

# Cours 9 - Éléments empiriques (1/4) (12 mars 2025)

## Alignement pédagogique

### Mise en application

*Objectifs visés et activités associées*

**Objectif général : 4. Appliquer les méthodes de base en sciences de l'information pour analyser des données**

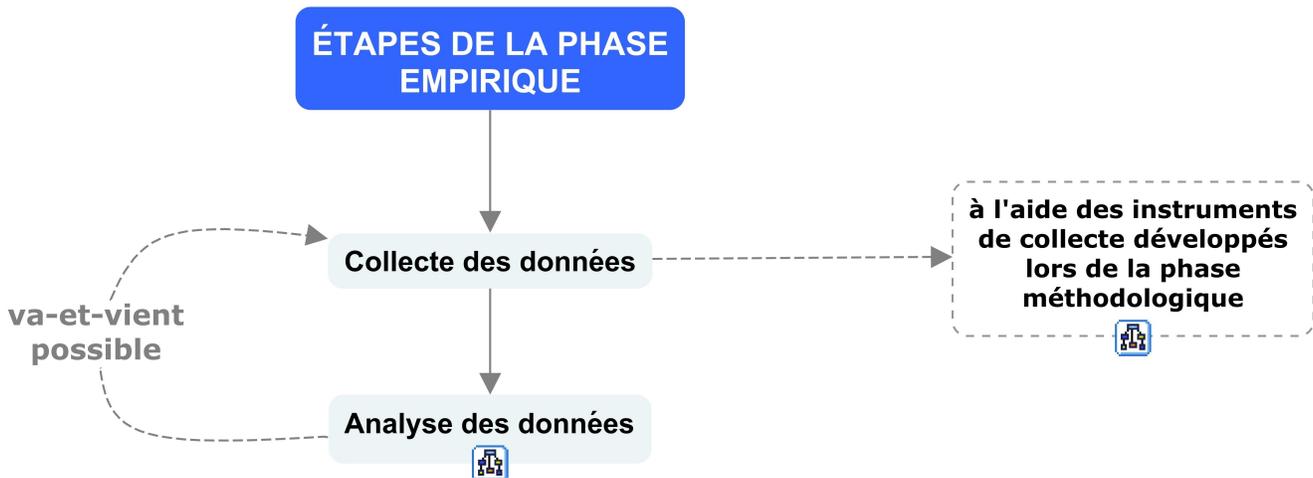
Objectif spécifique : 4b. Appliquer les méthodes statistiques de base pour analyser des données quantitatives

*Activité : TP3 et Devoir 3*

# Cours 9 - Éléments empiriques (1/4) (12 mars 2025)

## Cartes conceptuelles

Phase 3 : Éléments empiriques > Étapes de la phase empirique

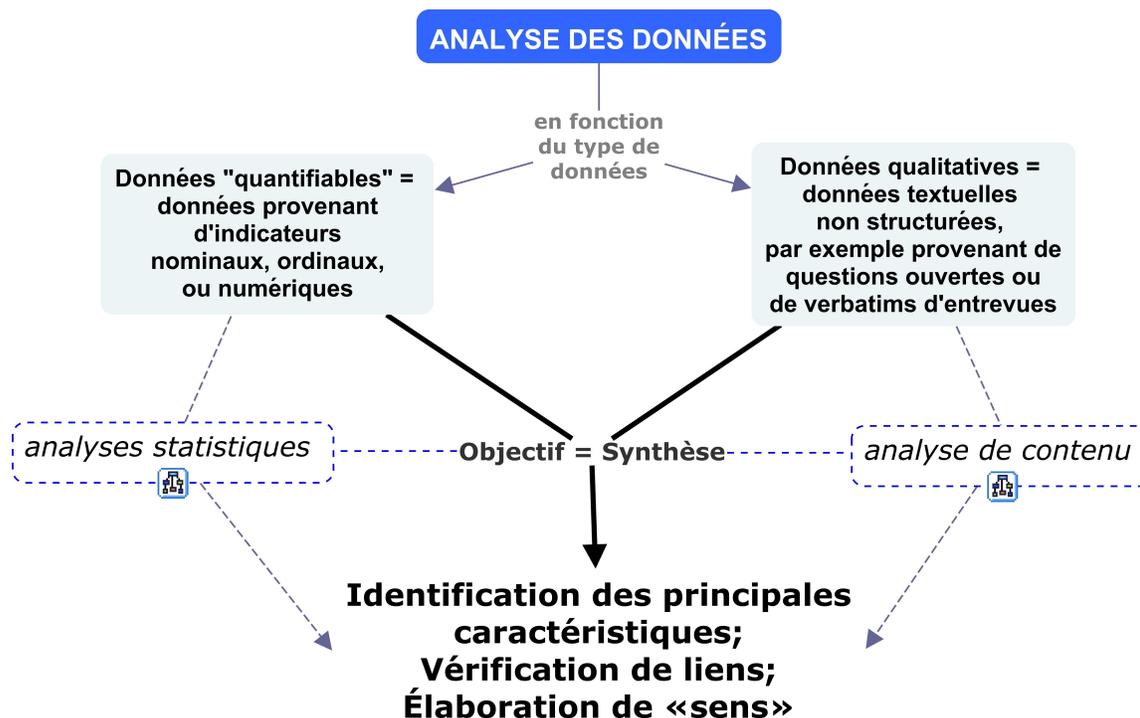


© Christine Dufour 2005-2018, 2024, EBSI, Université de Montréal  
(Modifications Virginie Wenglenski et Marielle St-Germain © 2023)

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c07\\_phaseempirique.html](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c07_phaseempirique.html)

### Synopsis

La phase empirique, qui suit la phase méthodologique, est le moment de mettre en pratique ce qui a été planifié dans les deux premières phases. On y retrouve la collecte des données, à partir des instruments préparés à la phase méthodologique, ainsi que l'analyse des données.

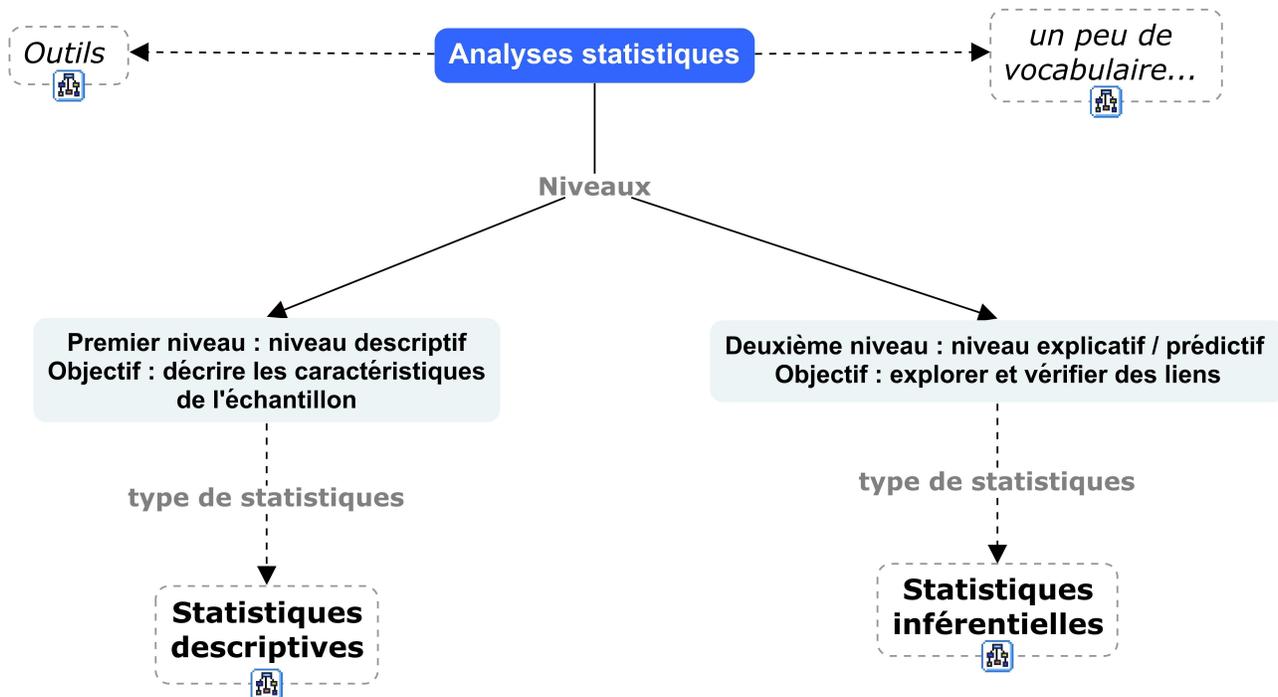


© Christine Dufour 2005-2018, 2024, EBSI, Université de Montréal  
(Modifications Virginie Wenglenski et Marielle St-Germain © 2023)

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c07\\_pe\\_analyse.html](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c07_pe_analyse.html)

### Synopsis

L'analyse des données, peu importe le type de données, est un exercice de synthèse. On veut réussir à en résumer les principales caractéristiques. Les techniques d'analyse vont varier selon que les données soient quantifiables (par exemple, provenant d'indicateurs nominaux, ordinaux ou numériques) ou qu'il s'agit de données textuelles non structurées. Dans le premier cas, il s'agira d'analyses statistiques et, dans le second, d'analyse de contenu.

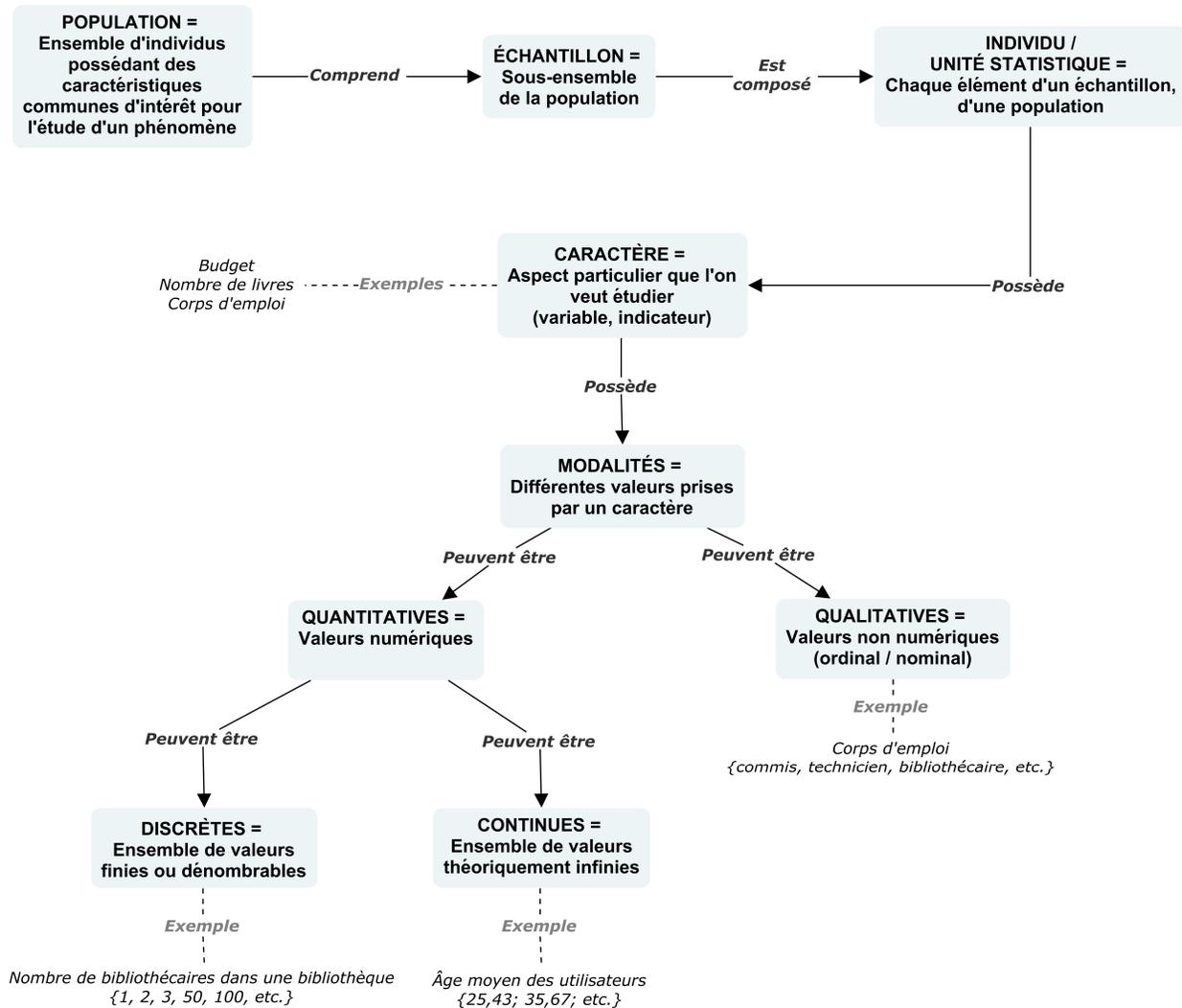


© Christine Dufour 2005-2018, 2024, EBSI, Université de Montréal  
(Modifications Virginie Wenglenski et Marielle St-Germain © 2023)

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c07\\_pe\\_analyse\\_statistique.html](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c07_pe_analyse_statistique.html)

### Synopsis

Selon le niveau d'analyse visé, il sera possible de faire des analyses statistiques dites descriptives (pour une recherche descriptive) ou inférentielles (pour des recherches au niveau explicatif ou prédictif).



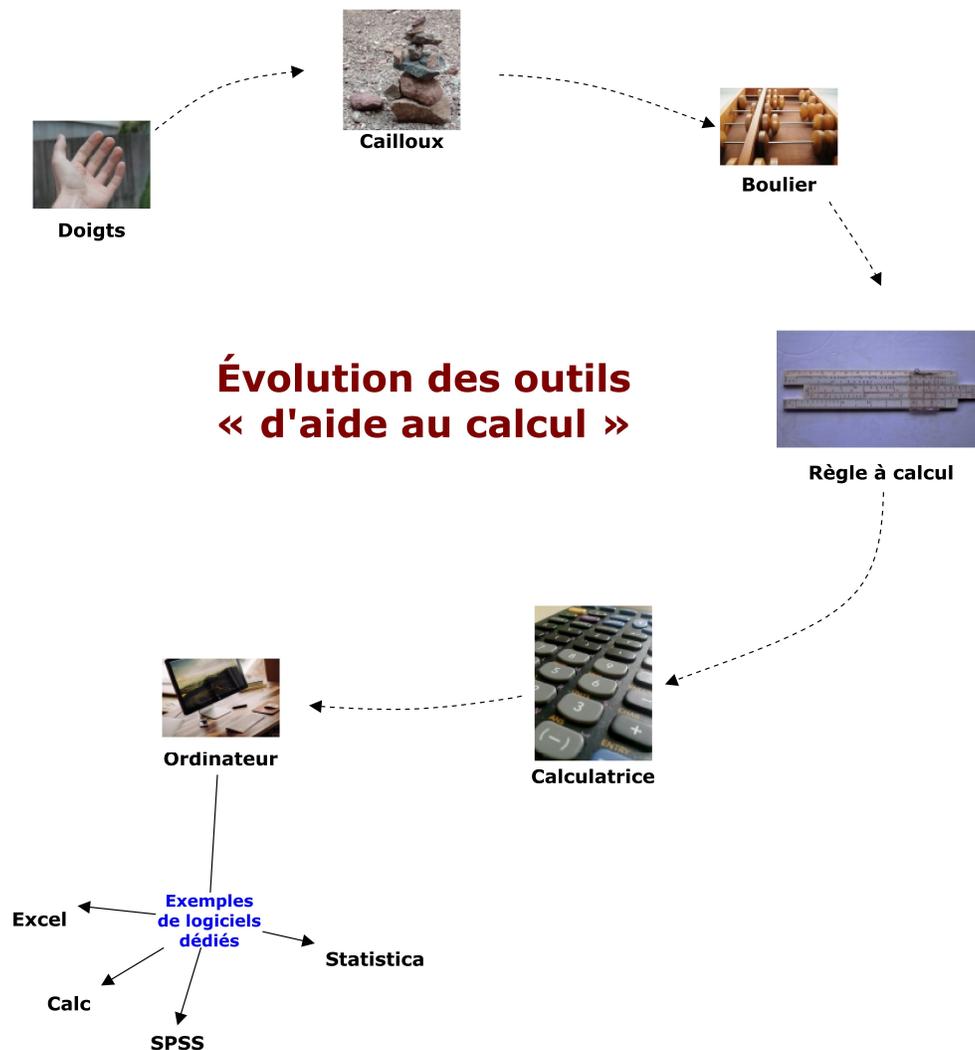
© Christine Dufour 2005-2018, 2024, EBSI, Université de Montréal (Modifications Virginie Wenglenski et Marielle St-Germain © 2023)

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c07\\_stats\\_concepts%20de%20base.html](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c07_stats_concepts%20de%20base.html)

### Synopsis

Un vocabulaire particulier est à maîtriser lorsque l'on parle d'analyse statistique. Les notions de population et d'échantillon sont les mêmes que nous avons déjà rencontrées à la phase méthodologique. S'y ajoute la notion d'individu ou d'unité statistique pour désigner chaque item composant l'échantillon.

Ces individus possèdent des caractéristiques que l'on veut étudier, ce que l'on peut appeler un caractère. Chaque caractère peut prendre différentes valeurs, donc différentes modalités. Ces dernières peuvent être quantitatives (numériques discrètes ou continues) ou non numériques (ordinales ou nominales).

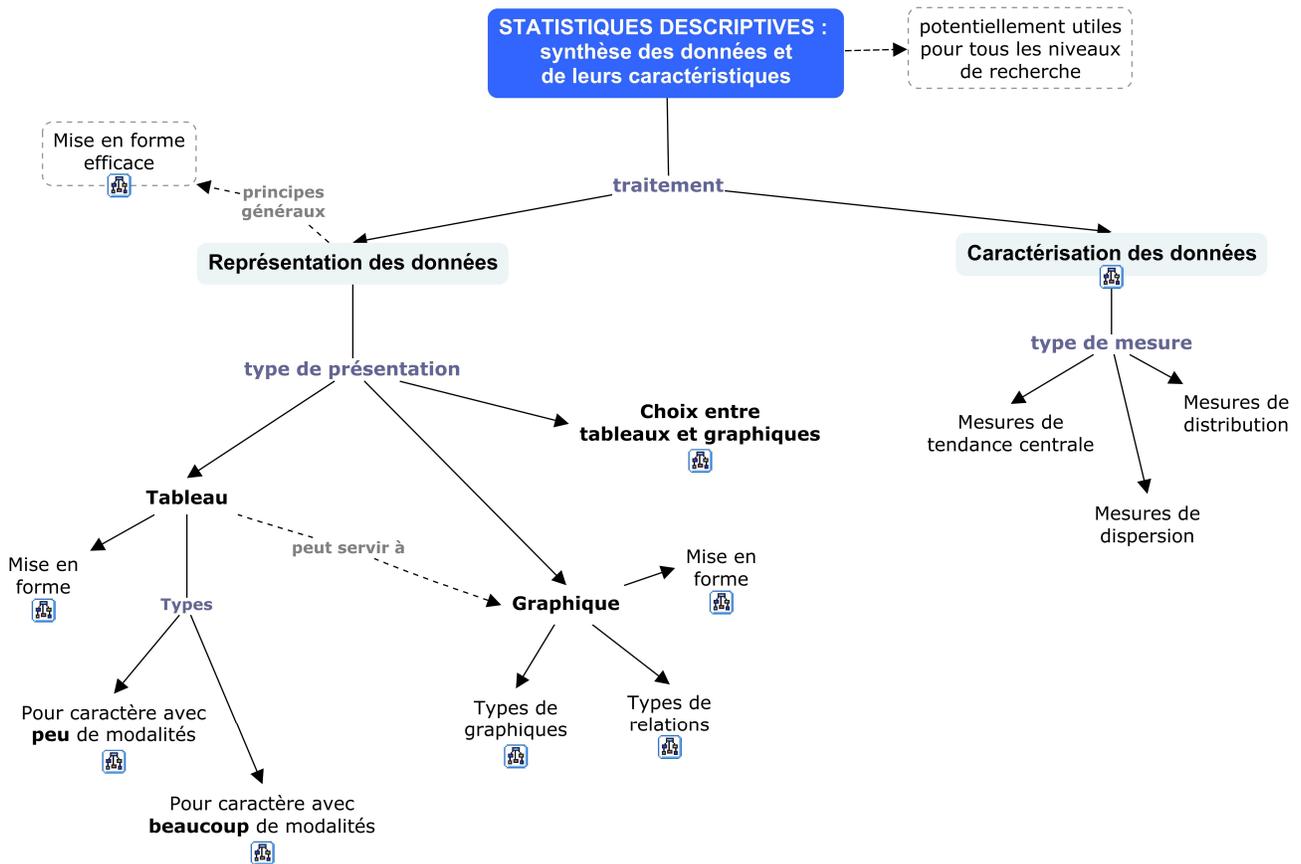


© Christine Dufour 2005-2018, 2024, EBSI, Université de Montréal  
(Modifications Virginie Wenglenski et Marielle St-Germain © 2023)

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c07\\_pe\\_analyse\\_statistique\\_outils.html](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c07_pe_analyse_statistique_outils.html)

### Synopsis

Les outils pour faire des statistiques ont évolué, bien heureusement pour nous! Différents logiciels sont maintenant accessibles pour faire des analyses statistiques, que ce soit des tableurs comme Excel (Microsoft) ou Calc (LibreOffice), pour faire des statistiques de base, ou des logiciels plus avancés comme SPSS ou Statistica.



© Christine Dufour 2005-2018, 2024, EBSI, Université de Montréal  
 (Modifications Virginie Wenglenski et Marielle St-Germain © 2023)

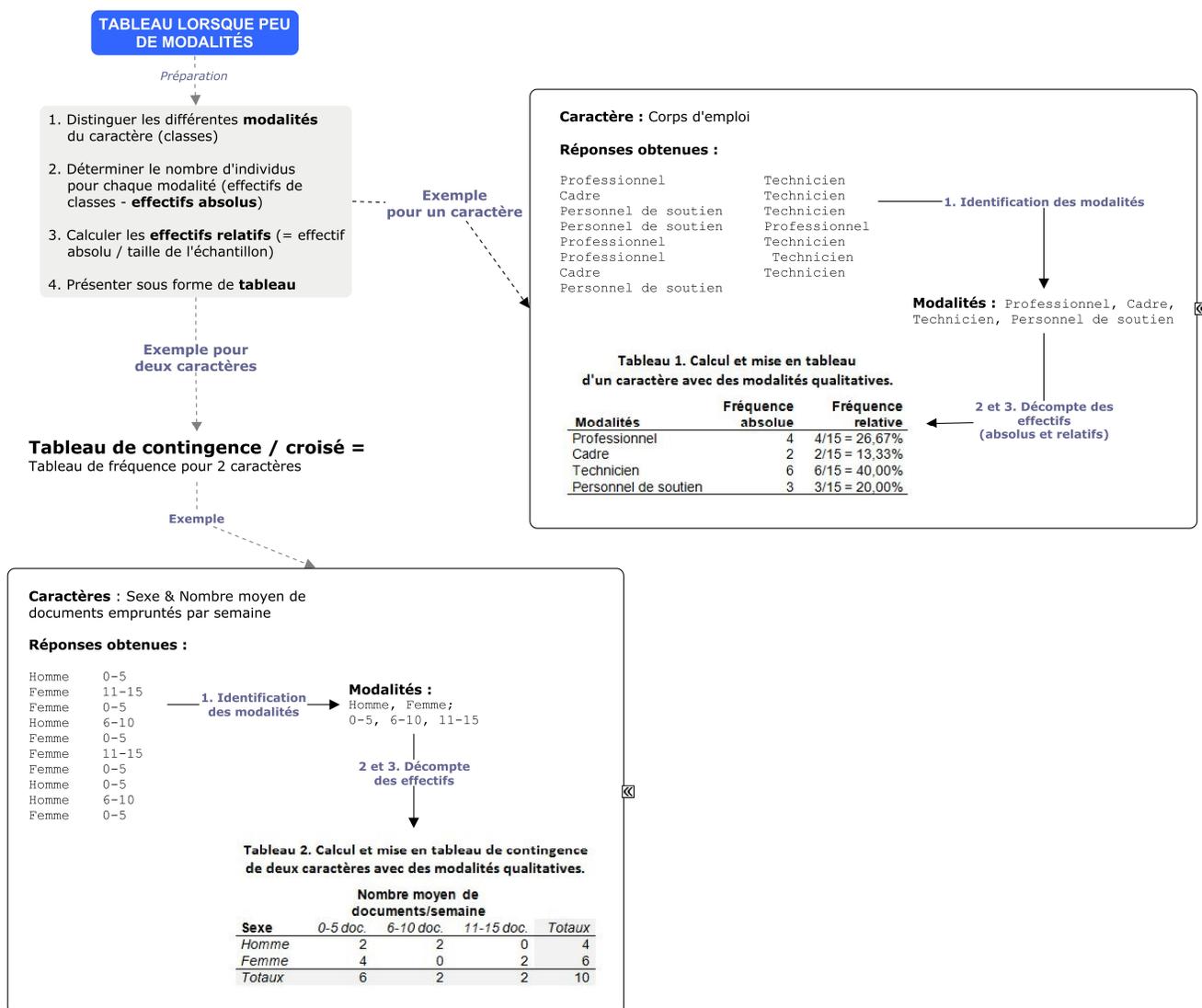
[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c09\\_pe\\_analyse\\_statistique\\_descriptives.html](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c09_pe_analyse_statistique_descriptives.html)

### Synopsis

Le premier niveau d'analyse statistique est celui des statistiques descriptives. Si elles sont utiles pour les recherches descriptives, elles peuvent aussi être exploitées pour des recherches prédictives ou explicatives. Comme leur nom l'indique, elles visent à synthétiser les données en les représentant soit en tableau ou en graphique, ou en en caractérisant la centralité ou la dispersion avec des mesures statistiques.

Le tableau permet de représenter textuellement les données de différentes manières selon que le caractère possède peu ou beaucoup de modalités. Le graphique exploite la dimension visuelle pour représenter différents types de relations. De manière générale, ainsi que spécifiquement pour les tableaux et les graphiques, il importe de respecter des principes de mise en forme pour en améliorer l'ergonomie.

Phase 3 : Éléments empiriques > Analyse des données > Analyses statistiques > Statistiques descriptives > Présentation en tableau > Caractère qualitatif OU quantitatif discret avec peu de modalités



© Christine Dufour 2005-2018, 2024, EBSI, Université de Montréal (Modifications Virginie Wenglenski et Marielle St-Germain © 2023)

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c09\\_pe\\_analyse\\_statistique\\_descriptives\\_tableau\\_qualitatif.html](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c09_pe_analyse_statistique_descriptives_tableau_qualitatif.html)

### Synopsis

Certains caractères ne prennent qu'un petit nombre de modalités distinctes. En ce cas, la synthèse en tableau consiste à identifier ces modalités pour ensuite calculer leurs fréquences dans les données. Les fréquences absolues sont les dénombrements effectués. Ces dernières peuvent utilement être accompagnées des fréquences relatives qui, en les convertissant en pourcentage, facilitent la comparaison.

Il arrive que l'on veuille regarder en même temps les fréquences de deux caractères. On procède à faire ce que l'on appelle un tableau de contingence ou tableau croisé. On y retrouvera en ligne un caractère et, en colonne, l'autre caractère. Les fréquences seront calculées pour chaque combinaison des deux caractères.

Phase 3 : Éléments empiriques > Analyse des données > Analyses statistiques > Statistiques descriptives > Présentation en tableau > Caractère quantitatif continu OU discret avec beaucoup de modalités

**TABLEAU LORSQUE BEAUCOUP DE MODALITÉS**

Exemple d'une méthode manuelle

1. Calculer **étendue** (valeur maximale – valeur minimale)
2. Choisir le **nombre de classes** (entre 5 et 15): peut être arbitraire, imposé ou basé sur une méthode mathématique (par ex. la règle de Sturges = valeur entière la plus proche de  $1 + 3,3 \log_{10}n$  où  $n$  est la taille de l'échantillon)
3. Calculer la **longueur des classes** (intervalle de classe) : = étendue / nombre de classes  
*Note* : pour la longueur des classes on retient le même nombre de chiffres après la virgule que les données mais sans arrondir et on ajoute 1 au dernier chiffre. Par ex., si la longueur des classes est 12,042 et que les données ont deux chiffres après la virgule, la longueur des classes est  $12,04 + 0,01 = 12,05$
4. Fixer la **limite inférieure** de la première classe (plus petite que la valeur minimale mais assez proche, une décimale de plus que les données pour éviter que des données arrivent sur une des limites)
5. Déduire les **autres limites** en additionnant la longueur des classes jusqu'à inclure la valeur maximale
6. Déterminer les **effectifs** de classe en comptant le nombre de données qui sont incluses dans chacune des classes

**Caractère:** Nombre de documents empruntés par mois

**Réponses obtenues:**

14	4	5	21	17	23	11
3	6	10	18	11	20	2
6	12	12	3	8	5	4
2	7	2	9	2	21	15
23	12	34	0	31	3	8
5	1	3	2	27	8	1
0	23	12	13	42	10	12
12	0	6	44	6	46	2
33	45	11	39	17	41	36

1. Calcul de l'étendue  
 $\text{maximum} - \text{minimum} = 46 - 0 = 46$
2. Définition du nombre de classes  
 7 (selon Sturges)
3. Définition de la longueur des classes  
 $\text{étendue} / \text{nombre de classes} = 46 / 7 = 6,5714$ , même nombre de décimales que les données sans arrondir + 1 =  $6 + 1 = 7$
4. Choix de la limite inférieure  
 -0,5 (plus petite valeur = 0)
5. Calcul des autres limites  
 Limite inférieure -0,5  
 $-0,5 + 7$  (longueur des classes) = **6,5**  
 $6,5 + 7 = \mathbf{13,5}$   
 $13,5 + 7 = \mathbf{20,5}$   
 $20,5 + 7 = \mathbf{27,5}$   
 $27,5 + 7 = \mathbf{34,5}$   
 $34,5 + 7 = \mathbf{41,5}$   
 $41,5 + 7 = \mathbf{48,5}$
6. Calcul des effectifs (absolus et relatifs)

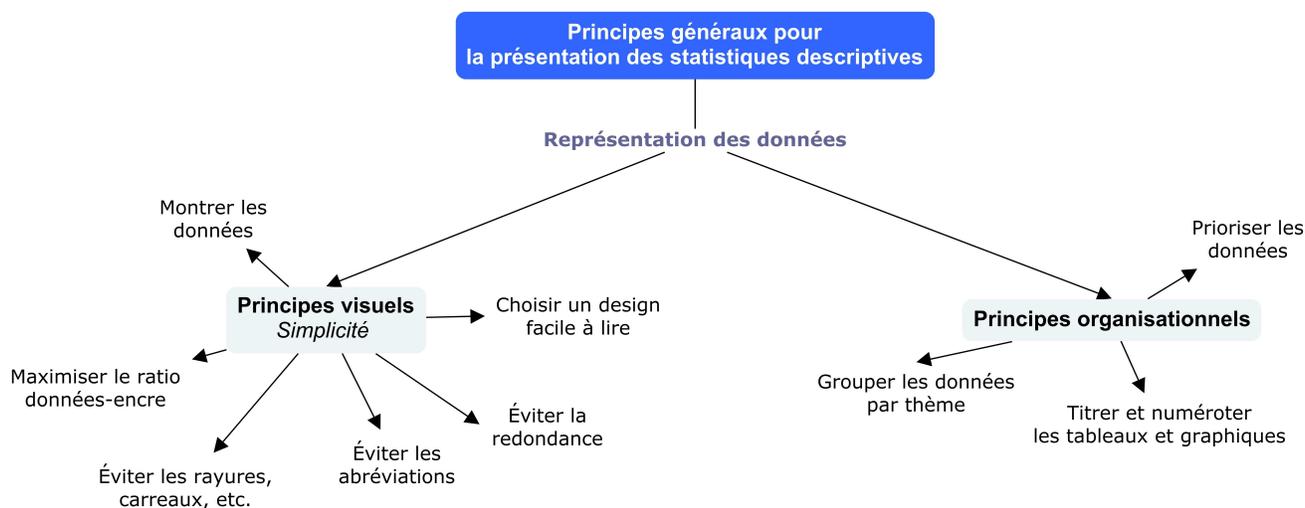
Classes	Fréquences absolues	Fréquences relatives
-0,5 - 6,5	24	38,10%
6,5 - 13,5	17	26,90%
13,5 - 20,5	6	9,52%
20,5 - 27,5	6	9,52%
27,5 - 34,5	3	4,76%
34,5 - 41,5	3	4,76%
41,5 - 48,5	4	6,35%

© Christine Dufour 2005-2018, 2024, EBSI, Université de Montréal  
 (Modifications Virginie Wenglenski et Marielle St-Germain © 2023)

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c09\\_pe\\_analyse\\_statistique\\_descriptives\\_tableau\\_quantitatif.html](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c09_pe_analyse_statistique_descriptives_tableau_quantitatif.html)

### Synopsis

Dans certains cas, un caractère peut prendre un grand nombre de modalités distinctes. La synthèse par le truchement de la liste des modalités et de leurs fréquences n'est alors pas utile, comme elle demeure trop détaillée. Il faut plutôt procéder à regrouper les modalités en classes, c'est-à-dire par rapport à des intervalles de valeurs. Différentes méthodes peuvent être utilisées pour le faire. Celle qui est proposée à titre d'exemple vise notamment à éviter que des modalités tombent à cheval sur la frontière entre deux classes.

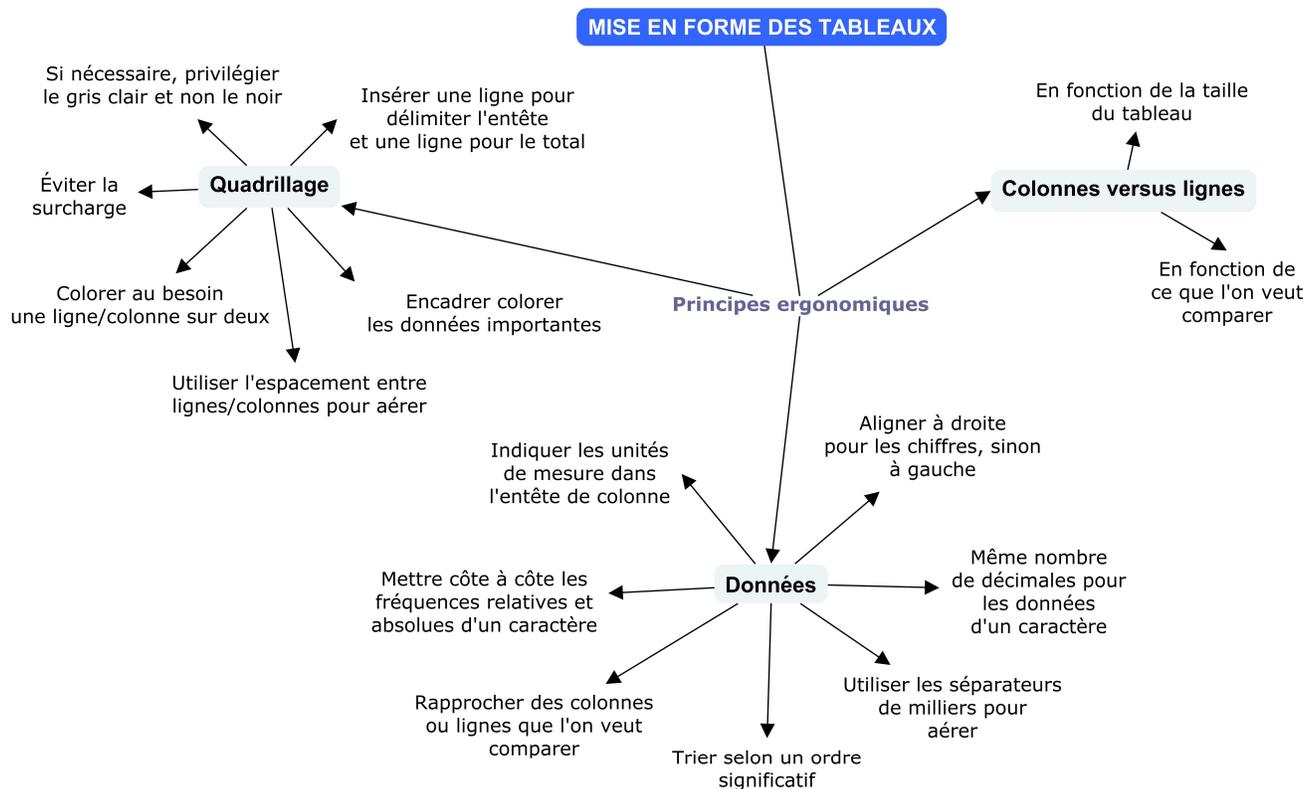


© Christine Dufour 2024, EBSI, Université de Montréal

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c09\\_pe\\_analyse\\_statistique\\_descriptives\\_principes\\_presentation.html](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c09_pe_analyse_statistique_descriptives_principes_presentation.html)

## Synopsis

Certains principes généraux sont à retenir pour améliorer la représentation des données. La simplicité est de mise! L'accent doit être sur les données comme telles et non sur des éléments décoratifs. Il faut aussi penser à faciliter le travail cognitif des lecteurs et lectrices par un choix de design facile à lire, en évitant les abréviations, et en regroupant les données par thème par exemple.

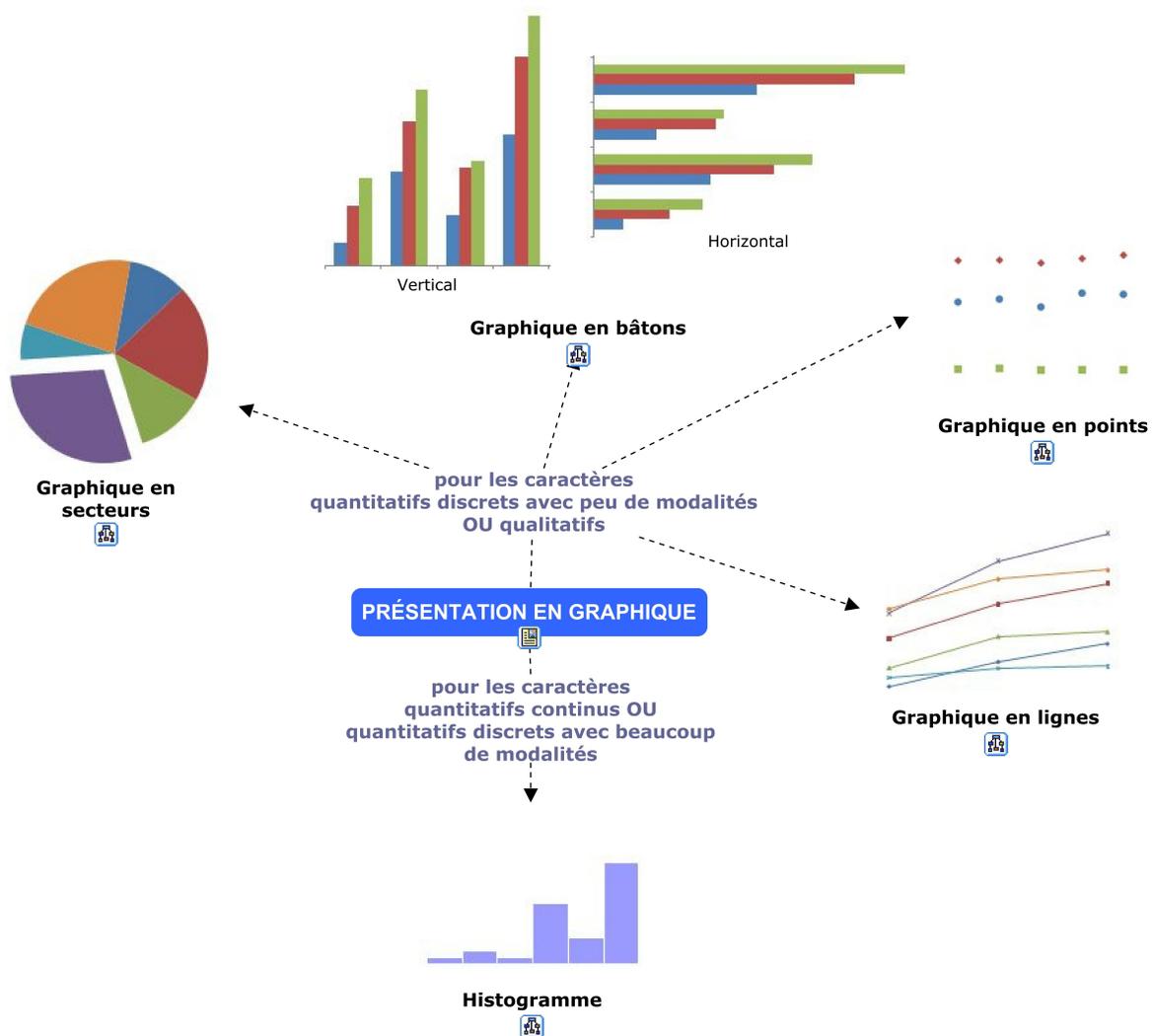


© Christine Dufour 2024, EBSI, Université de Montréal

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c09\\_pe\\_analyse\\_statistique\\_descriptives\\_principes\\_presentation\\_tableaux.ht](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c09_pe_analyse_statistique_descriptives_principes_presentation_tableaux.ht)

### Synopsis

Plus particulièrement, la mise en forme efficace des tableaux s'attarde à trois éléments principaux. D'une part, il faut minimiser le quadrillage des tableaux pour ne pas nuire aux données. D'autre part, les données doivent être présentées de manière cohérente et uniforme, par exemple pour l'alignement et les décimales ainsi que pour l'ordre de présentation. Finalement, il faut faire un choix pertinent quant à ce que l'on met en colonne et en ligne.

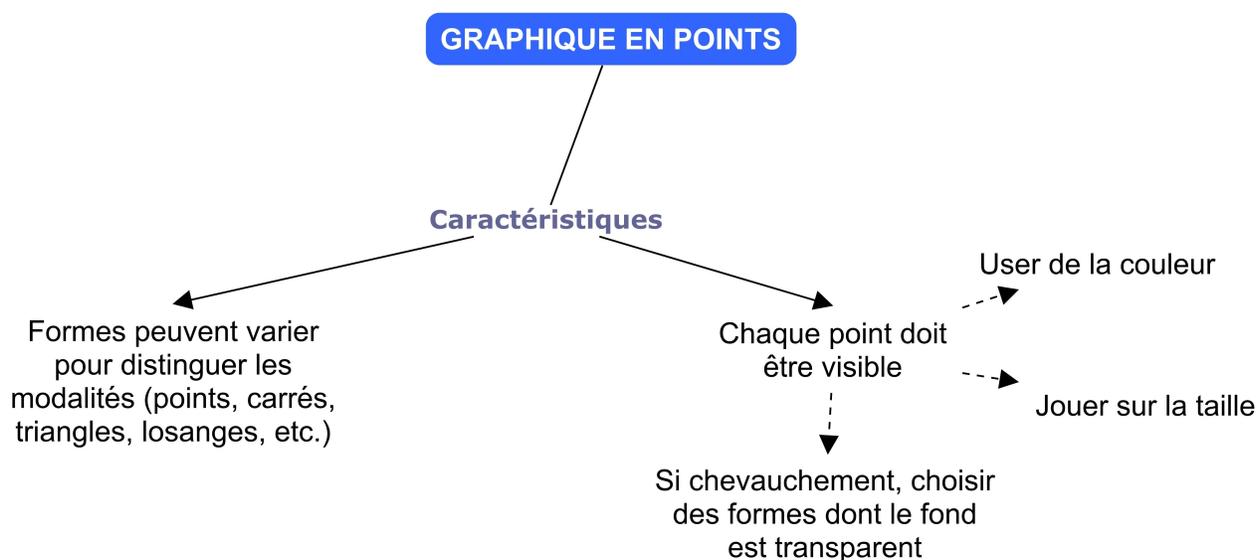
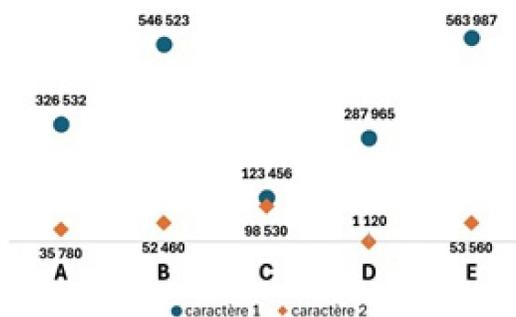


© Christine Dufour 2005-2018, 2024, EBSI, Université de Montréal  
(Modifications Virginie Wenglenski et Marielle St-Germain © 2023)

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c09\\_pe\\_analyse\\_statistique\\_descriptives\\_graphique.html](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c09_pe_analyse_statistique_descriptives_graphique.html)

## Synopsis

Les graphiques sont un autre outil pour synthétiser les données dans une visée descriptive. Plusieurs types de graphique existent et peuvent être utilisés en fonction du type de caractère.

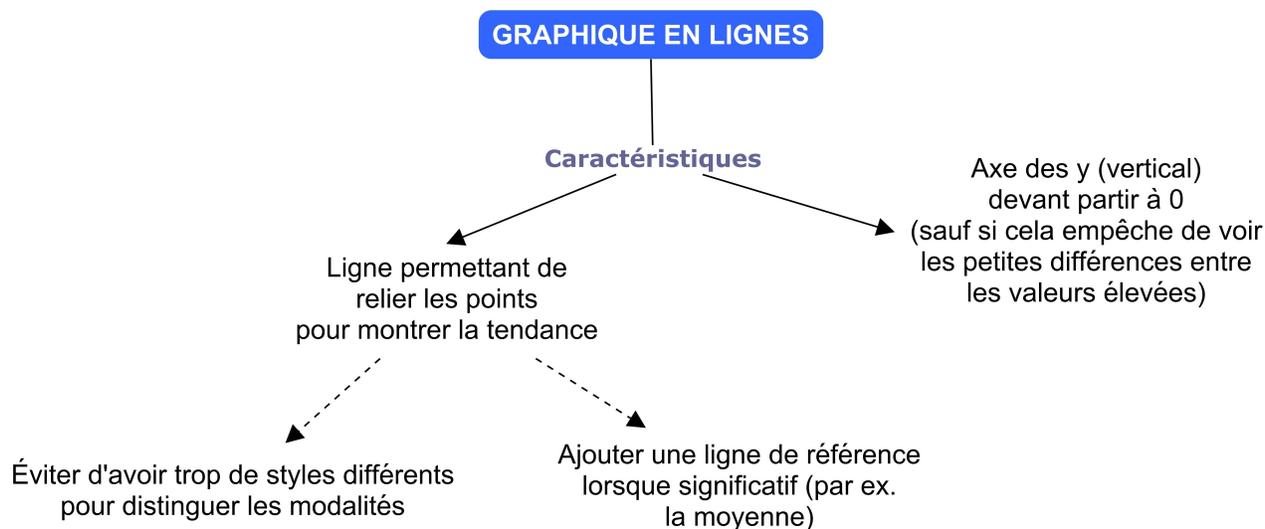


© Christine Dufour 2024, EBSI, Université de Montréal

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c09\\_pe\\_analyse\\_statistique\\_descriptives\\_graphique\\_points.html](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c09_pe_analyse_statistique_descriptives_graphique_points.html)

### Synopsis

Le graphique en points présente un point pour chaque donnée représentée. Si plusieurs catégories sont représentées, différentes formes et couleurs peuvent être utilisées pour les distinguer.

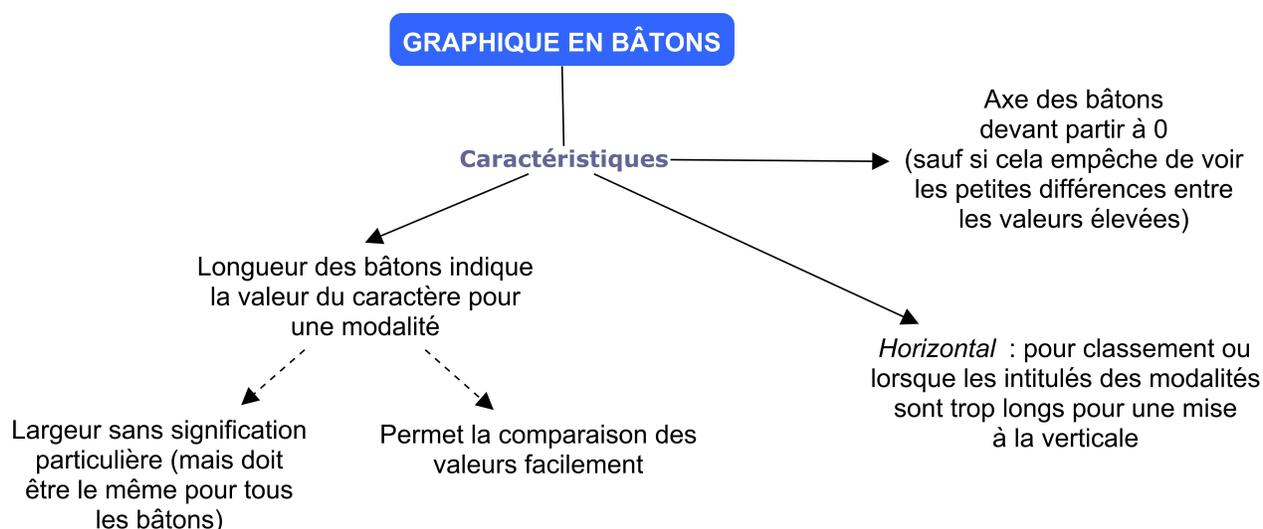
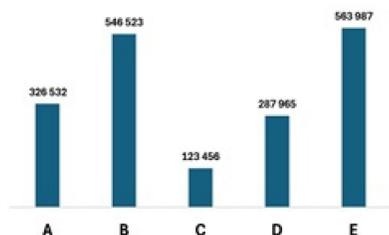


© Christine Dufour 2024, EBSI, Université de Montréal

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c09\\_pe\\_analyse\\_statistique\\_descriptives\\_graphique\\_lignes.html](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c09_pe_analyse_statistique_descriptives_graphique_lignes.html)

### Synopsis

Le graphique en lignes se base sur un graphique en points en ajoutant une ligne entre chaque point. Il permet de mieux faire ressortir certaines tendances.

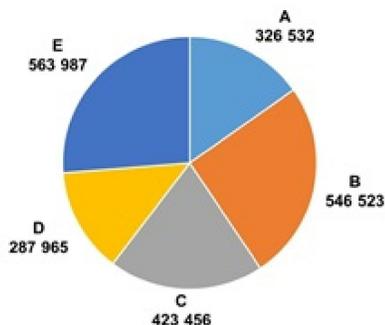


© Christine Dufour 2024, EBSI, Université de Montréal

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c09\\_pe\\_analyse\\_statistique\\_descriptives\\_graphique\\_batons.html](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c09_pe_analyse_statistique_descriptives_graphique_batons.html)

### Synopsis

Le graphique en bâtons présente un bâton par modalité représentée, la hauteur du bâton étant la fréquence de cette modalité.



**GRAPHIQUE EN SECTEURS**

**Caractéristiques**

Permet de mettre en avant une part par rapport aux autres

Ajouter les valeurs des parts pour ne pas que l'évaluation ne repose que sur la comparaison physique des angles des parts

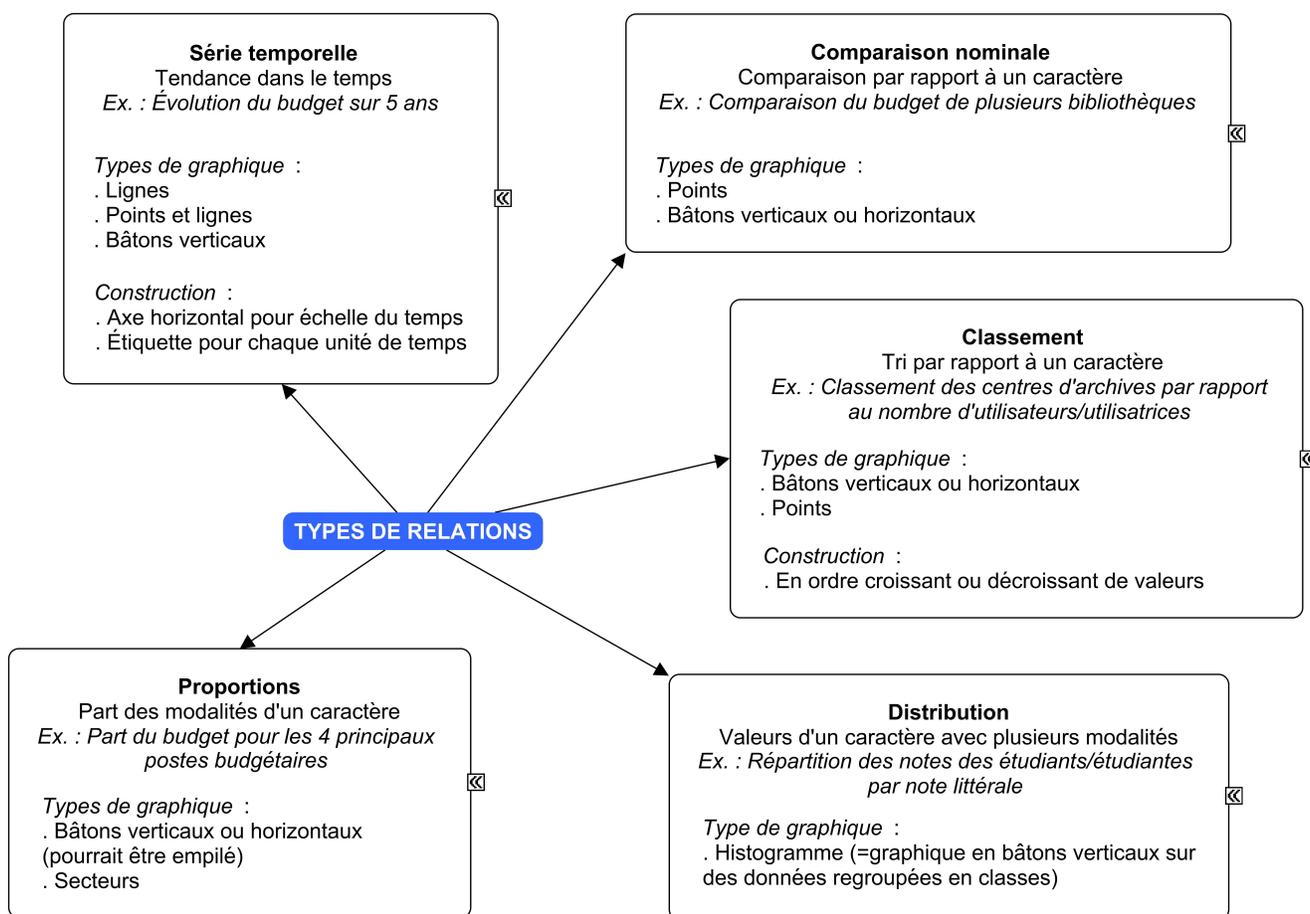
Éviter d'avoir un trop grand nombre de parts (4-5 environ)

© Christine Dufour 2024, EBSI, Université de Montréal

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c09\\_pe\\_analyse\\_statistique\\_descriptives\\_graphique\\_secteurs.html](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c09_pe_analyse_statistique_descriptives_graphique_secteurs.html)

**Synopsis**

Le graphique en secteurs (ou pointes de tarte/camembert!) fait ressortir dans un cercle la proportion représentée par chaque modalité pour un caractère. D'un point de vue ergonomique, certains auteurs le rejettent du fait que la comparaison entre les modalités se base sur la comparaison d'angles (ce qui n'est pas facile). La manière de contourner cette limite est de prendre soin d'afficher les valeurs numériques dans le graphique.

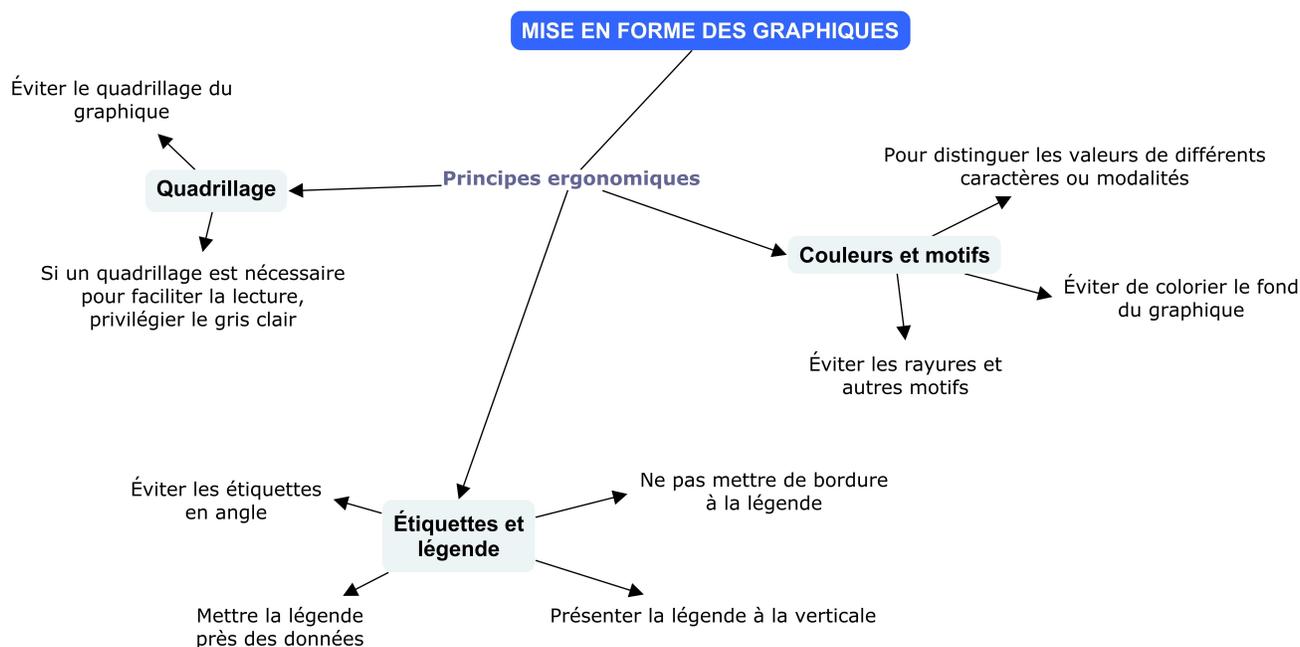


© Christine Dufour 2024, EBSI, Université de Montréal

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c09\\_pe\\_analyse\\_statistique\\_descriptives\\_graphique\\_relations.html](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c09_pe_analyse_statistique_descriptives_graphique_relations.html)

## Synopsis

Les graphiques permettent de représenter différents types de relation (série temporelle, comparaison nominale, classement, proportions, distribution). Pour chaque relation, on retrouve des types de graphique à privilégier. Ce que l'on remarque est que le graphique en bâtons semble le plus polyvalent!

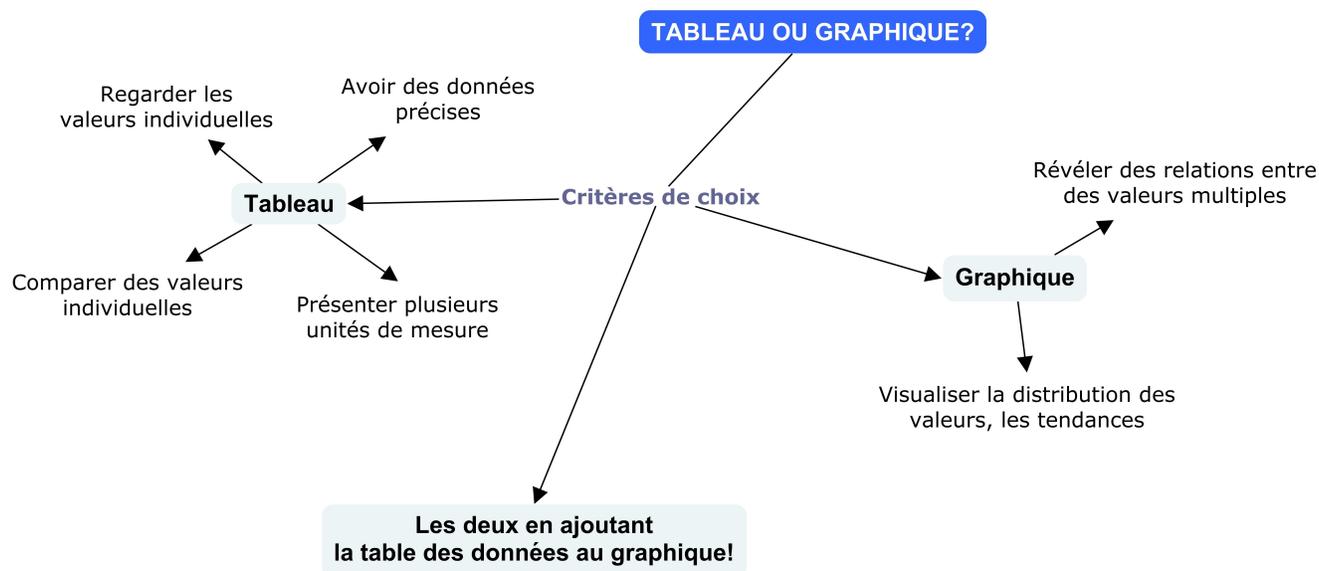


© Christine Dufour 2024, EBSI, Université de Montréal

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c09\\_pe\\_analyse\\_statistique\\_descriptives\\_principes\\_presentation\\_graphique](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c09_pe_analyse_statistique_descriptives_principes_presentation_graphique)

### Synopsis

La mise en forme des graphiques doit aussi respecter certains principes pour en améliorer l'ergonomie. Tout comme pour les tableaux, il faut éviter la surcharge du quadrillage, pour ne pas nuire aux données. Les couleurs peuvent être utilisées pour distinguer des éléments, mais ici aussi il faut éviter la surcharge. Les étiquettes doivent demeurer lisibles (donc horizontales).



© Christine Dufour 2024, EBSI, Université de Montréal

[https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007\\_c09\\_pe\\_analyse\\_statistique\\_descriptives\\_principes\\_choix\\_tableau\\_graphique](https://cours.ebsi.umontreal.ca/sci6007/cartes/sci6007_c09_pe_analyse_statistique_descriptives_principes_choix_tableau_graphique)

### Synopsis

Tableaux et graphiques sont deux types de représentation utiles pour les statistiques descriptives. Les tableaux sont à privilégier tout particulièrement lorsque l'on veut des valeurs précises. Les graphiques, par leur dimension visuelle, peuvent permettre entre autres de mieux faire ressortir les tendances. Cela dit, on peut dans certains cas utiliser l'un tout comme l'autre, voire même les deux! Il faut aussi se rappeler que certaines personnes sont plus textuelles, et préféreront les tableaux, tandis que d'autres sont visuelles et aimeront mieux les graphiques. En essayant de varier entre tableaux et graphiques, vous réussirez ainsi à rejoindre le plus grand nombre de personnes dans leurs préférences.