



Gestion de Projet Big Data & Développement d'applications Big Data

EDAH Kodjo
Consultant Systèmes d'Information, Big-Data



Objectifs

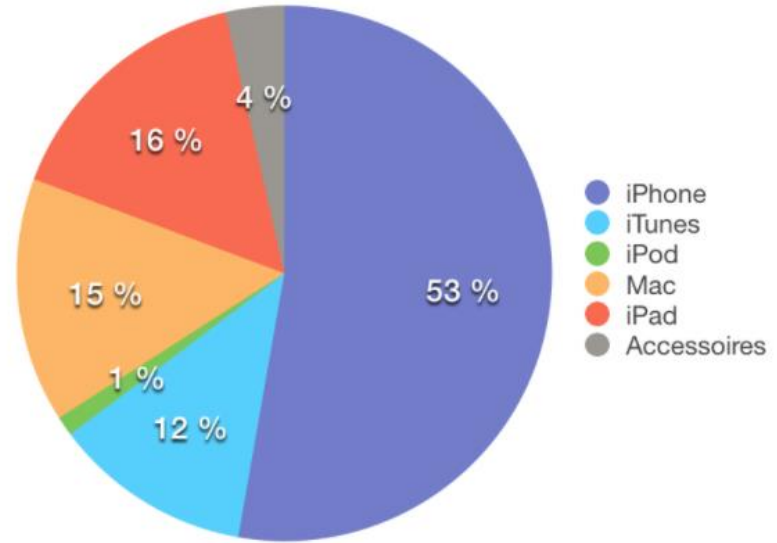
- Comprendre la notion et les spécificités du Big Data
- Connaître les outils de collecte, de traitement et d'exploitation des données
- Savoir utiliser les outils de visualisation des données (Dataviz)
- Piloter et maîtriser les risques des projets



Partie 3 : La visualisation des données

Visualisation des données : c'est quoi la dataviz

- ❑ La data visualisation représente graphiquement les données.
- ❑ Elle permet d'appréhender des concepts, tendances ou motifs difficiles à voir via les données brutes et de communiquer l'information



Visualisation des données : dataviz

❑ Combien de chiffres 6 comptez-vous ?

3 5 1 0 8 0 4 2 3 5 1 0 9 4 2 3 2 1 4 8 3 5 1 2 3 2 1 2 3 9 1
0 4 2 3 5 1 0 9 4 2 3 5 0 9 4 0 9 4 2 3 4 2 3 4 0 9 4 4 0 9 4
4 9 8 1 8 5 0 4 2 3 5 1 0 9 4 2 3 8 1 4 2 3 5 4 2 3 7 4 2 3 0
4 2 3 5 1 0 8 0 7 8 9 2 3 0 9 0 0 3 2 1 7 8 9 9 0 0 3 9 0 0 3
4 2 3 5 1 0 1 2 3 4 5 9 4 2 3 5 0 9 4 0 3 4 5 3 5 0 9 3 5 0 9
1 5 9 4 2 3 5 1 0 8 3 7 8 9 2 7 1 7 2 3 0 8 0 9 7 8 7 5 7 1 7
2 3 0 9 0 0 1 6 9 8 2 8 5 0 4 2 3 5 6 0 9 8 5 4 2 3 5 4 2 3 5
9 4 2 3 5 6 5 0 2 3 5 1 0 8 1 7 6 9 2 3 2 3 5 1 7 8 9 1 7 8 9
1 0 9 4 2 3 3 1 4 8 3 5 1 0 8 0 4 2 3 5 4 8 3 8 0 4 2 8 0 4 2
3 5 0 9 4 0 9 4 2 3 0 4 2 3 5 1 0 9 4 2 2 3 0 5 1 0 9 5 1 0 9
5 1 0 9 4 2 3 5 1 4 4 9 8 4 8 5 0 4 2 3 1 4 4 8 5 0 4 8 5 0 4
9 2 3 0 9 0 0 3 2 1 4 2 3 6 1 0 6 9 7 8 2 1 4 6 0 8 3 1 0 8 0
5 9 4 2 3 5 0 9 4 0 4 2 3 5 1 0 1 2 3 4 4 0 4 1 0 1 2 1 0 1 2
0 7 8 9 1 7 3 7 2 3 1 5 9 4 2 3 5 1 0 8 2 3 1 2 3 5 1 2 3 5 1
0 8 5 0 4 2 3 5 1 0 2 3 0 9 0 0 1 9 9 8 1 0 2 0 0 1 9 0 0 1 9
5 1 0 8 7 7 8 9 2 3 9 4 2 3 5 0 5 0 2 6 2 3 9 5 0 5 0 5 0 5 0
3 9 1 4 8 3 5 1 0 9 4 2 3 7 0 4 2 3 5 1 0 9 4 0 4 2 3 0 4 2 3
0 9 4 2 3 0 6 3 5 0 9 4 0 9 1 0 9 4 2 3 5 0 9 1 0 9 4 1 0 9 4
2 3 7 1 4 4 9 5 1 0 9 4 2 3 5 0 4 2 3 5 1 0 9 5 0 4 2 5 0 4 2
0 0 3 2 1 4 2 9 2 3 0 9 0 0 0 8 9 7 8 9 2 3 0 0 8 1 7 0 8 4 7

Source : toucantoco.com



Visualisation des données : dataviz

❑ Combien de chiffres 6 comptez-vous ?

3 5 1 0 8 0 4 2 3 5 1 0 9 4 2 3 2 1 4 8 3 5 1 2 3 2 1 2 3 9 1
0 4 2 3 5 1 0 9 4 2 3 5 0 9 4 0 9 4 2 3 4 2 3 4 0 9 4 4 0 9 4
4 9 8 1 8 5 0 4 2 3 5 1 0 9 4 2 3 8 1 4 2 3 5 4 2 3 7 4 2 3 0
4 2 3 5 1 0 8 0 7 8 9 2 3 0 9 0 0 3 2 1 7 8 9 9 0 0 3 9 0 0 3
4 2 3 5 1 0 1 2 3 4 5 9 4 2 3 5 0 9 4 0 3 4 5 3 5 0 9 3 5 0 9
1 5 9 4 2 3 5 1 0 8 3 7 8 9 2 7 1 7 2 3 0 8 0 9 7 8 7 5 7 1 7
2 3 0 9 0 0 1 6 9 8 2 8 5 0 4 2 3 5 6 0 9 8 5 4 2 3 5 4 2 3 5
9 4 2 3 5 6 5 0 2 3 5 1 0 8 1 7 6 9 2 3 2 3 5 1 7 8 9 1 7 8 9
1 0 9 4 2 3 3 1 4 8 3 5 1 0 8 0 4 2 3 5 4 8 3 8 0 4 2 8 0 4 2
3 5 0 9 4 0 9 4 2 3 0 4 2 3 5 1 0 9 4 2 2 3 0 5 1 0 9 5 1 0 9
5 1 0 9 4 2 3 5 1 4 4 9 8 4 8 5 0 4 2 3 1 4 4 8 5 0 4 8 5 0 4
9 2 3 0 9 0 0 3 2 1 4 2 3 6 1 0 6 9 7 8 2 1 4 6 0 8 3 1 0 8 0
5 9 4 2 3 5 0 9 4 0 4 2 3 5 1 0 1 2 3 4 4 0 4 1 0 1 2 1 0 1 2
0 7 8 9 1 7 3 7 2 3 1 5 9 4 2 3 5 1 0 8 2 3 1 2 3 5 1 2 3 5 1
0 8 5 0 4 2 3 5 1 0 2 3 0 9 0 0 1 9 9 8 1 0 2 0 0 1 9 0 0 1 9
5 1 0 8 7 7 8 9 2 3 9 4 2 3 5 0 5 0 2 6 2 3 9 5 0 5 0 5 0 5 0
3 9 1 4 8 3 5 1 0 9 4 2 3 7 0 4 2 3 5 1 0 9 4 0 4 2 3 0 4 2 3
0 9 4 2 3 0 6 3 5 0 9 4 0 9 1 0 9 4 2 3 5 0 9 1 0 9 4 1 0 9 4
2 3 7 1 4 4 9 5 1 0 9 4 2 3 5 0 4 2 3 5 1 0 9 5 0 4 2 5 0 4 2
0 0 3 2 1 4 2 9 2 3 0 9 0 0 0 8 9 7 8 9 2 3 0 0 8 1 7 0 8 4 7

Source : toucantoco.com



Visualisation des données : dataviz

❑ Quel pays est le plus visité ?

Visits per countries

	2010	2011	2012	2013	2014
U.K.	30	35	40	25	30
Belgium	10	15	20	20	15
France	35	40	20	20	25
Italy	10	10	15	10	20
Norway	20	25	25	35	45
Spain	5	15	10	15	20
Sweden	20	30	30	45	40
Germany	40	50	40	35	40
Finland	35	40	40	35	45
Danemark	5	5	15	20	20

Source : toucantoco.com



Visualisation des données : dataviz

❑ Quel pays est le plus visité ?

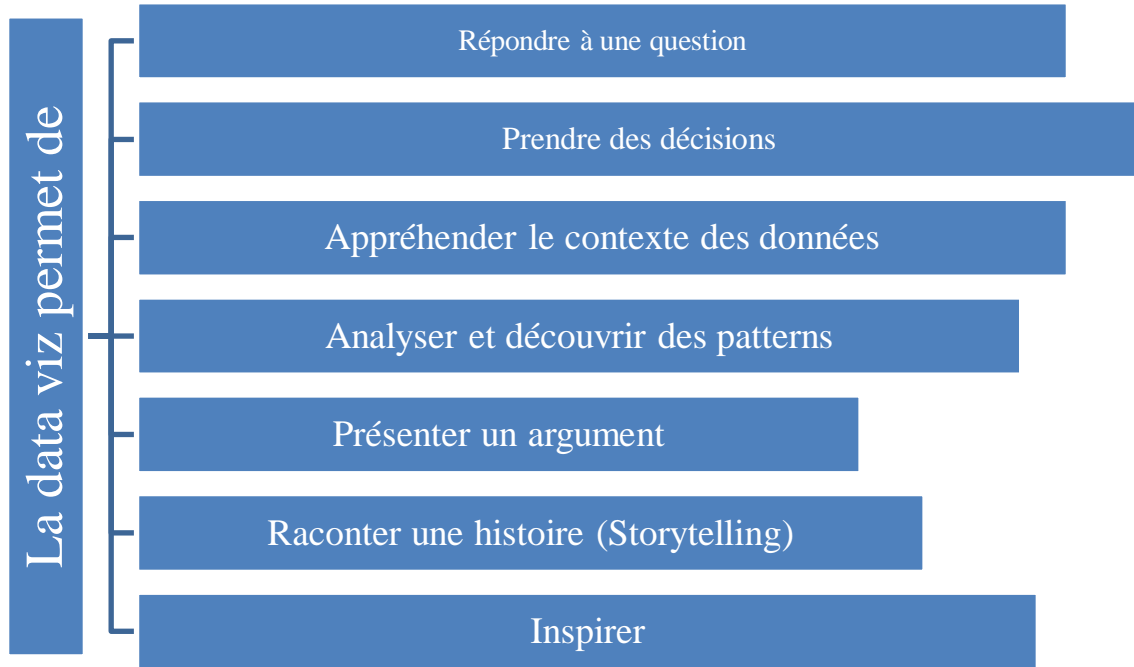


Source : toucantoco.com



Pourquoi la dataviz

- ❑ En Big Data, la volumétrie de données rend sa compréhension encore plus difficile



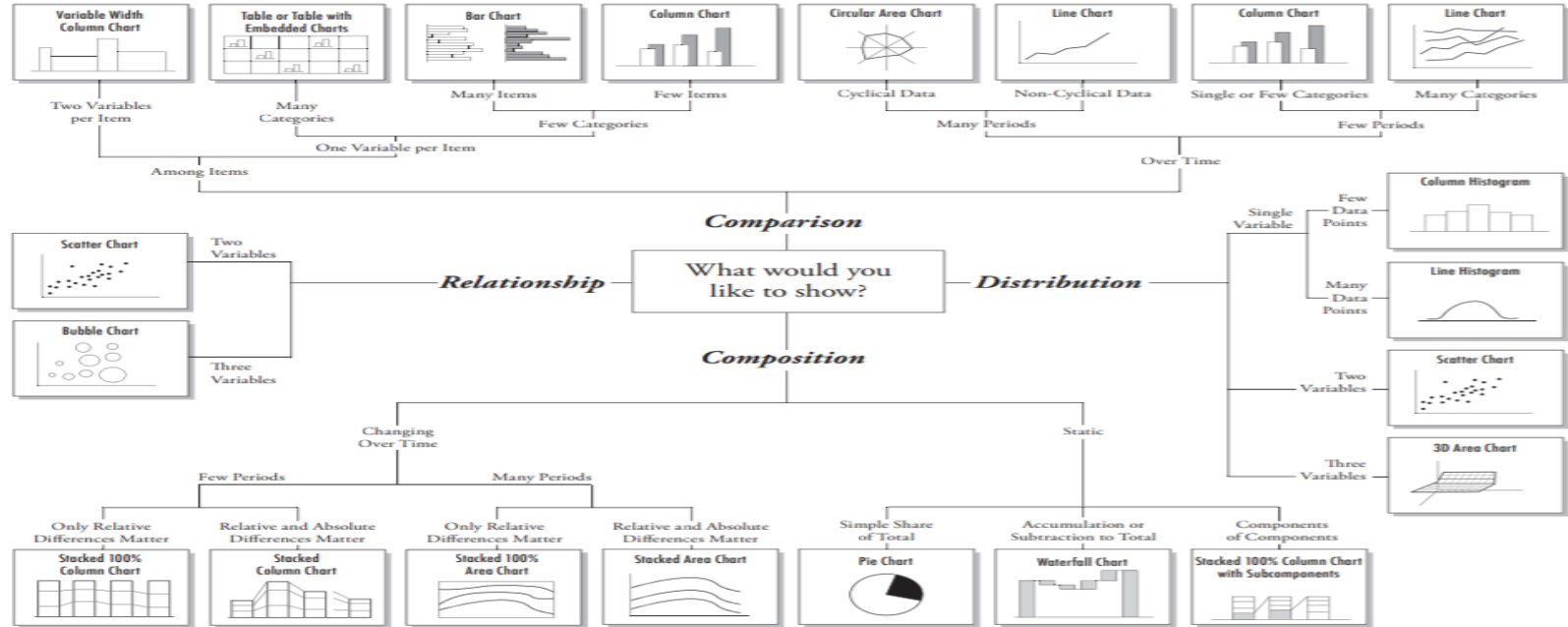
Les questions à se poser ?

- ❑ Quelle est l'information à communiquer ?
 - ❑ Mettre en valeur une information en particulier ?
 - ❑ Nécessité d'utiliser plusieurs graphiques ?
- ❑ Quels formats utiliser ?
 - ❑ Les données numériques sont importantes ?
 - ❑ Montrer une tendance ?
- ❑ Conception du design
 - ❑ Axes
 - ❑ Légendes
 - ❑ Couleurs
 - ❑ Grilles



Quel type de présentation utilisé ?

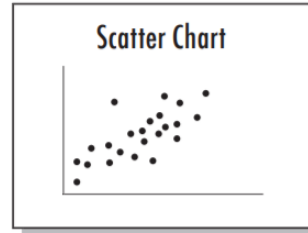
Chart Suggestions—A Thought-Starter



Relation

❑ Scatter chart : nuage de point

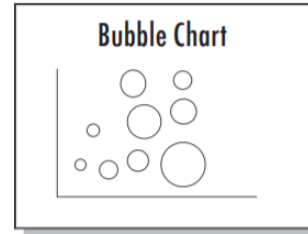
- ❑ Mettre en évidence le degré de corrélation entre au moins deux variables liées.
- ❑ Visualiser des tendances, des dépendances, des relations (positives, négatives, directes, indirectes ou inverses), des répartitions plus ou moins homogènes, des données aberrantes s'écartant de l'écart type.



Two
Variables

❑ Bubble chart

- ❑ Graphique à bulles
- ❑ Affichage de 3 dimensions (v_1 , v_2 , v_3), Deux v_i sur les axes x et y, la troisième dimension la taille de la bulle



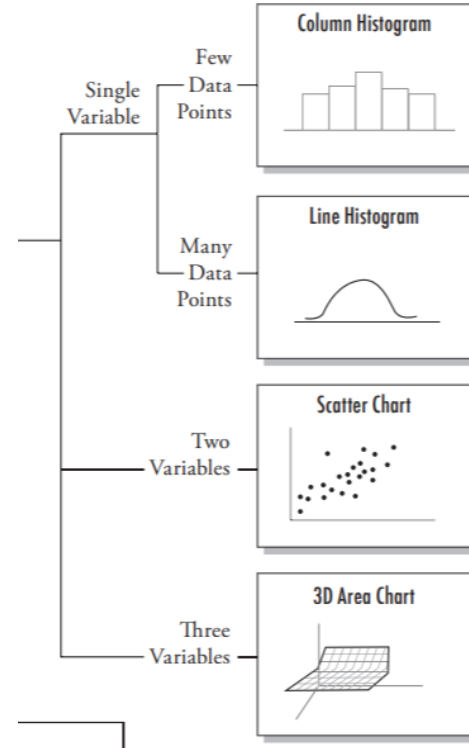
Three
Variables

Relationship

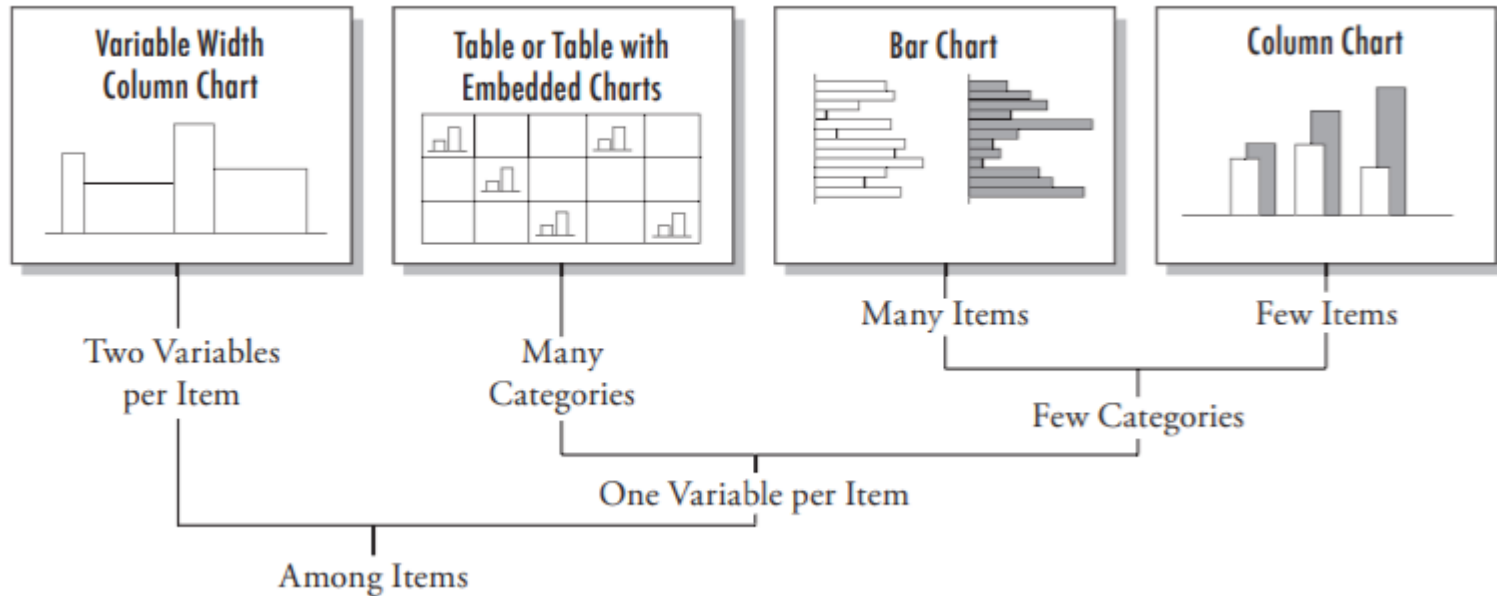


Distribution

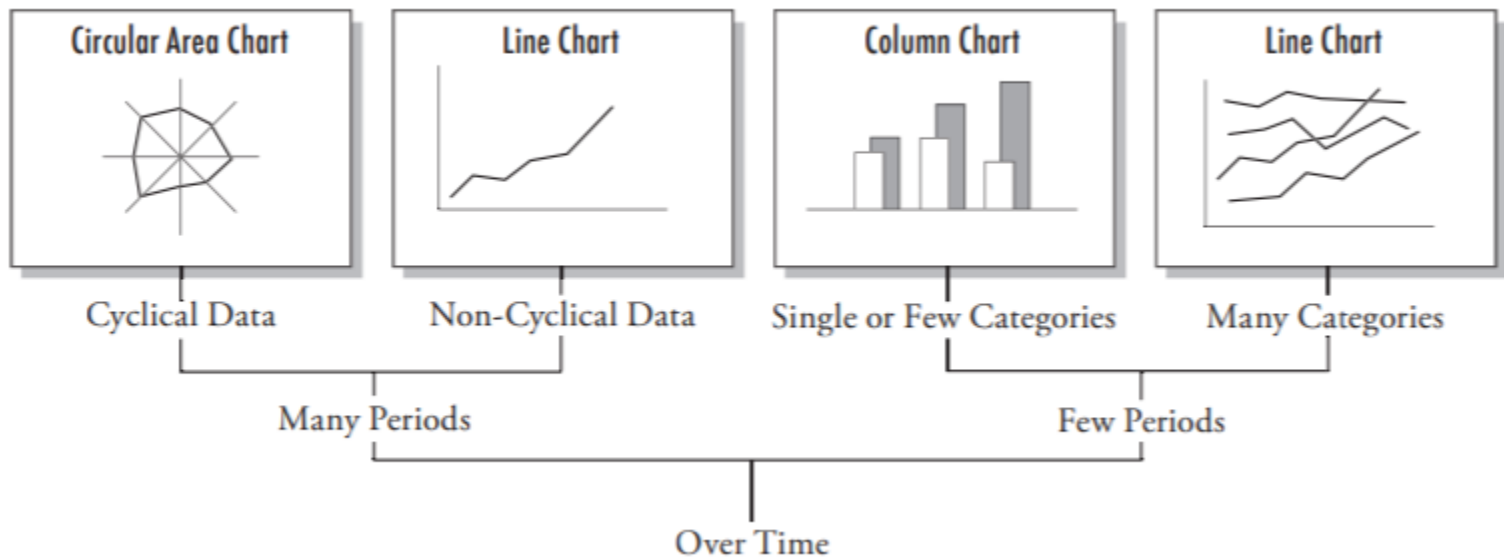
- ❑ Column histogram : histogramme de colonnes
 - ❑ Représenter la répartition d'une variable continue en la représentant avec des colonnes
- ❑ Line histogram : alternative de Column histogram en cas de plusieurs points
- ❑ Scatter chart : distribution entre deux variables
- ❑ 3D Area chart : distribution entre trois variables



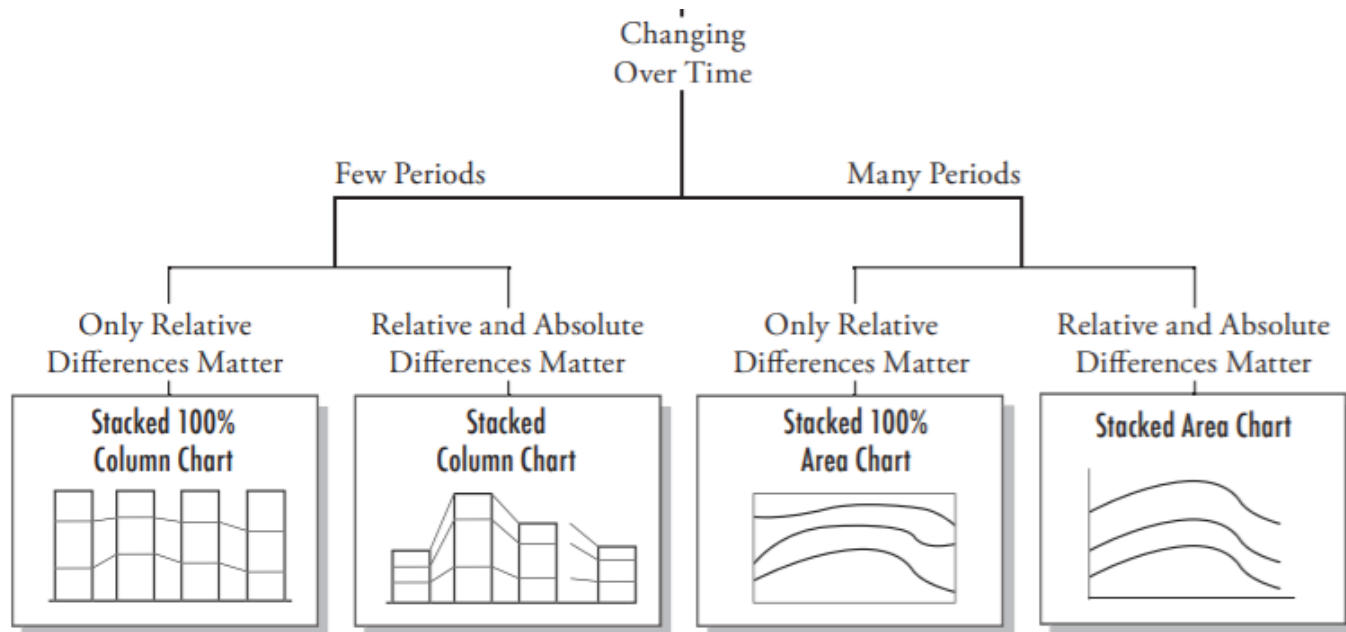
Comparaison : sur les items



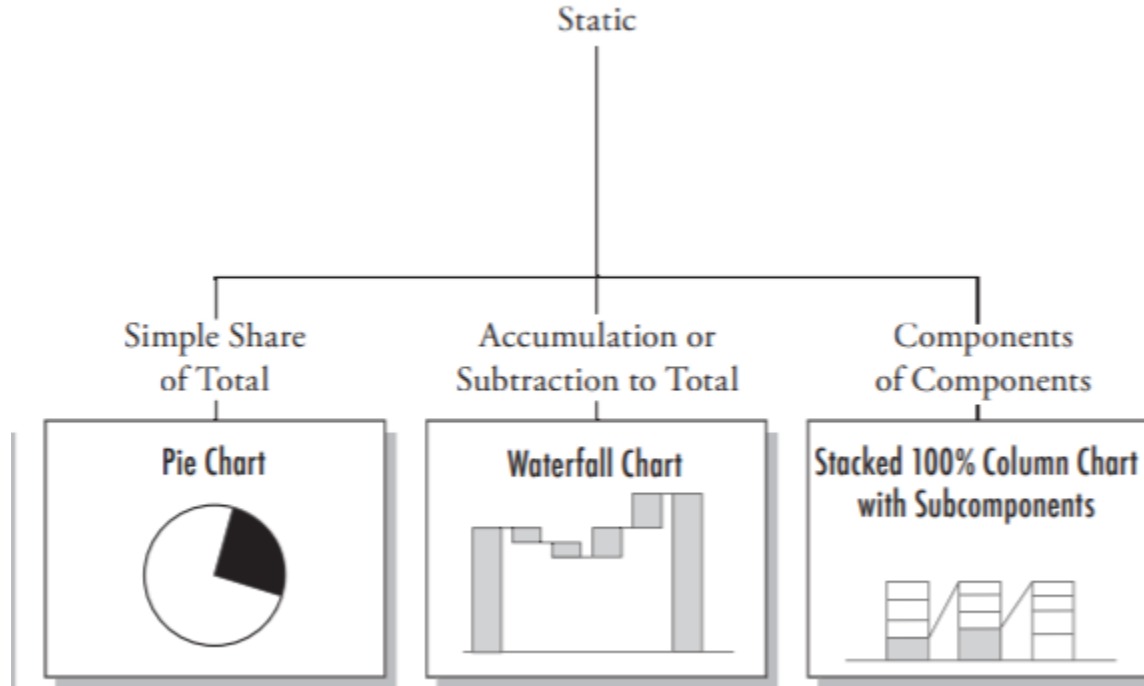
Comparaison : sur le temps



Composition : changement par rapport au temps



Composition : static



Les règles à respecter

- ❑ La lisibilité
- ❑ L'échelle
- ❑ Les proportions
- ❑ Les couleurs



Les librairies



Questions ?

Merci

