



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

EDAH Kodjo  
Consultant Systèmes d'Information, Big-Data



# Objectifs

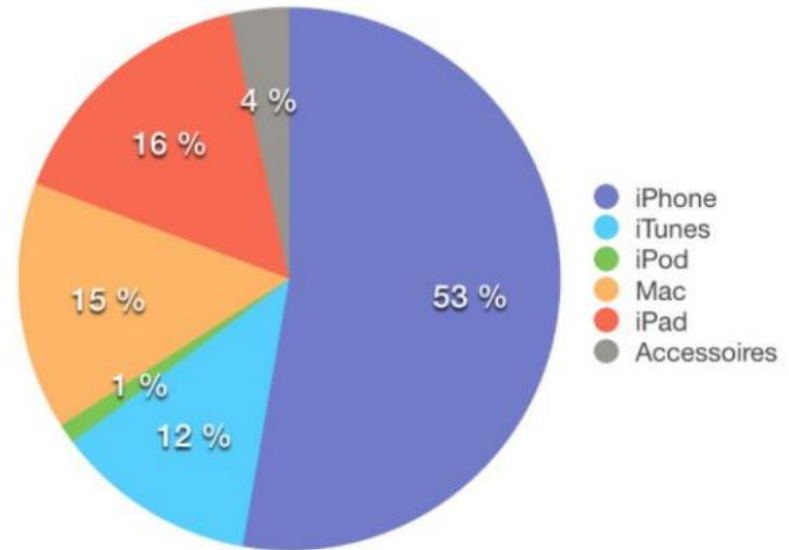
- Comprendre la notion et les spécificités du Big Data
- Connaître les outils de collecte, de traitement et d'exploitation des données
- Savoir utiliser les outils de visualisation des données (Dataviz)
- Piloter et maîtriser les risques des projets



## 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





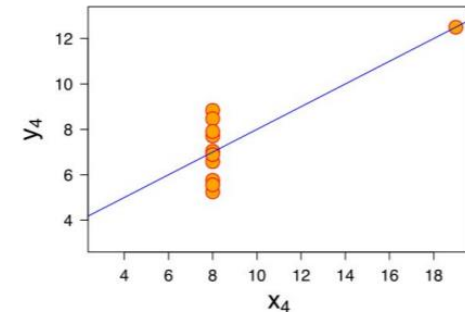
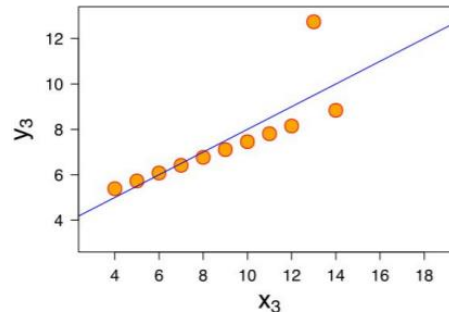
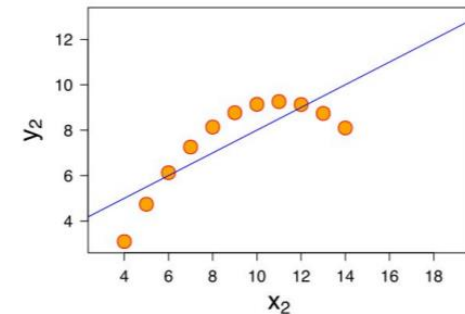
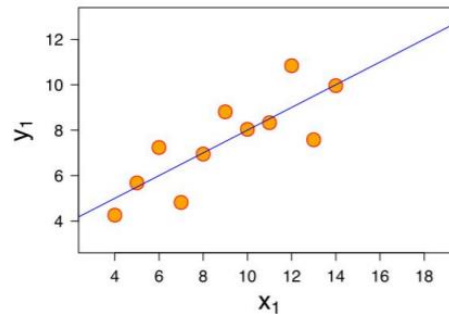
# Visualisation des données : c'est quoi la dataviz

—

**Pourquoi ne pas  
s'appuyer sur  
l'analyse de  
données ?**

Le Quartet d'Anscombe

[https://en.wikipedia.org/wiki/Anscombe%27s\\_quartet](https://en.wikipedia.org/wiki/Anscombe%27s_quartet)



# 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](http://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





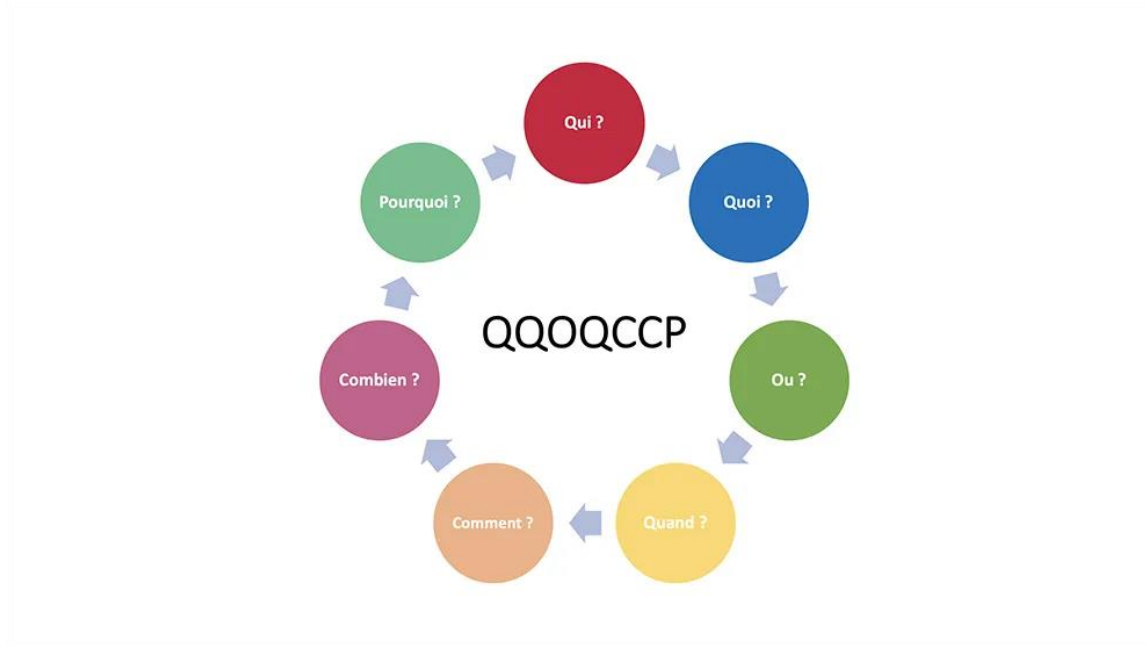
# Visualisation des données : dataviz

❑ Quel pays est le plus visité ?



Source : toucantoco.com

# Pourquoi la dataviz



<https://www.logiciels.pro/wp-content/uploads/2022/01/methode-qqoqccp-definition-exemples.webp>

# Pourquoi la dataviz

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

La data viz permet de

Répondre à une question

Prendre des décisions

Appréhender le contexte des données

Analyser et découvrir des patterns

Présenter un argument

Raconter une histoire (Storytelling)

Inspirer



# Objectif dataviz



Communication  
facile



Meilleure  
mémorisation



Ludique





# Les questions à se poser ?

- ❑ Quelle est l'information à communiquer ?
  - ❑ Mettre en valeur une information en particulier ?
  - ❑ Nécessité d'utiliser plusieurs graphiques ?



# Les questions à se poser ?

☐ Quels formats utiliser ?

☐ Les données numériques sont importantes ?

☐ Montrer une tendance ?



# Les questions à se poser ?

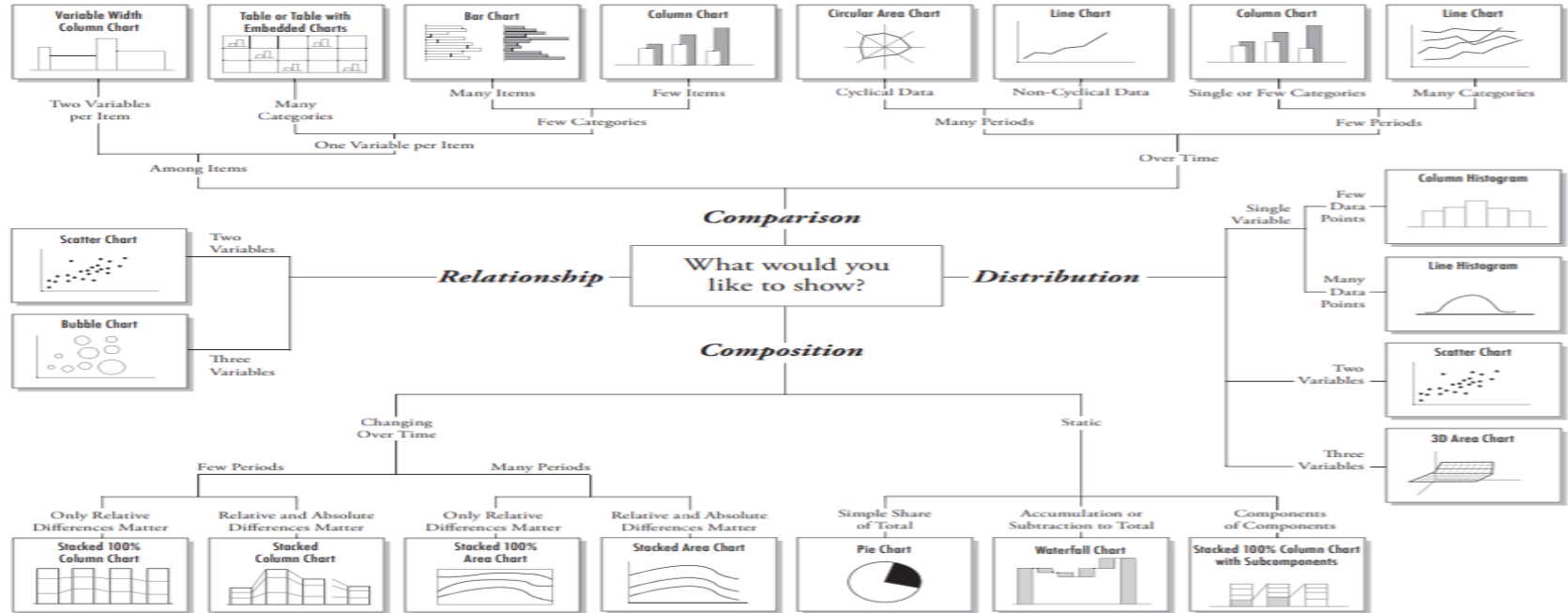
## ☐ Conception du design

- ☐ Axes
- ☐ Légendes
- ☐ Couleurs
- ☐ Grilles



# Quel type de présentation utilisé ?

## Chart Suggestions—A Thought-Starter

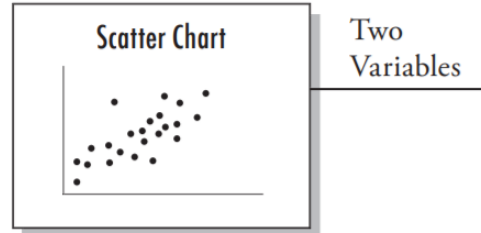




# Relation

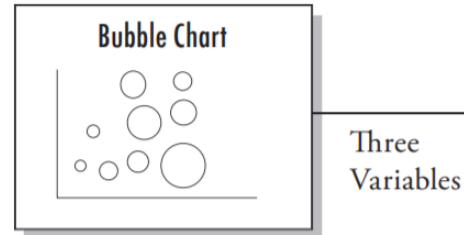
## ❑ Scatter chart : nuage de point

- ❑ 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.



## ❑ Bubble chart

- ❑ Graphique à bulles
- ❑ Affichage de 3 dimensions ( v1, v2, v3) , Deux v<sub>i</sub> sur les axes x et y, la troisième dimension la taille de la bulle

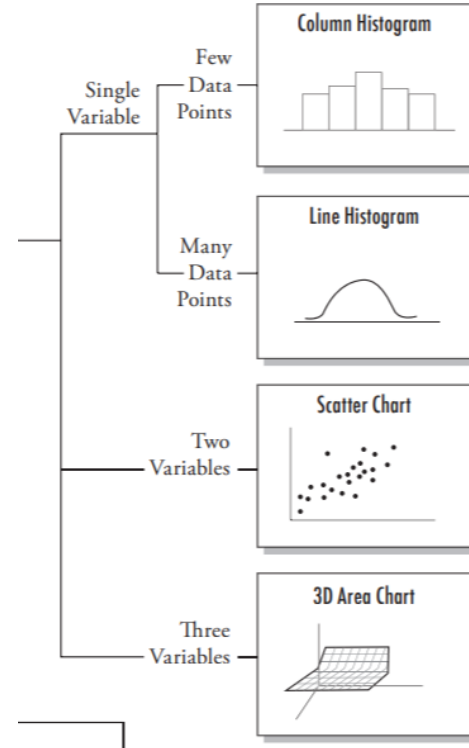


*Relationship*

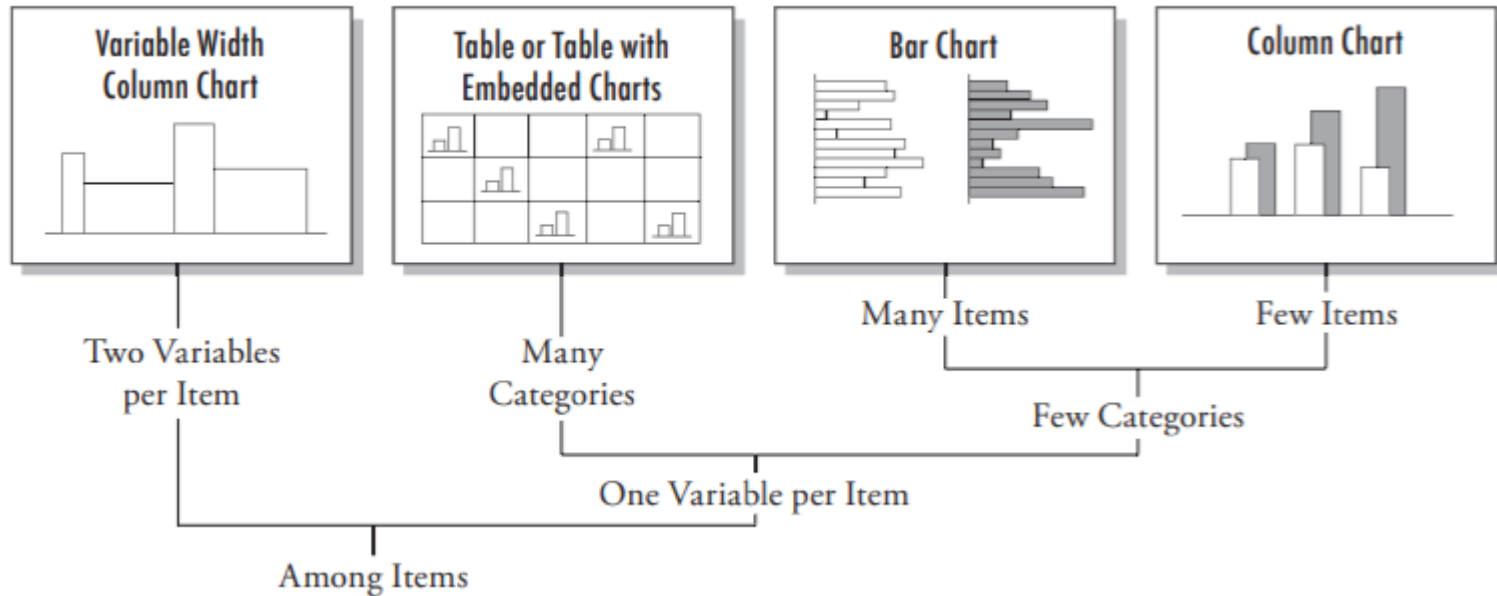


# Distribution

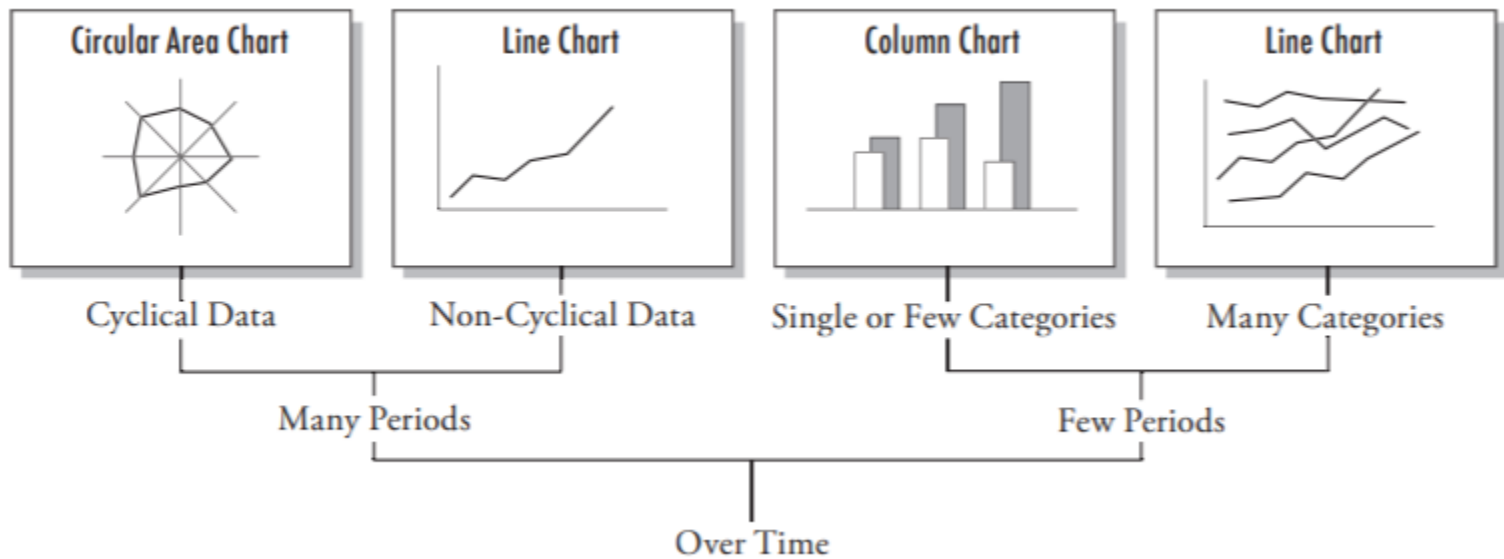
- ❑ Column histogram : histogramme de colonnes
  - ❑ Représenter la répartition d'une variable continue en la représentant avec des colonnes
- ❑ Line histogram : alternative de Column histogram en cas de plusieurs points
- ❑ Scatter chart : distribution entre deux variables
- ❑ 3D Area chart : distribution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



# Comparaison : sur les items

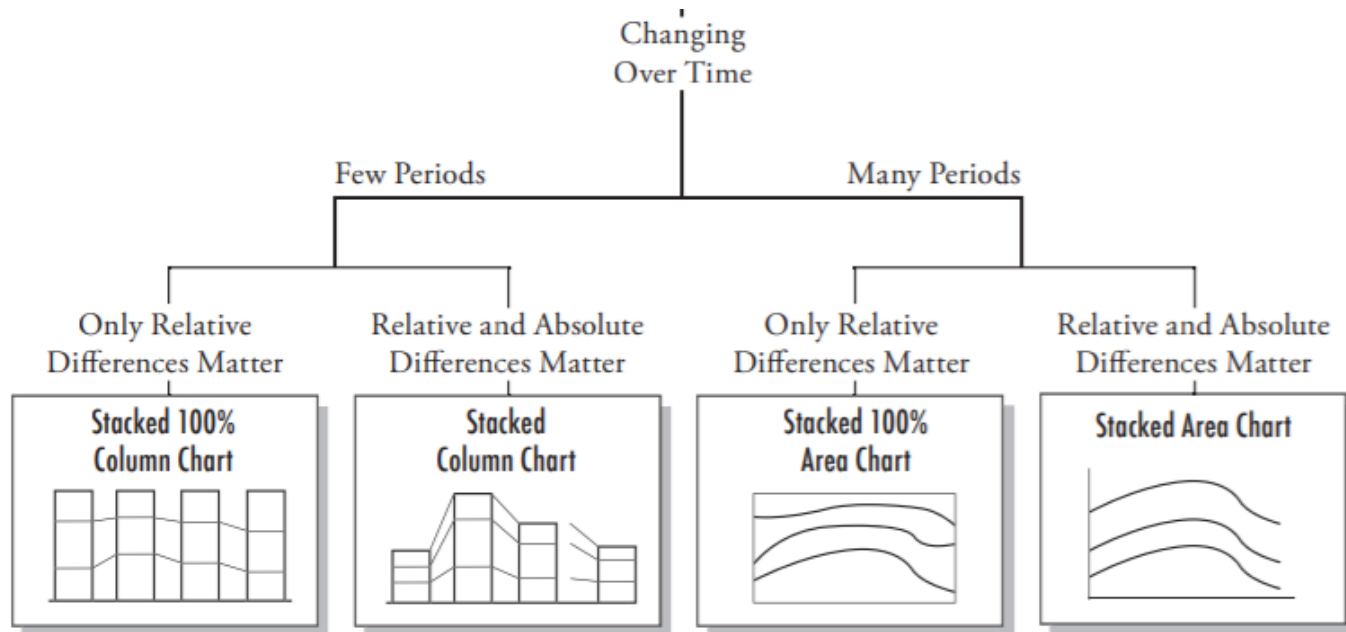


# Comparaison : sur le temps

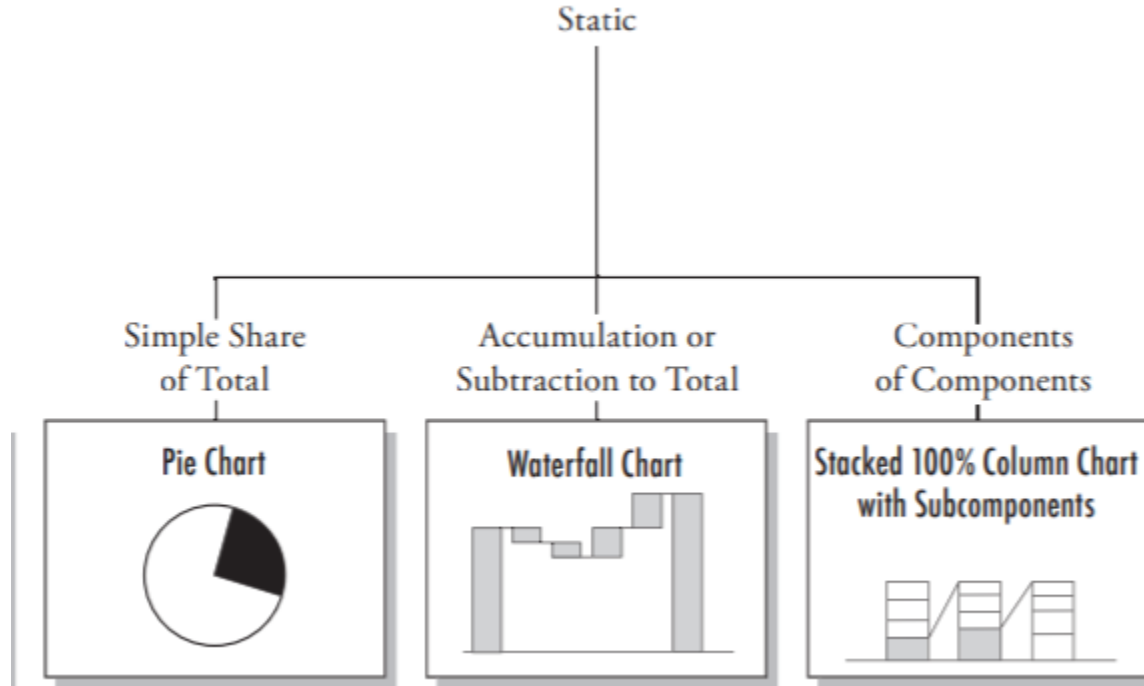




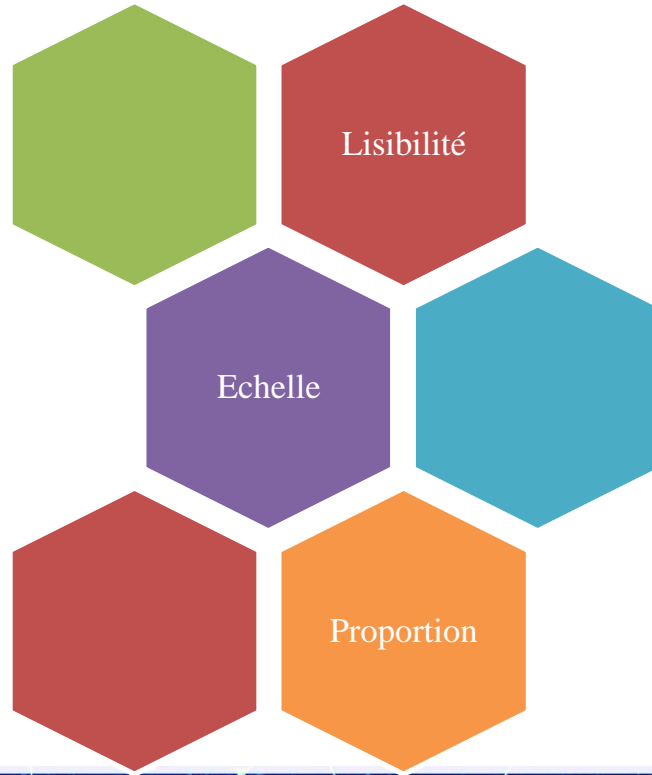
# Composition : changement par rapport au temps



# Composition : static



# Les règles à respecter



# Les librairies





# TP : ELK

❑ Elasticsearch est un moteur de recherche et d'analyse RESTful distribué

- ❑ centralise le stockage de vos données
- ❑ assure une recherche ultra-rapide
- ❑ scalables



- ❑ Logstash est un pipeline côté serveur destiné au traitement des données
  - ❑ Ingérer des données provenant d'une multitude de sources
  - ❑ Transformer
  - ❑ Envoyer vers votre système de stockage préféré (Elasticsearch .....



**Logstash**



- ❑ Kibana est une interface utilisateur qui vous permet de visualiser vos données Elasticsearch et de naviguer dans la Suite Elastic



☐ Récupérer des données et les traiter



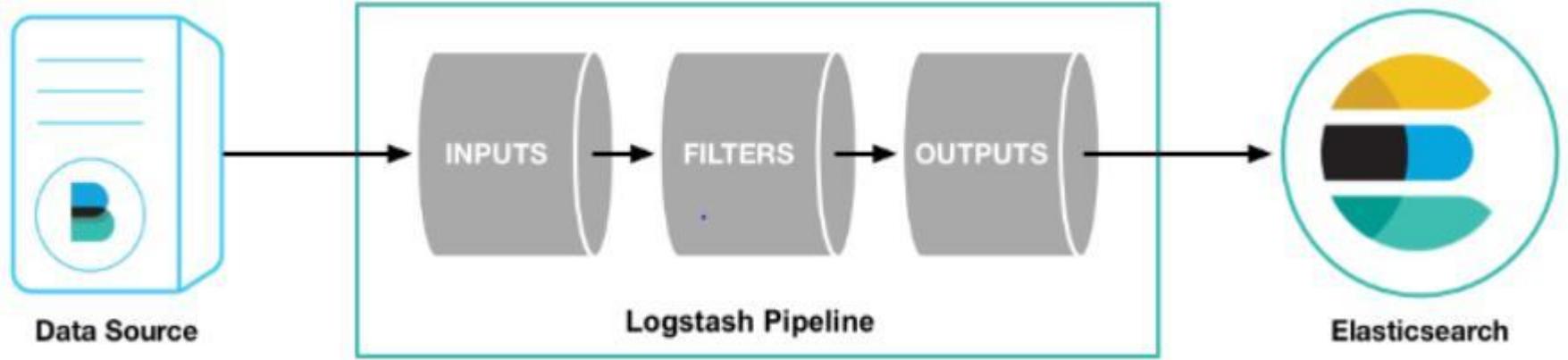
☐ Les stocker dans un moteur de recherche



☐ Les afficher et les analyser sous forme de graphe



# ELK



Source : <https://b3m5i7j6.rocketcdn.me/wp-content/uploads/article-elastic-invivoo-blog-2.jpg>





# ELK : Cas d'utilisation

- ☐ Traitement de fichiers de logs applicatifs en temps réel
- ☐ Utilisation dans un processus d'ETL pour de l'informatique décisionnelle
- ☐ Supervision d'un parc informatique
- ☐ Configuration Management DataBase : Base de données regroupant les machines d'un parc informatique
- ☐ IOT Affichage des données issues des objets connectés



# Logstash : Extract

## Configuration des Entrées

```
1 # Input à partir d'un fichier
2 input {
3   file {
4     path => ...
5   }
6 }
```

Entrées	Filtres	Output
file	date	elasticsearch
jdbc	csv	csv
log4j	geoip	email
rss	grok	nagios
sqlite	json	irc
syslog	ruby	stdout
tcp/udp	xml	mongodb
redis	aggregate	file



# Logstash : Extract

## Configuration des Entrées

```
1 # Input à partir d'une base de données MySQL
2 input {
3   jdbc {
4     jdbc_driver_library => "/path/to/mysql-connector-java-5.1.33-bin.jar"
5     jdbc_driver_class => "com.mysql.jdbc.Driver"
6     jdbc_connection_string => "jdbc:mysql://host:port/database"
7     jdbc_user => "user"
8     jdbc_password => "password"
9     statement => "SELECT ..."
10    jdbc_paging_enabled => "true"
11    jdbc_page_size => "50000"
12  }
13 }
```



# Logstash : Transform

## Configuration des Filtres

```
1 filter {  
2   # Lecture d'un fichier csv  
3   csv {  
4     columns => ["col1","col2","col3"]  
5     separator => ";"  
6   }  
7 }
```

```
1 filter {  
2   mutate {  
3     gsub => [  
4       # Remplace tous les slashes par des underscore  
5       "fieldname", "/", "_",  
6       # Remplace les backslashes, point d'interrogation, dièse, et tiret  
7       # par un point "."  
8       "fieldname2", "[\\?#-]", "."  
9     ]  
10  }  
11 }
```



# Logstash : Load

## Configuration des Sorties

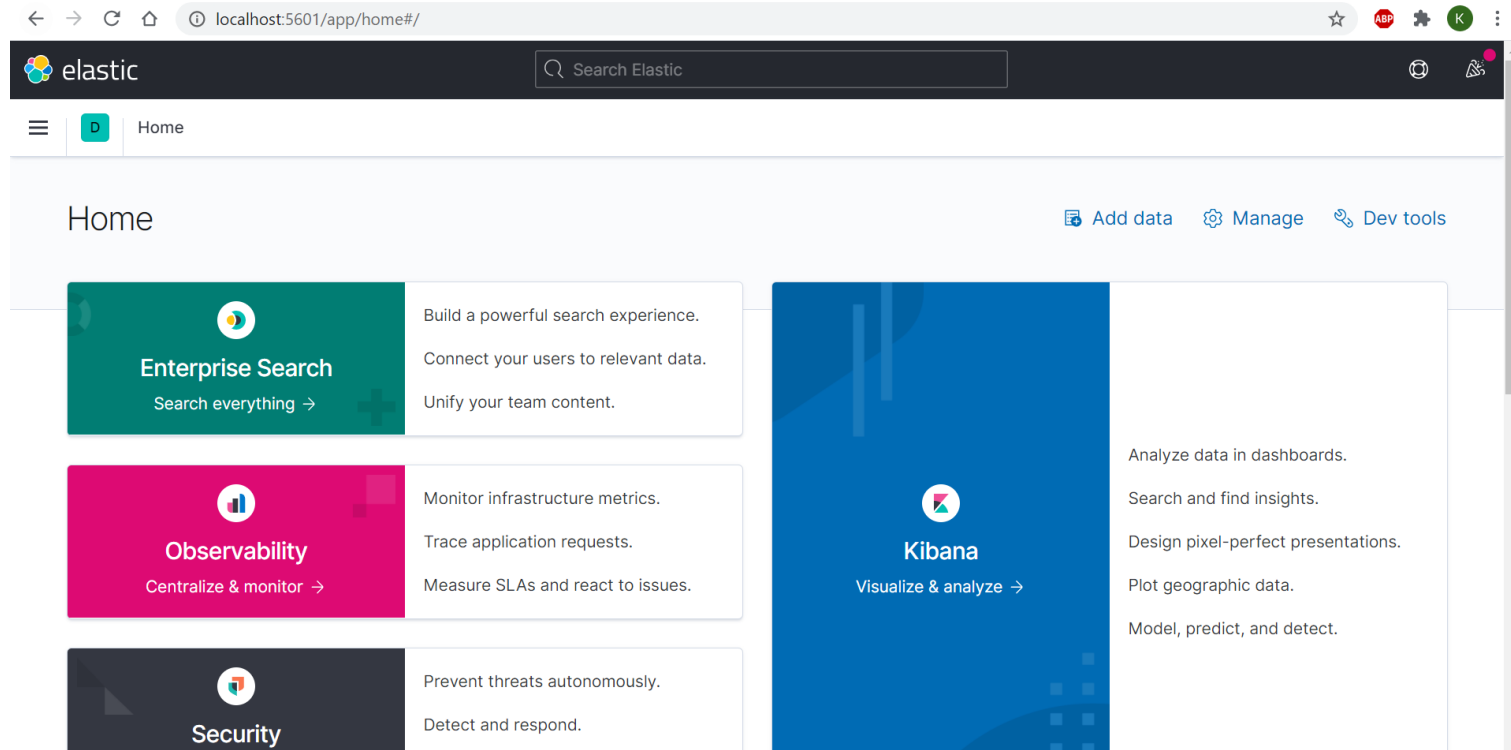
```
1 # Sortie vers la console
2 output {
3   stdout { codec => rubydebug }
4 }
```

```
1 # Sortie vers une base elasticsearch
2 output {
3   elasticsearch {
4     host => "elasticsearch_host"
5     cluster => "elastic_cluster"
6     port => "9300"
7     index => "mon-index-%{+YYYY.MM.dd}"
8     protocol => "transport"
9   }
10 }
```





# Kibana : Visualiser



The screenshot shows the Kibana web interface in a browser window. The address bar displays 'localhost:5601/app/home#/'. The page header features the 'elastic' logo, a search bar labeled 'Search Elastic', and user profile icons. A navigation bar below the header includes a menu icon, a 'D' icon, and the text 'Home'. The main content area is titled 'Home' and contains three primary action links: 'Add data', 'Manage', and 'Dev tools'. Below these links are three large, colorful cards representing different Kibana features: 'Enterprise Search' (green), 'Observability' (pink), and 'Security' (dark grey). Each card includes an icon, a title, a brief description, and a 'Visualize & analyze' link. To the right of these cards is a large blue card for 'Kibana' with a list of capabilities: 'Analyze data in dashboards.', 'Search and find insights.', 'Design pixel-perfect presentations.', 'Plot geographic data.', and 'Model, predict, and detect.'

elastic Search Elastic

Home

Add data Manage Dev tools

**Enterprise Search**  
Search everything →  
Build a powerful search experience.  
Connect your users to relevant data.  
Unify your team content.

**Observability**  
Centralize & monitor →  
Monitor infrastructure metrics.  
Trace application requests.  
Measure SLAs and react to issues.

**Security**  
Prevent threats autonomously.  
Detect and respond.

**Kibana**  
Visualize & analyze →  
Analyze data in dashboards.  
Search and find insights.  
Design pixel-perfect presentations.  
Plot geographic data.  
Model, predict, and detect.

Questions ?

---

Merci

