

Statistique graphique



Ecole d'Été sur
**L'INTELLIGENCE
ARTIFICIELLE**

24 juillet 2023

Plan

1- Qu'est-ce-que la statistique graphique?

2- Objectifs de la statistique graphique

3- Notions importantes pour lire, interpréter ou construire les graphiques

4- Eléments de décision

5- Quelle histoire souhaite t-on raconter?

6- Cas d'utilisation particuliers des graphiques

7- Qu'est ce qu'on peut faire avec un graphique?

8- Quelques bonnes pratiques

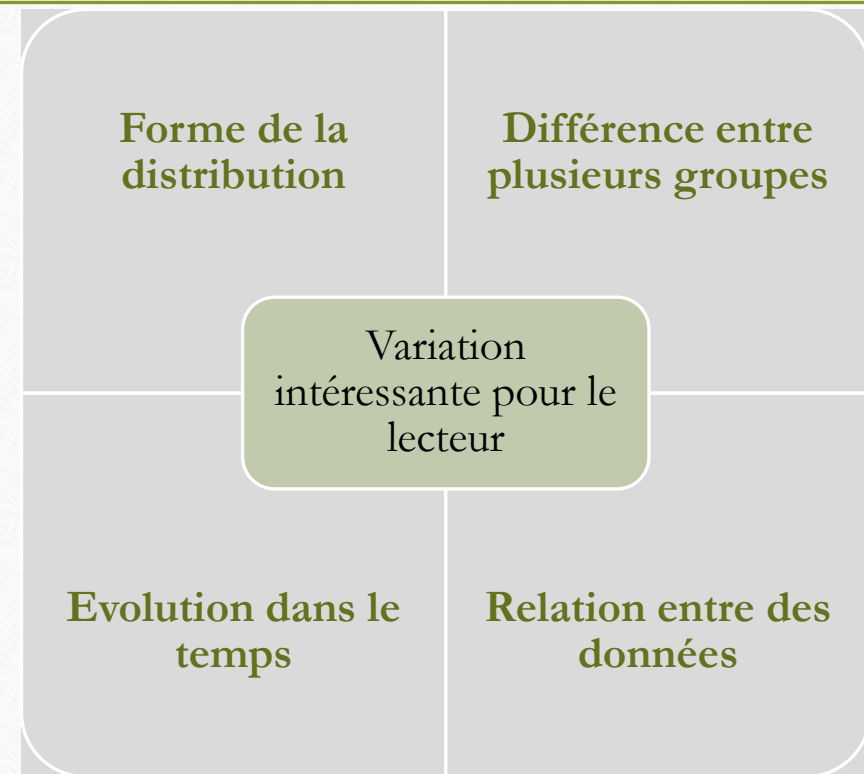
9- Quels outils & comment ?

1- Qu'est-ce-que la statistique graphique?

- **Un graphique statistique** : figure fondée sur des données recueillies dans des populations ou des échantillons – Représentation visuelle des variables d'une population



- Variation **nouvelle** ou **inattendue** mise en lumière par la représentation en partant de données complexes



2- Objectifs de la statistique graphique

- **Synthétiser:** En présence de nombreuses données, le cerveau humain a beaucoup de mal à en dégager la configuration. En revanche, avec une représentation visuelle, tout s'éclaire en une seconde !
- **Dimension visuelle:** Mémorisation plus simple des faits illustrés car 60 % de visuels parmi les humain et parce que les faits présentés sont sélectionnés
- Accompagne un ou des tableau (x)

3- Notions importantes pour lire, interpréter ou construire les graphiques

Population: Ensemble d'éléments similaires faisant l'objet d'un phénomène particulier

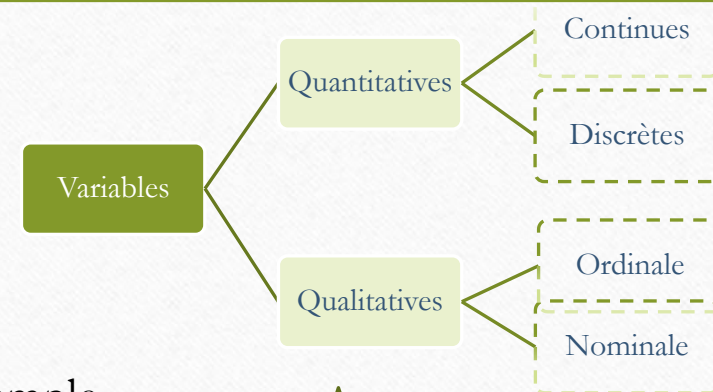
Variable / Caractère: Caractéristique qui permet de différencier les individus de la population

Spécifique aux variables qualitatives

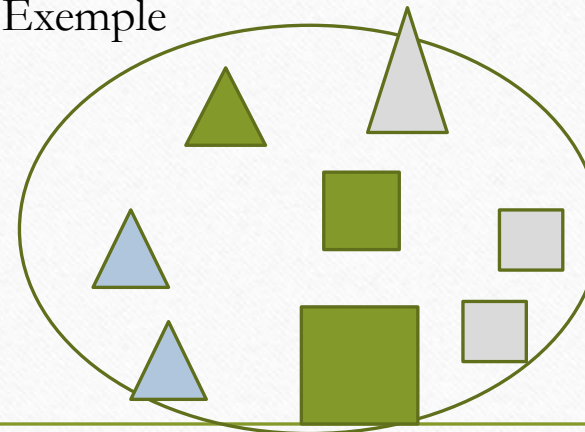
Modalités: Valeurs possibles d'une variable qualitative

Effectifs: Nombre d'individus présentant une modalité

Fréquence: Effectif d'une modalité / Effectif total



Exemple



3- Notions : Les paramètres de tendances centrale – Variable quantitative

- Permettent de résumer l'ensemble de la distribution d'une variable en un seul nombre
- **Mode** : Valeur la plus représentée
- **Médiane** : Valeur qui sépare la distribution en deux parties égales
- **Moyenne** : Valeur de la variable si toutes les observations étaient similaires

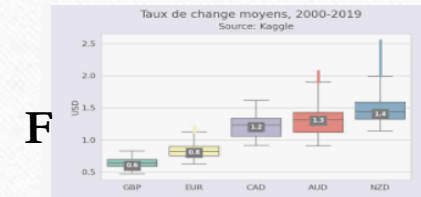
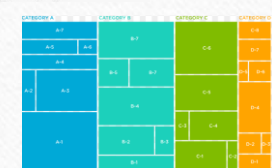
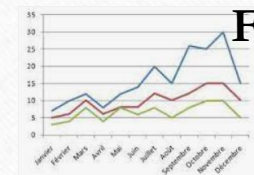
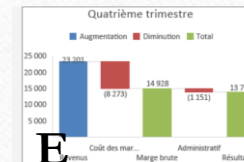
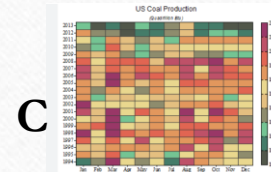
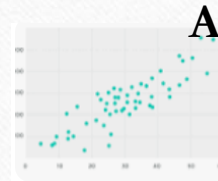
Individu	Variable
A	12
B	13
C	18
D	15
E	13

4- Éléments de décision

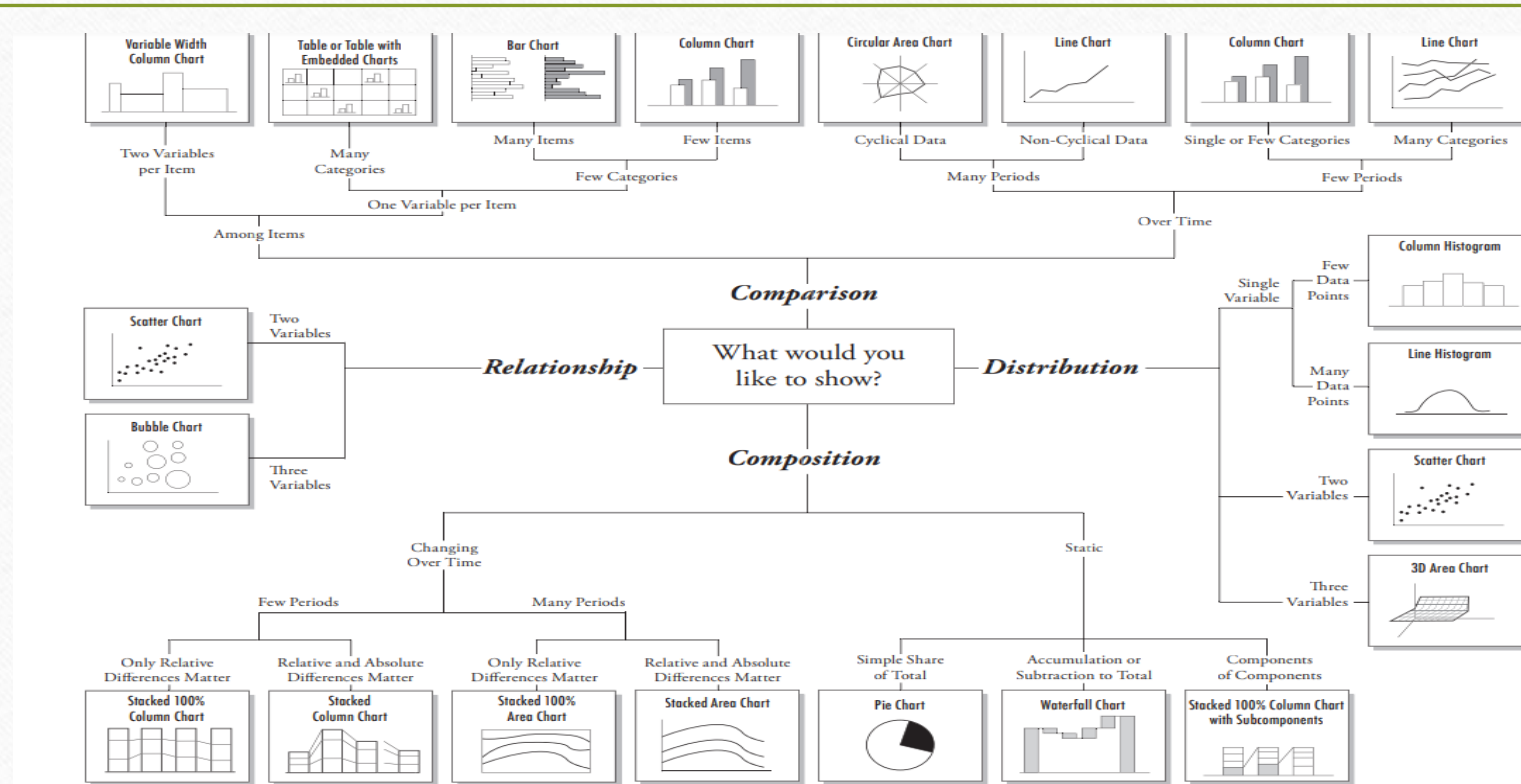
- Type de variables
- Nombre de variables
- Lien entre les variables
- Question métier

5- L'histoire - Les représentations graphiques

- Diagramme en bâtons , circulaire, histogramme
- Courbe, surface
- Nuage de points
- Treemap/ Graphique hiérarchique
- Graphique en cascade
- Carte de chaleur
- Boîtes à moustaches



5- L'histoire - Que raconte-t-on?

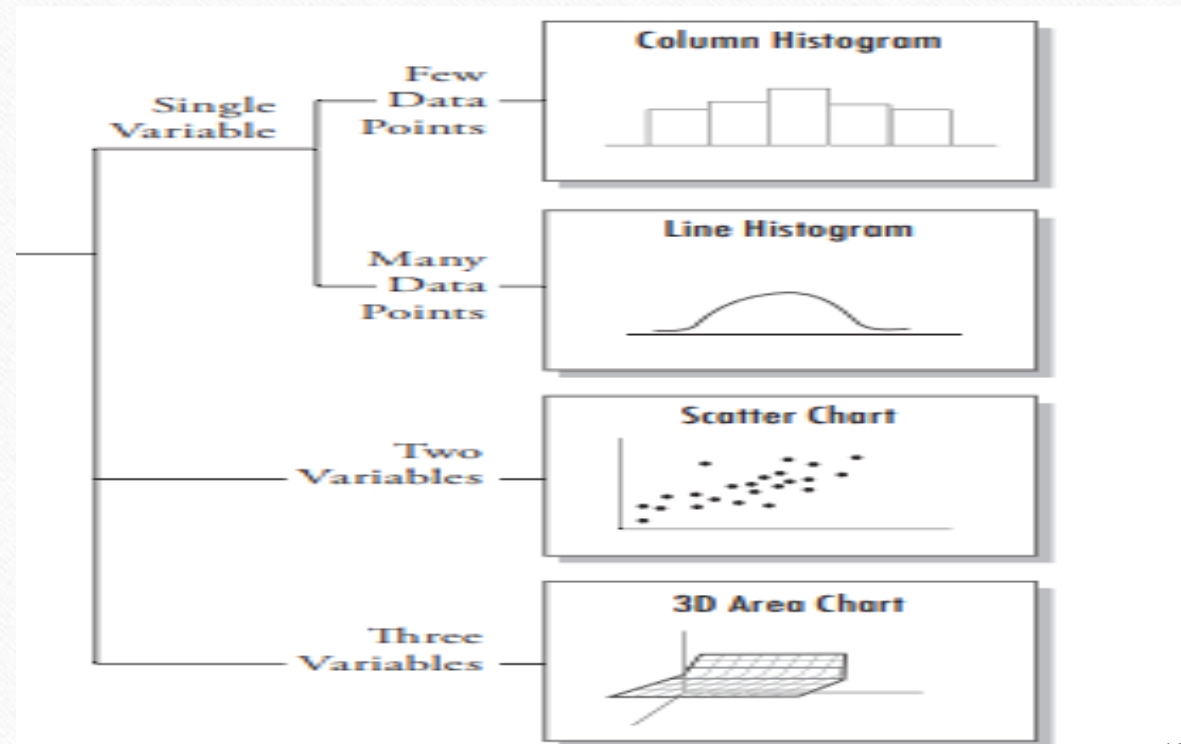


5- L'histoire – Cas d'une distribution

Questions métier associées:

- Comment se réparti....?
- Quel est le maximum, le minimum, le mode?
- Quelle est la progression de la fréquence de ma grandeur?
- Est-ce que ma(mes) variable(s) suit (vent) une loi connue?

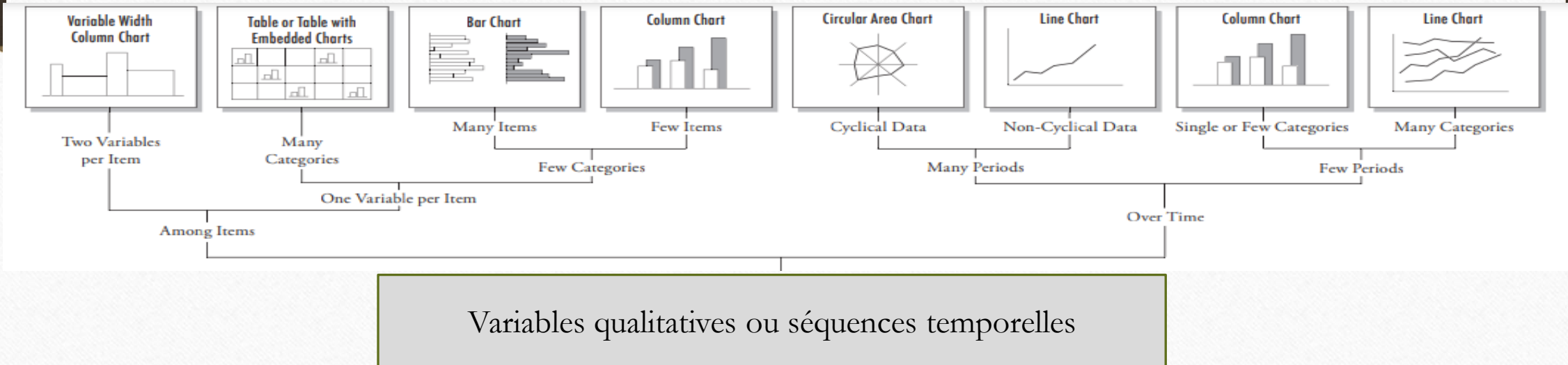
Variables quantitatives continues



5- L'histoire – Cas d'une comparaison

Questions métier associées:

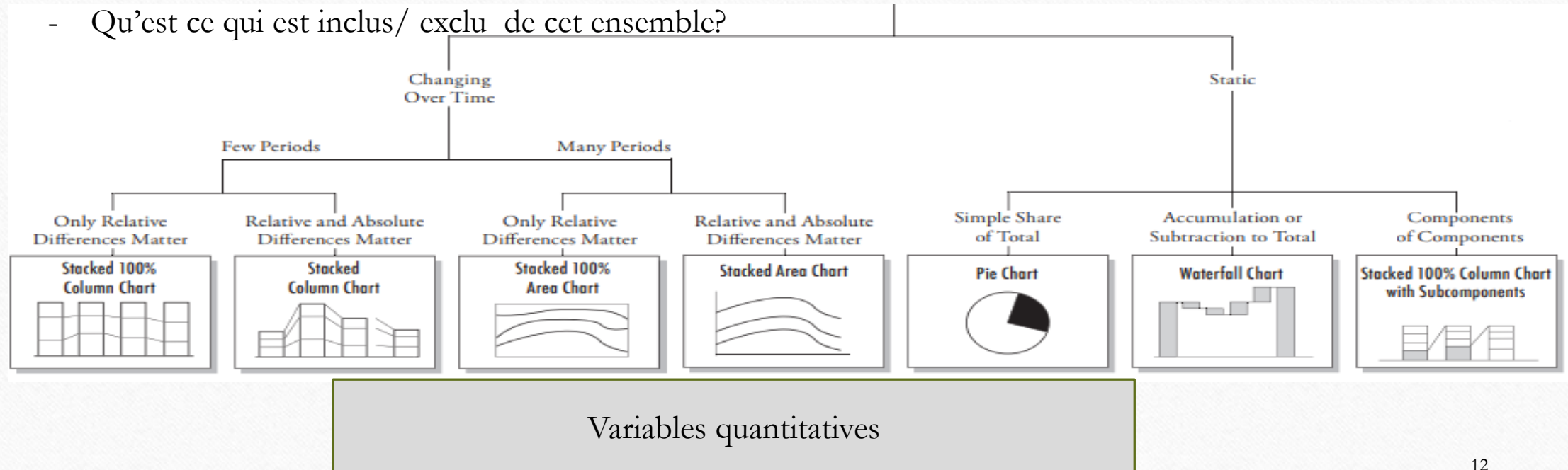
Quel est le groupe ou la période le (la) plus/ le moins (grand, favorable, productif...)?



5- L'histoire – Cas d'une composition

Questions métier associées:

- Que contient X?
- Qu'est ce qui est inclus/ exclu de cet ensemble?



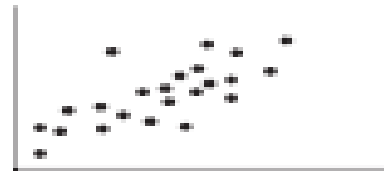
5- L'histoire – Cas d'une relation

Questions métier associées:

- Quel est l'impact ou l'effet de X sur Y?(croissance ou décroissance)
- Qu'est ce qui entraîne ceci?
- Est-ce que X est lié à Y?

Variables quantitatives continues et qualitatives

Scatter Chart



Two Variables

Bubble Chart



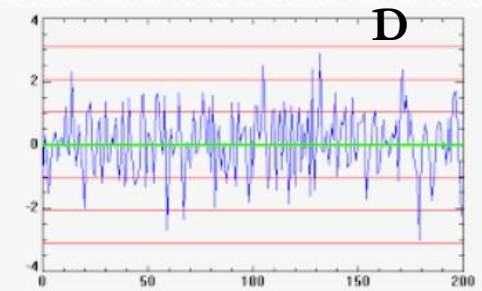
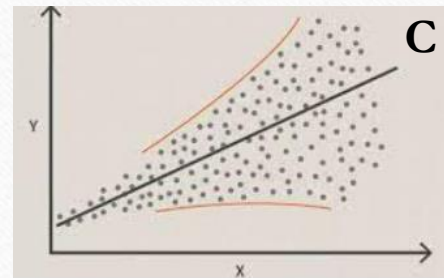
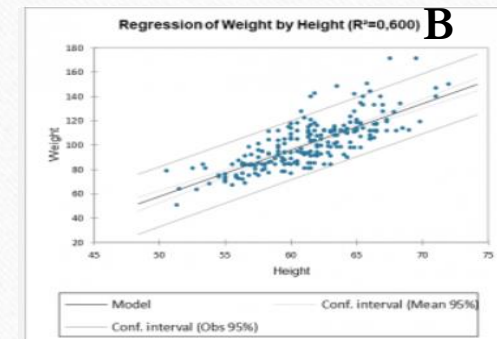
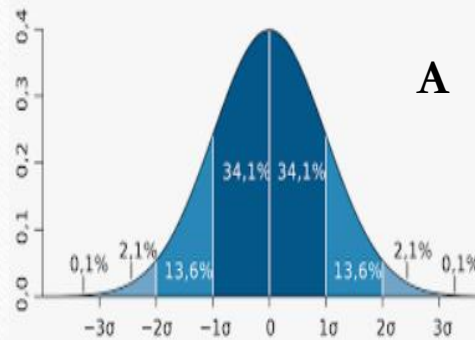
Three Variables

6- Cas d'utilisation particuliers des graphiques

- Le graphique en radar et le graphique sémantique différentiel en marketing et en psychologie.
- Le graphique en aires montre l'évolution d'un volume comme dans le cas du suivi de la masse salariale
- Les chandeliers sont des graphiques typiquement boursiers.
- Les pyramides des âges sont des sortes de doubles histogrammes
- Graphiques mixtes, superposant par exemple une courbe et un diagramme en barres montrent deux séries différentes dans un même repère.

6- Représentations caractéristiques en statistique

- Loi normale
- Régression linéaire
- Distribution symétrique
- Hétéroscédasticité
- Bruit blanc



7- Qu'est ce qu'on peut faire avec un graphique?

- Calculer un mode
- Déterminer une droite et un coefficient d'ajustement
- Identifier des symétries / dyssymétrie dans une distribution, des biais
- Hétéroscédasticité / analyse de la variance
- Extraire des données d'un histogramme par exemple
- Déterminer la médiane avec les courbes d'effectifs cumulés croissant et décroissant
- Identifier une corrélation entre deux variables

8- Quelques bonnes pratiques

- Proportions => Attention à la taille de l'échantillon, aux données manquantes
- Aux échelles multiplicatives ou additives, à respecter
- Variable dépendante en ordonnée
- Niveau de mesure et de continuité de la variable / Nature quantitative et qualitative à respecter

- Couleurs
- Annotations
- Unités

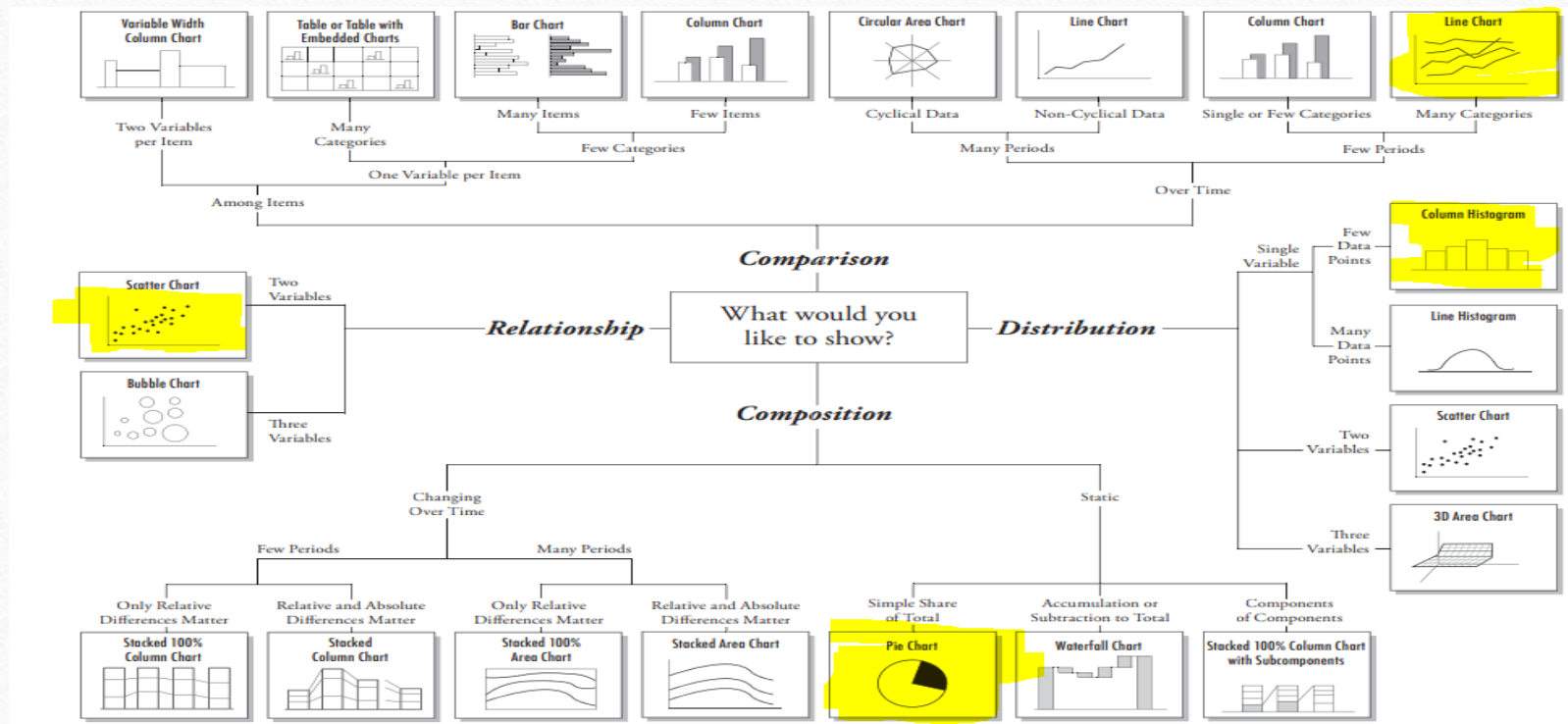
9- Quels outils & comment ?

- Excel, R, Python, les outils de datavisualisation (par exemple Power BI)

- Les librairies importantes sous Python

Panda, Matplotlib, Seaborn

9- Quels outils & comment ?



Jeu de données

Sujet: Le réchauffement climatique

Area Code	Area	Months Code	Months	Element	Element	Unit	Y1961
2	Afghanistan	7001	January	7271	Tempera	°C	0.777000
2	Afghanistan	7001	January	6078	Standard	°C	1.950000
2	Afghanistan	7002	February	7271	Tempera	°C	-1.743000
2	Afghanistan	7002	February	6078	Standard	°C	2.597000
2	Afghanistan	7003	March	7271	Tempera	°C	0.516000
2	Afghanistan	7003	March	6078	Standard	°C	1.513000

Variables quantitatives	Variables qualitatives

Courbe

Sujet: Le réchauffement climatique

Area Code	Area	Months Code	Months	Elem	Element	Unit	Y1961
2	Afghanistan	7001	January	7271	Tempera	°C	0.777000
2	Afghanistan	7001	January	6078	Standard	°C	1.950000
2	Afghanistan	7002	February	7271	Tempera	°C	-1.743000
2	Afghanistan	7002	February	6078	Standard	°C	2.597000
2	Afghanistan	7003	March	7271	Tempera	°C	0.516000
2	Afghanistan	7003	March	6078	Standard	°C	1.512000



Area Code	Area	Months Code	Months	Année	Temperature
2	Afghanistan	7001	January	1961	0,78
2	Afghanistan	7002	February	1961	-1,74
2	Afghanistan	7003	March	1961	0,52
2	Afghanistan	7004	April	1961	-1,71
2	Afghanistan	7005	May	1961	1,41
2	Afghanistan	7006	June	1961	-0,06
2	Afghanistan	7007	July	1961	0,88

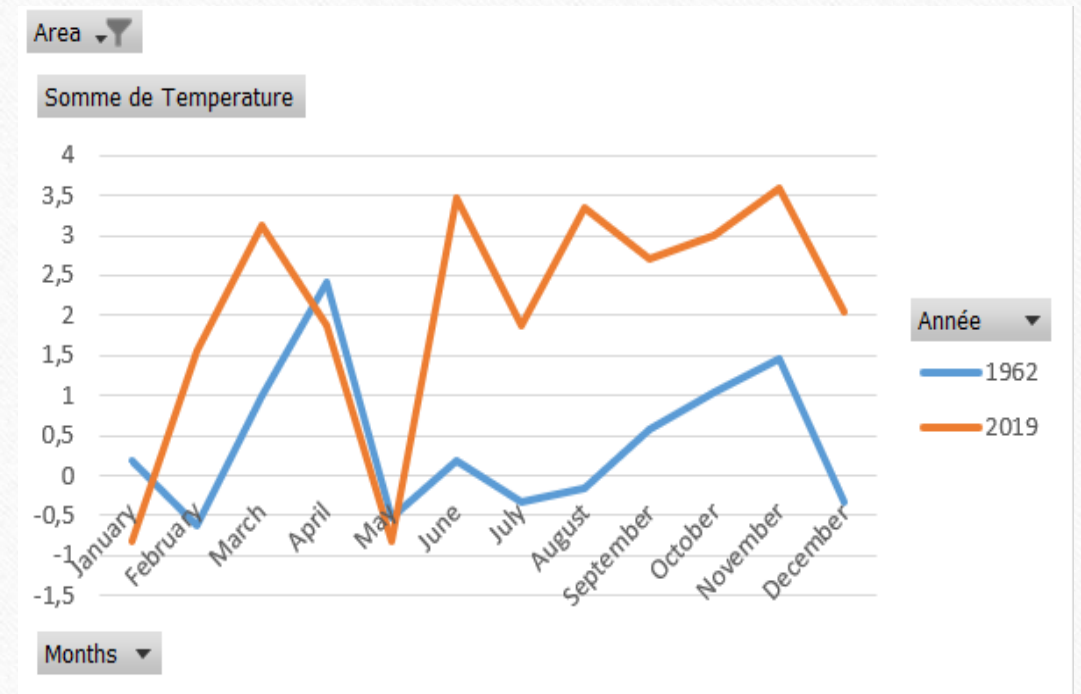


Diagramme circulaire

Sujet: La qualité du sommeil

Person ID	Gender	Age	Occupation	Sleep Duration	Quality
1	Male	27	Software Engineer		6,10
2	Male	28	Doctor		6,20
3	Male	28	Doctor		6,20
4	Male	28	Sales Representative		5,90
5	Male	28	Sales Representative		5,90

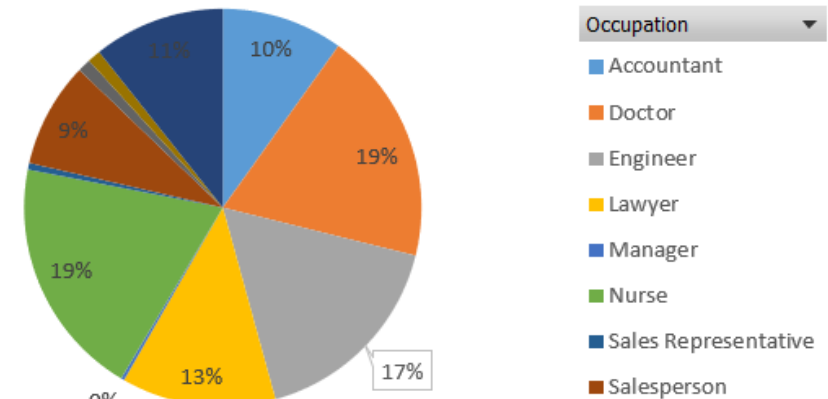


Tableau croisé dynamique

Étiquettes de lignes	Nombre de Occupation	Nombre de Occupation2
Accountant	37	9,89%
Doctor	71	18,98%

Nombre de Occupation Nombre de Occupation2

Effectif en fonction de la profession



Valeurs

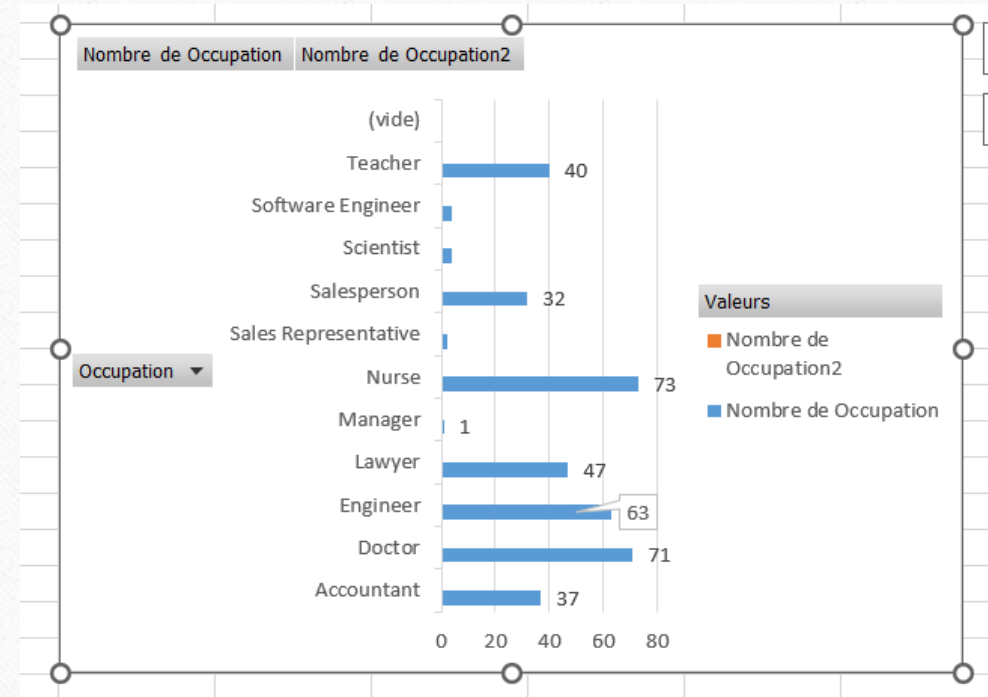
Diagramme en colonne

Sujet: La qualité du sommeil

Person ID	Gender	Age	Occupation	Sleep Duration	Quality
1	Male	27	Software Engineer	6,10	
2	Male	28	Doctor	6,20	
3	Male	28	Doctor	6,20	
4	Male	28	Sales Representative	5,90	
5	Male	28	Sales Representative	5,90	



Étiquettes de lignes	Nombre de Occupation	Nombre de Occupation2
Accountant	37	9,89%
Doctor	71	18,98%

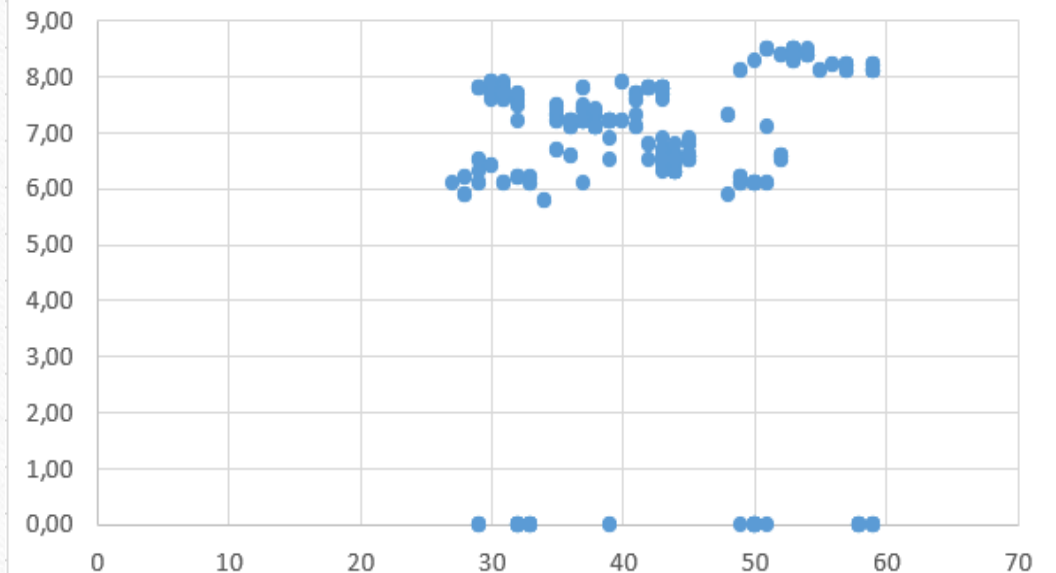


Nuage de points

Sujet: La qualité du sommeil

Person ID	Gender	Age	Occupation	Sleep Duration
1	Male	27	Software Engineer	6,10
2	Male	28	Doctor	6,20
3	Male	28	Doctor	6,20
4	Male	28	Sales Representative	5,90
5	Male	28	Sales Representative	5,90
6	Male	28	Software Engineer	5,90
7	Male	29	Teacher	6,30
8	Male	29	Doctor	7,80
9	Male	29	Doctor	7,80
10	Male	29	Doctor	7,80
11	Male	29	Doctor	6,10

Qualité du sommeil en fonction de l'age



Boite à moustaches

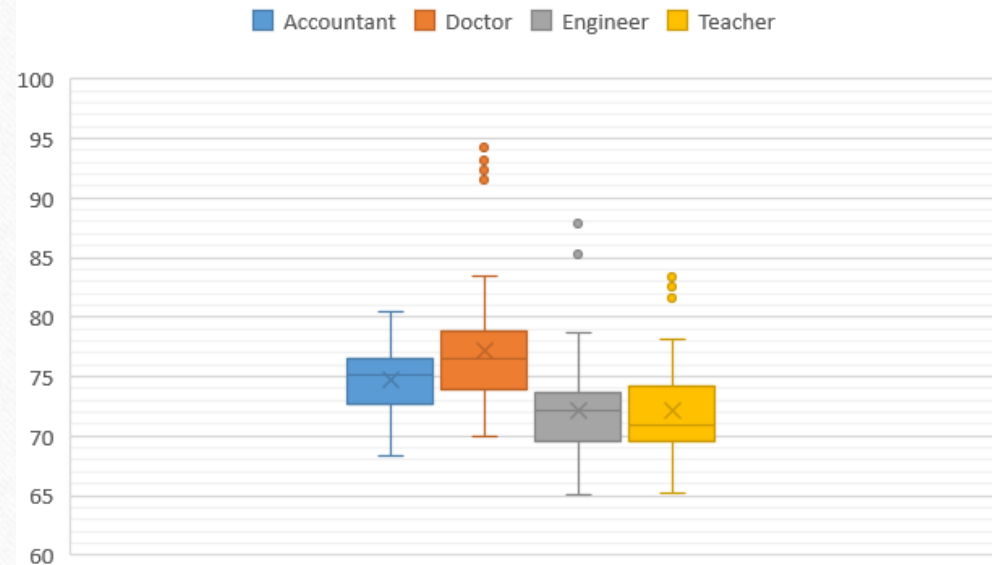
Sujet: La qualité du sommeil

Occupation	Heart Rate
Software Engineer	77
Doctor	75
Doctor	75
Sales Representative	85



Accountant	Doctor	Engineer	Teacher
75,6	83,4	71,9	82,8
71,5	76,3	78,3	73,7
75,1	73,5	69,4	75,3
74	78,7	73,3	69,5
71,5	77	71,3	69,6

Distribution du nombre de battements cardiaques en fonction de la profession



Démo Power BI

Boite à moustaches

Sujet: La qualité du sommeil

