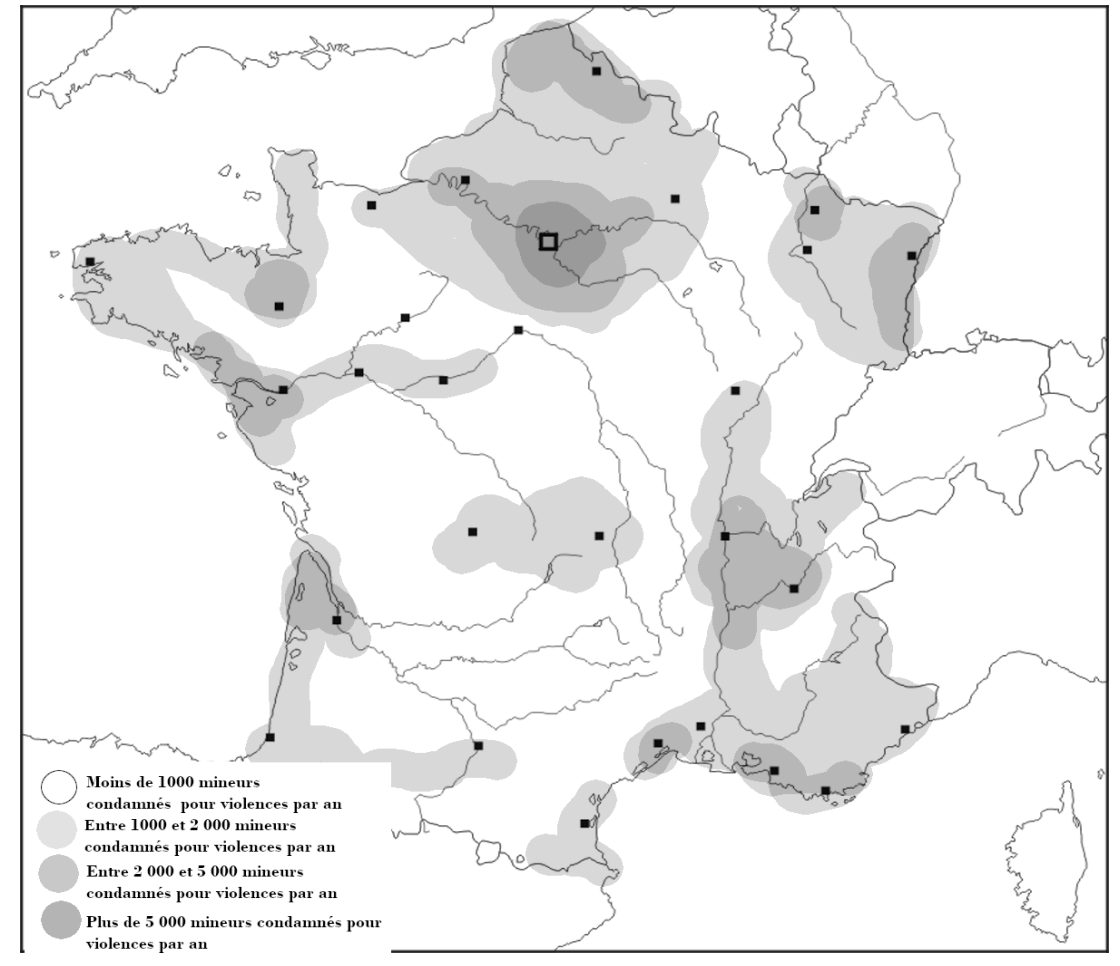
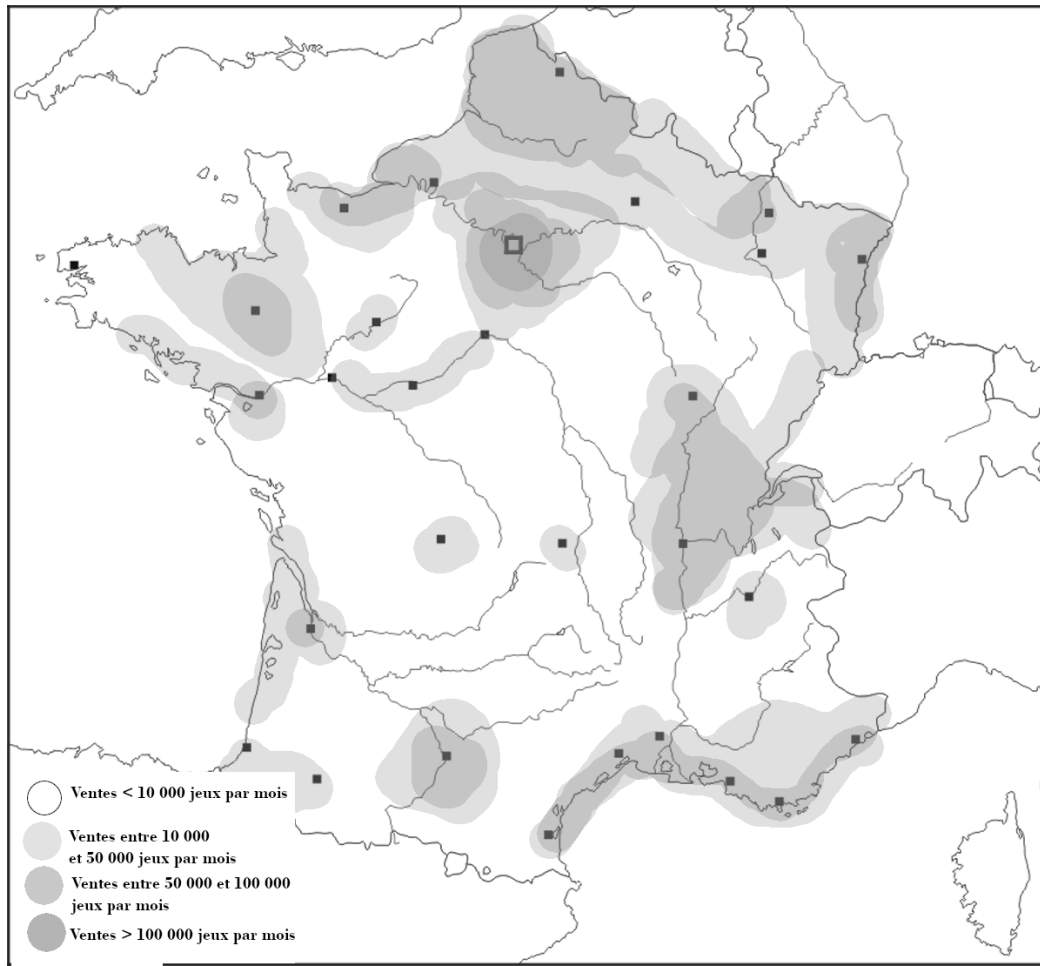
« Les faits sont têtus. Il est plus facile de s'arranger avec les statistiques. »

« Il y a trois sortes de mensonges : les mensonges, les gros mensonges et les statistiques. »

Attribué à Mark Twain

Activité 1 – Violence et jeux vidéo

A partir de ces deux cartes, le groupe 1 veut rédiger un article qui montre que les jeunes qui jouent aux jeux vidéo ont tendance à être plus violents. Le groupe 2 veut rédiger un article qui montre qu'il n'y a pas de lien entre violence des jeunes et pratique des jeux vidéo.



Activité 2 – les médicaments

Voici les données sur deux médicaments qui permettent de traiter deux maladies: le rhume et la bronchite Le groupe 1 veut rédiger un article qui prouve que le médicament A est le plus efficace, le groupe 2 veut rédiger un article qui prouve que le médicament B est le plus efficace.

Données sur le rhume		Données sur la bronchite	
Médicament A	Médicament B	Médicament A	Médicament B
81 guérisons sur 87 cas	234 guérisons sur 270 cas	192 guérisons sur 263 cas	55 guérisons sur 80 cas

Total sur les deux maladies	
Médicament A	Médicament B
273 guérisons sur 350 cas	289 guérisons sur 350 cas

Activité 3 – le procès

Un procès a lieu pour juger une personne qui a été identifiée par une caméra de surveillance comme étant une criminelle. Le groupe 1 veut écrire un article pour prouver qu'il est fortement probable que cette personne soit une criminelle. Le groupe 2 veut écrire un article qui prouve qu'il est fortement probable qu'elle soit innocente.

Caractéristiques de la caméra.

La caméra utilise un logiciel de reconnaissance faciale. Elle scanne les visages des passants et les compare à une base de données où se trouvent les visages de 100 personnes recherchées.

Elle a un taux d'erreur de 1%, c'est-à-dire que quand elle scanne un visage, il y a 1% de chance qu'elle identifie une personne innocente comme étant un criminel, et 1% de chance qu'elle n'identifie pas un criminel.

Position de la caméra

La caméra est dans une rue très passante. Elle a scanné toute la population de la ville depuis sa mise en service. Dans cette ville, il y a 100 000 personnes innocentes et 100 criminels.

Peut-on faire dire ce que l'on veut aux données chiffrées?

Il y a **toujours** plusieurs manières de **présenter** des données.

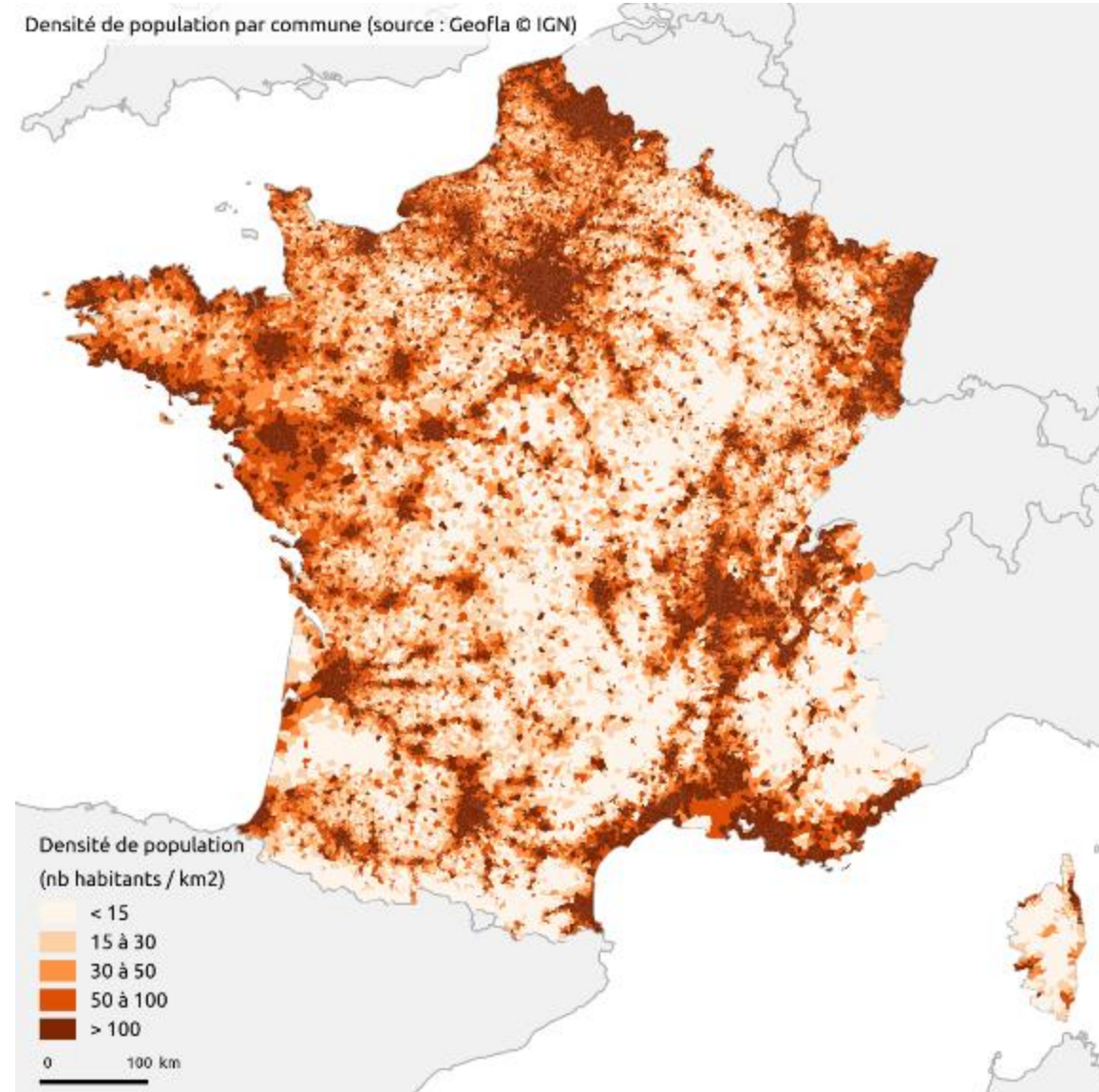
Mais est-ce qu'elles se valent toutes?

Est-ce que dans les activités que vous venez de faire il n'y avait pas une *meilleure* manière d'interpréter les données?

Activité 1 – Violence et jeux vidéo

Le groupe 2 avait raison: ces deux cartes ne prouvent rien parce qu'elles montrent une corrélation dont la cause commune était la densité de population.

Quelles auraient été les données pertinentes ?



Activité 2 – les médicaments

Le médicament A est plus efficace que le médicament B. Le groupe 1 avait raison. Mais comme le médicament B a été plus testé sur le rhume (qui est plus fréquent et plus simple à guérir) que sur la bronchite (qui est moins fréquente et plus difficile à guérir), les chiffres totaux donnent l'impression qu'il est plus efficace.

Données sur le rhume		Données sur la bronchite	
Médicament A	Médicament B	Médicament A	Médicament B
93% (81 guérisons sur 87 cas)	86% (234 guérisons sur 270 cas)	73% (192 guérisons sur 263 cas)	68% (55 guérisons sur 80 cas)

Total sur les deux maladies	
Médicament A	Médicament B
78% (273 guérisons sur 350 cas)	82% (289 guérisons sur 350 cas)

Activité 3 – le procès

Calculons le nombre de personnes qui ont été identifiées comme « criminelles » par la caméra.

Sur les 100 criminels : 99 personnes.

Sur les 100 000 de personnes innocentes : 1000 personnes.

Donc parmi les 1099 personnes identifiées comme criminelles, il n'y en a que 99 qui le sont réellement.

$$99/1099 = 0,09.$$

Il y a donc seulement 9 % de chance que la personne qui a été reconnue comme criminelle par la caméra le soit réellement. Le groupe 2 avait raison.

Conclusion: données et interprétation

Toute donnée doit être **interprétée** pour répondre à une question.

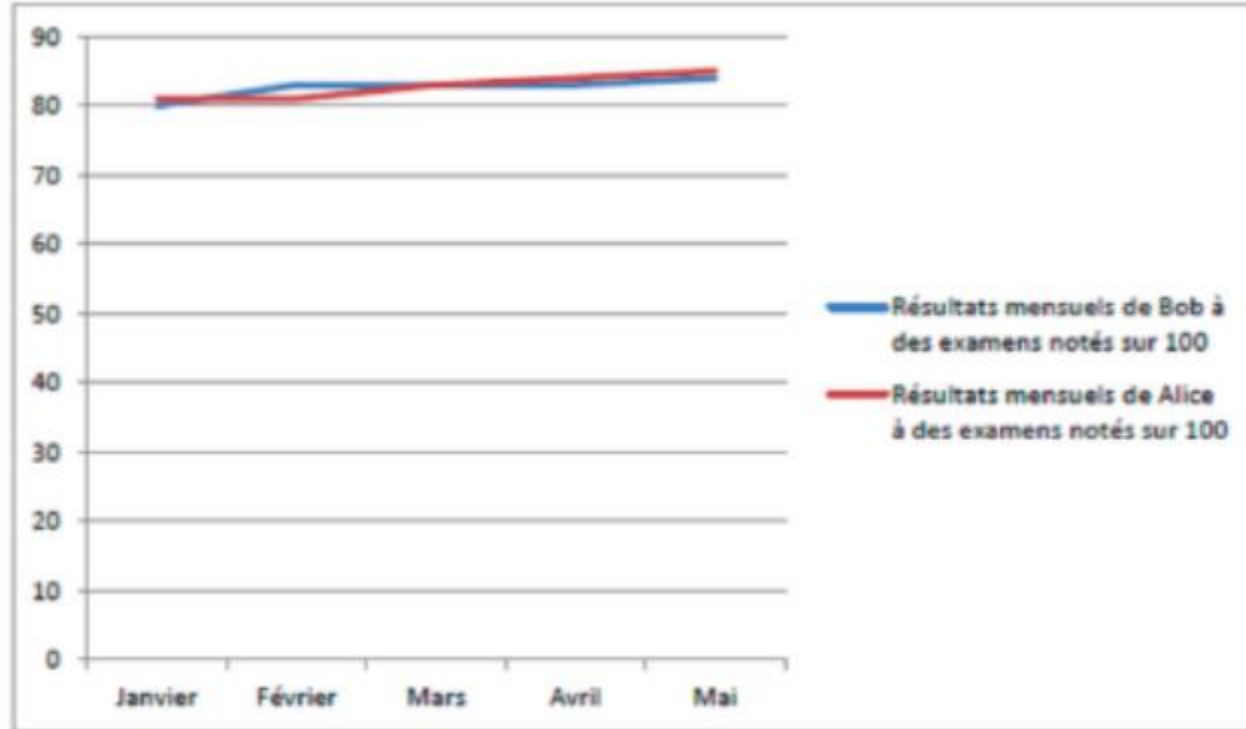
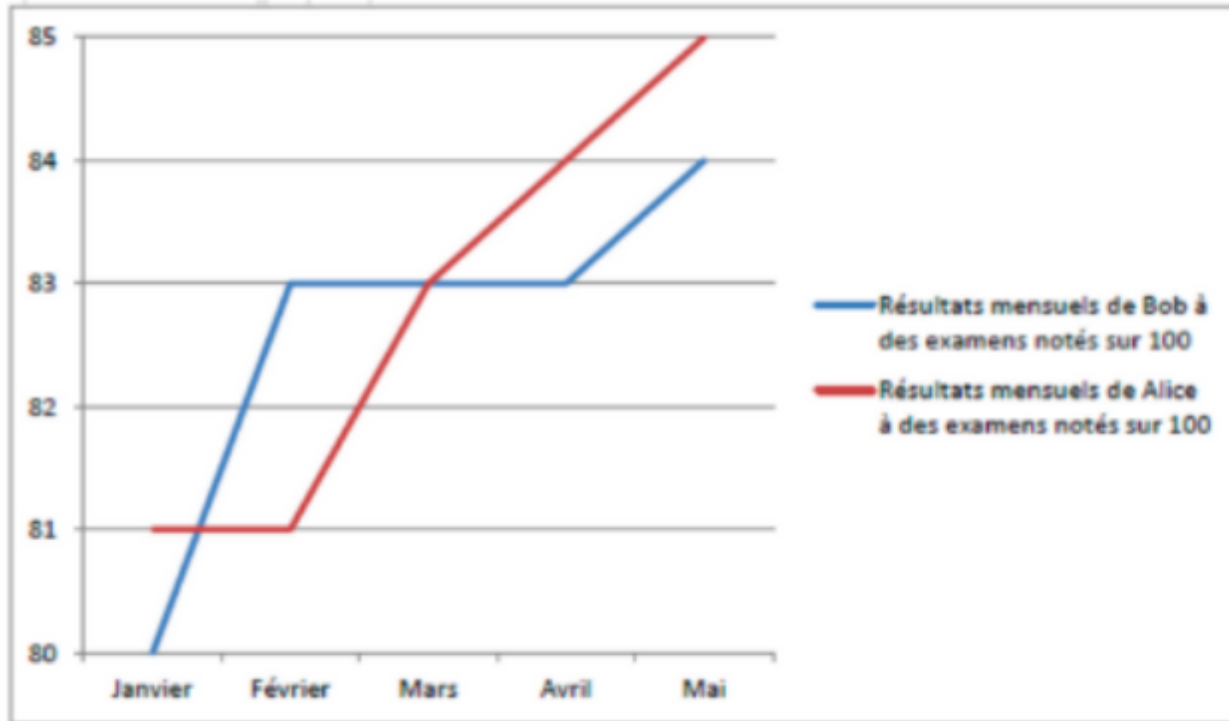
C'est le rôle des statistiques.

Interpréter = donner un *sens* à quelque chose en trouvant sa *cause*.

Il peut exister **plusieurs interprétations** des mêmes données mais elles ne se valent pas toutes.

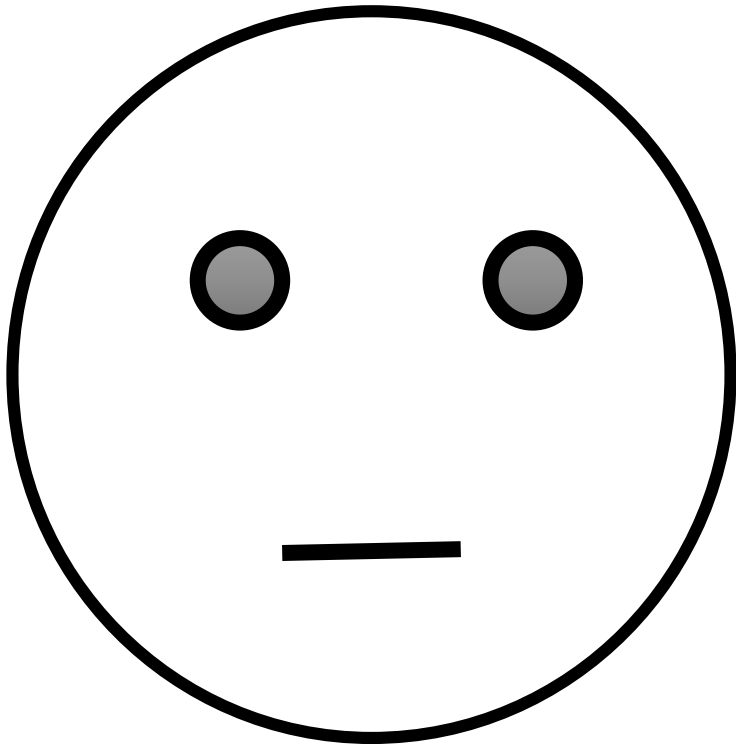
Donc **non**, on ne peut pas faire dire ce que l'on veut aux chiffres! Mais en changeant **la présentation** des données, **on peut influencer** leur interprétation.

Conclusion: données et interprétation



- Est-ce que ces deux graphiques représentent des données différentes ?
- Quelle est la différence dans la présentation ?
- Quelle interprétation a-t-on tendance à faire de chacun de ces graphiques ?

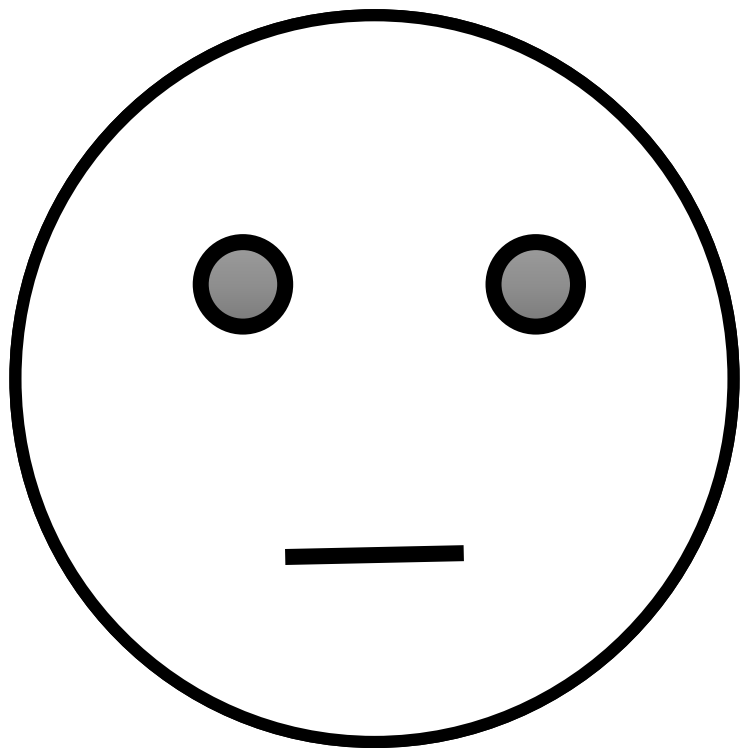
Conclusion: données et interprétation



Données :

Interprétation :

Conclusion: données et interprétation



Conclusion: données et interprétation

Attention : nous avons toutes et tous des **biais** d'interprétation, des tendances à interpréter les données de certaines manières qui peuvent nous mener à l'erreur ou à être manipulés.

Biais de confirmation : tendance à ne prendre en compte que les données qui confirment nos idées préconçues (préjugés).

Biais de causalité : tendance à voir des causes là où il n'y en a pas.