Festival Online de la Data

Synaltic

PROCHAINES DATES

Jeudi 04 Mars de 15h à 15h30
De Tableau Prep à la visualisation sur Tableau
Aya MHADHBI, Data Analyst, Synaltic

Jeudi 11 Mars de 15h à 15h30
Data-Asso
Marc SALLIERES, SYNALTIC

Jeudi 18 Mars de 15h à 15h30
ELASTICSEARCH - KIBANA SANS EFFORT
Galla TOPALIAN, Cheffe de projet Analytics, Synaltic
Alexandre NASRY, Solution Architect DevOps, Synaltic

Jeudi 25 Mars de 15h à 15h30
La Data Gouvernance par la pratique
Charly CLAIRMONT, CTO, Synaltic

SYNALTIC EN QUELQUES MOTS

Acteur innovant, dénicheur de solution et une équipe de collaborateurs engagés

Synaltic

30

15

160

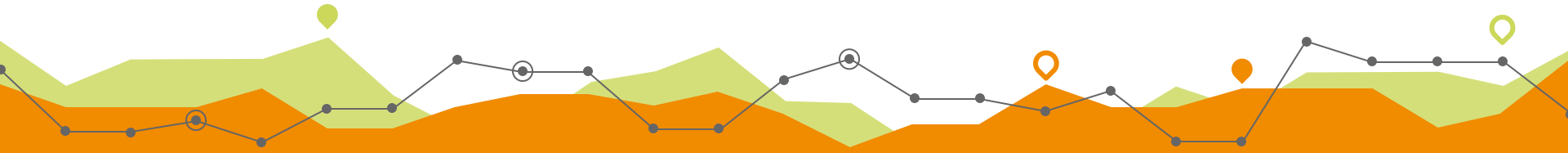
30 collaborateurs formés et certifiés contribuant aux communautés d'utilisateurs.

Spécialiste en Data Management depuis plus de 15 ans et plus de **250 projets** réalisés.

Plus de **160 clients** dont certains s'appuient sur nous avec succès depuis plus de **8 ans**.

Experts en Data Management, passionnés d'Open Source et d'Innovation !

TABLEAU + R



De la distribution statistique aux méthodes factorielles.

Février 2021

John BONTIT

John BONTIT

Data Analyst

Sommaire

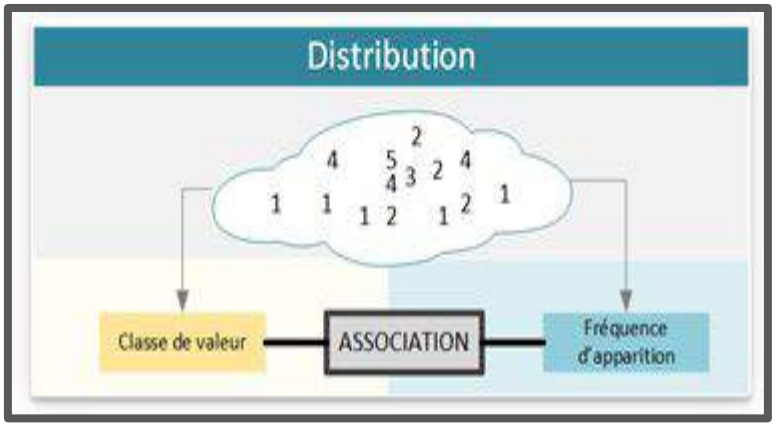
Introduction

Distribution

Statistique

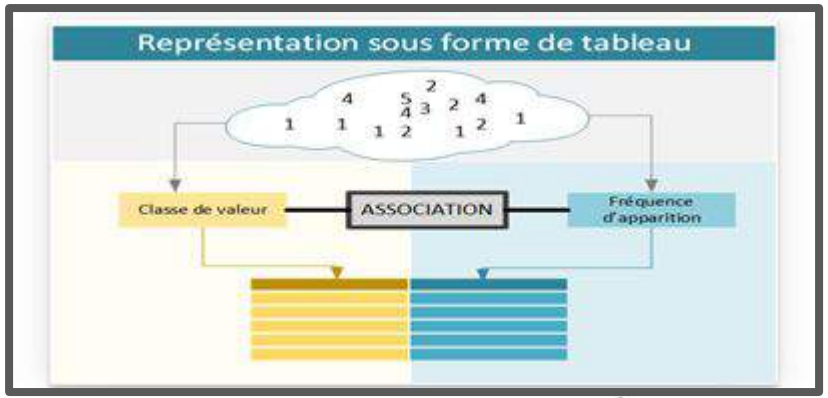
Méthodes factorielles

Introduction : une distribution



La distribution est une fonction qui associe une **fréquence d'apparition** à une **classe de valeur.**

Cette fonction permet de représenter l'information contenue dans un **ensemble de données**.



Introduction : la statistique

La statistique c'est l'étude des phénomènes relatifs à :

- La collecte des données et leur traitement
- Leur analyse et l'interprétation des résultats
- Leur restitution de manière compréhensible

Elle est donc une science, une **méthode** et un **ensemble de techniques**

Introduction : la statistique

Stats

VS

Dataviz

Classe 1

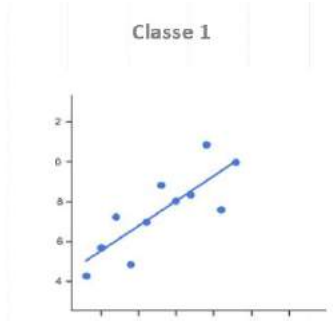
ID	Notes
A	2
B	6
C	3
D	5
AVG	4
VAR	2,5

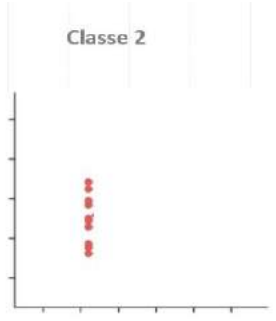
Classe 2

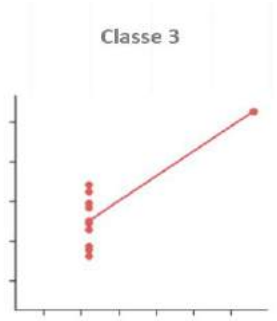
ID	Notes
A	4
B	4
C	4
D	4
AVG	4
VAR	0

Classe 3

ID	Notes
A	0
B	0
C	0
D	16
AVG	4
VAR	48

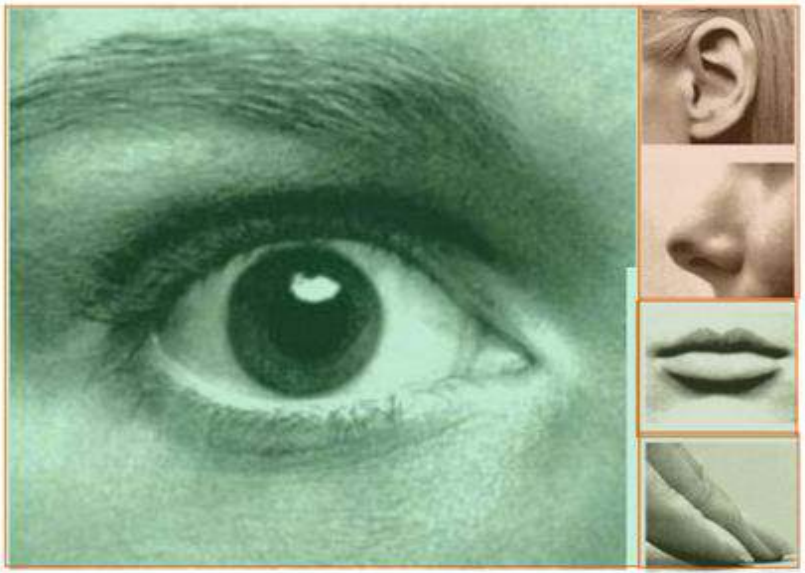






Introduction : les méthodes factorielles

Be Visual



70%

30%

Introduction : les méthodes factorielles

Mélange de statistique, de mathématiques appliquées et d'informatique

- Elles permettent d'utiliser les **facultés de perception** sur les graphiques.

- Ces représentations graphiques sont aussi un **moyen de communication** remarquable.

- Le résultat d'une analyse factorielle est **unique**, ce qui en assure la rigueur.

Méthodes univariées : statistiques à une variable

Collecte & traitement

Analyse et interprétation

Restitution

Méthodes univariées : collecte & traitement

Etudiant	Français
Charles	9
Paul	15
Eric	6
Abdou	14
François	7
Jo	7
Alex	10
Vanessa	16
Anna	12
Aya	11

Individus : 10 Synalticiens

Variable(s) : Note de Français / 20

Méthodes univariées : analyse & interprétation

Etudiant	Français
Charles	9
Paul	15
Eric	6
Abdou	14
François	7
Jo	7
Alex	10
Vanessa	16
Anna	12
Aya	11

- **Insuffisants :** Charles, Eric, François, Jo
- **Moyens :** Alex, Anna, Aya
- **Bons :** Paul, Abdou, Vanessa

Méthodes univariées : restitution & diffusion

- **Insuffisants :** Charles, Eric, François, Jo
- **Moyens :** Alex, Anna, Aya
- **Bon :** Paul, Abdou, Vanessa

Méthodes bivariées : statistique à 2 variables

Collecte & traitement

Analyse et interprétation

Dataviz et diffusion

Méthodes bivariées : collecte & traitement

Etudiant	Français	Maths	Physique
Charles	9	16	15
Paul	15	8	10
Eric	6	9	10
Abdou	14	17	16
Francois	7	13	12
Jo	7	14	15
Alex	10	8	8
Vanessa	16	15	16
Anna	12	10	9
Aya	11	11	11

Individus : 10 Synalticiens

Variables :

- **Note de Français / 20**
- **Note de Maths / 20**
- **Note de Physique / 20**

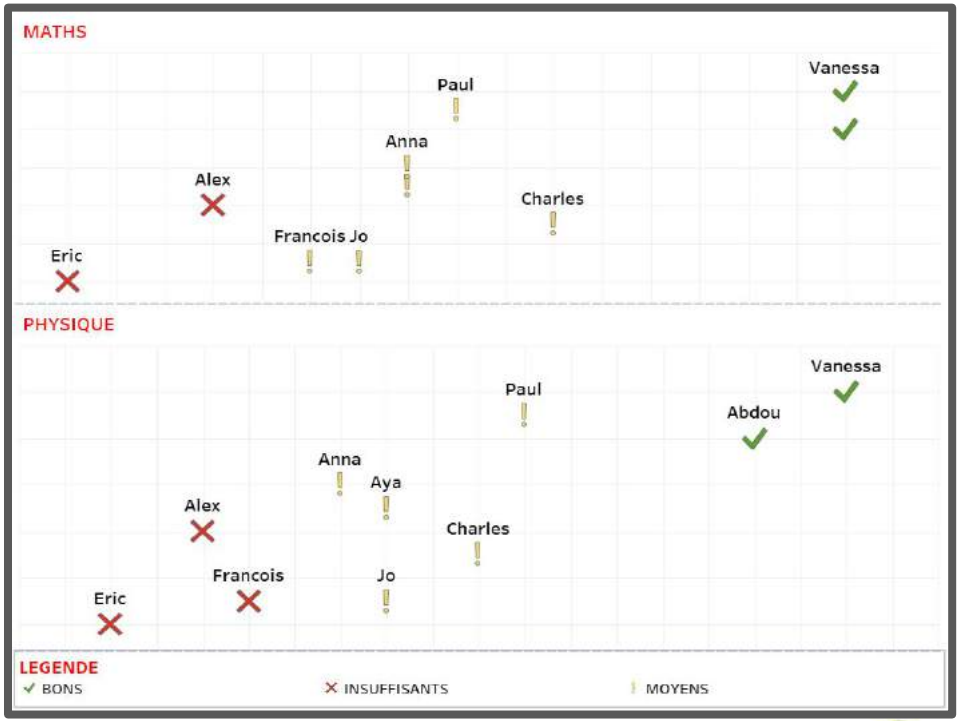
Méthodes bivariées : analyse & interprétation

Etudiant	Français	Maths	Physique
Charles	9	16	15
Paul	15	8	10
Eric	6	9	10
Abdou	14	17	16
Francois	7	13	12
Jo	7	14	15
Alex	10	8	8
Vanessa	16	15	16
Anna	12	10	9
Aya	11	11	11

CORRELATION Francais - Maths	10,63%
CORRELATION Francais - Physique	18,18%
CORRELATION Maths - Physique	94,51%

Méthodes bivariées : restitution & diffusion

CORRELATION Francais - Maths	10,63%
CORRELATION Francais - Physique	18,18%
CORRELATION Maths - Physique	94,51%



Méthodes multivariées : statistique à n variables

Problématique

Typologie des méthodes factorielles

Démonstration

Méthodes multivariées : problématique

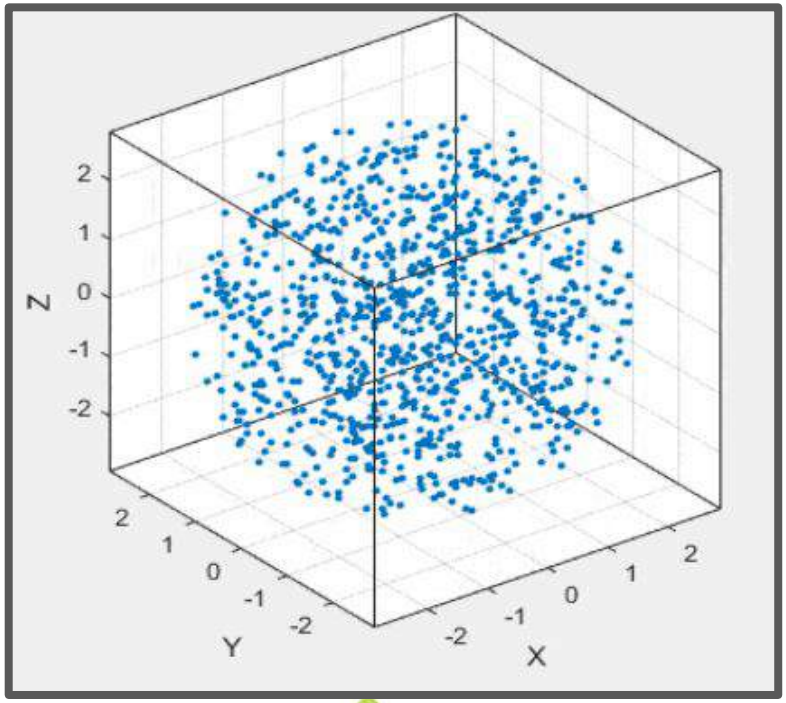
A

Etudiant	Français	Maths	Physique
Charles	9	16	15
Paul	15	8	10
Eric	6	9	10
Abdou	14	17	16
Francois	7	13	12
Jo	7	14	15
Alex	10	8	8
Vanessa	16	15	16
Anna	12	10	9
Aya	11	11	11

VS

B

Méthodes multivariées : problématique

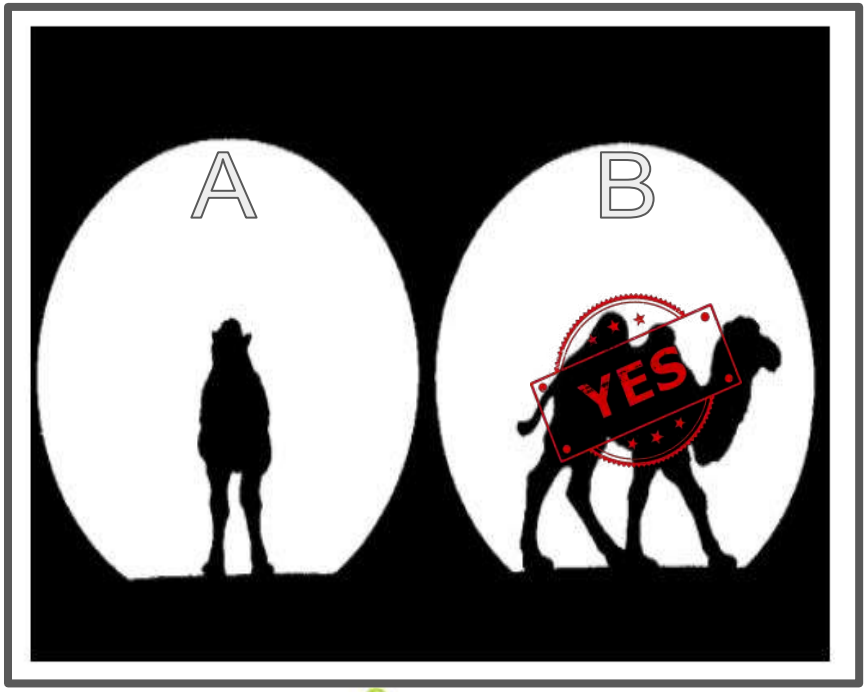


Représentation en 3D

- Pas grand intérêt à réaliser un nuage en 3D
- Analyse et interprétation pas aisé
- Restitution et diffusion

Que dire de la représentation lorsque n>3 ?

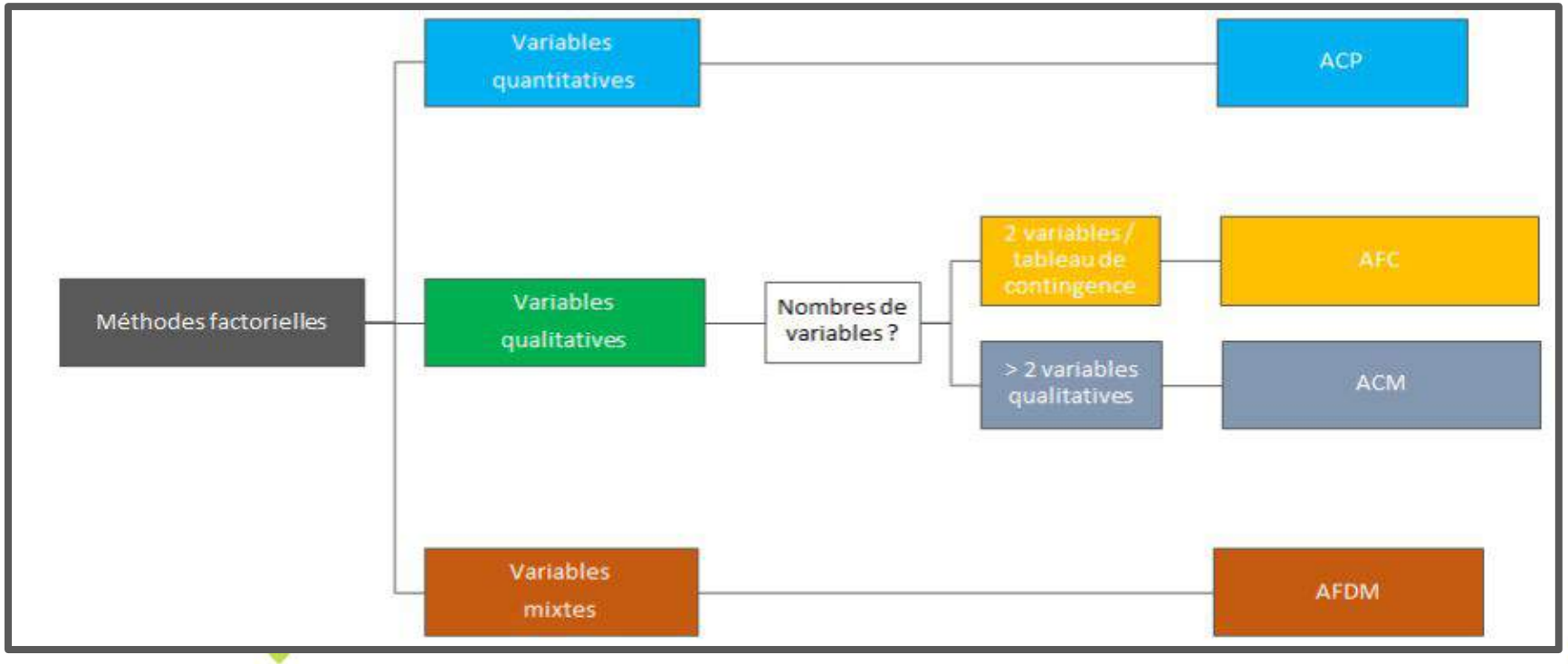
Méthodes multivariées : problématique



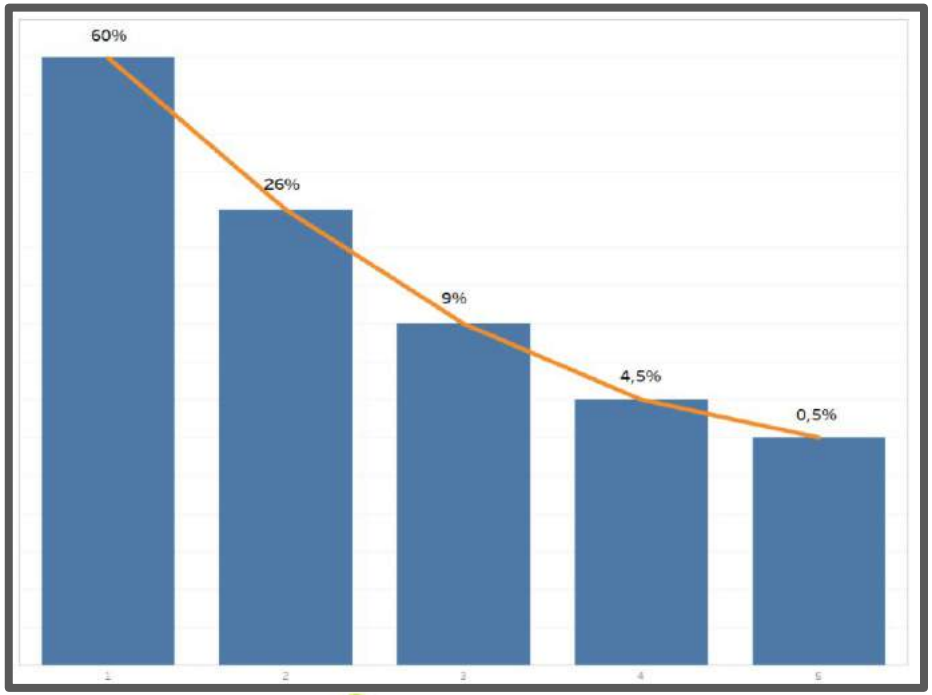
Be visual !!!

- **Image >3D :** il est quasi-impossible d'identifier la nature de l'animal sur la figure.
- **Image en 2D :** il est facile de visualiser mentalement et d'identifier l'animal.

Méthodes multivariées : typologie



Méthodes multivariées : typologie



Taux d'information

1ere CP = 60%

2e CP = 26%

3e CP = 9%

4e CP = 4,5%

5e CP = 0,5%

Méthodes multivariées : typologie

Marque	prix (€)	cylindrée	puissance	longueur	largeur	poids	vitesse	finition
A	15280	1350	79	393	161	870	165	B
B	20000	1588	85	468	177	1110	160	TB
C	14800	1294	68	424	168	1050	152	M
D	14100	1222	59	412	161	930	151	M
E	17450	1585	98	439	164	1105	165	B
F	17740	1297	82	429	169	1080	160	TB
G	16150	1796	79	449	169	1160	154	B
H	16000	1565	55	424	163	1010	140	B
I	23800	2664	128	452	173	1320	180	TB
J	13270	1166	55	399	157	815	140	M
K	21200	1570	109	428	162	1060	175	TB
L	17000	1798	82	445	172	1160	158	B
M	22000	1998	115	469	169	1370	160	TB
N	17500	1993	98	438	170	1080	167	B
O	19700	1442	80	431	166	1129	144	TB
P	14000	1769	83	440	165	1095	165	M
Q	16350	1979	100	459	173	1120	173	B
R	11050	1294	68	404	161	955	140	M

DEMO AFC

Méthodes multivariées : demo AFC

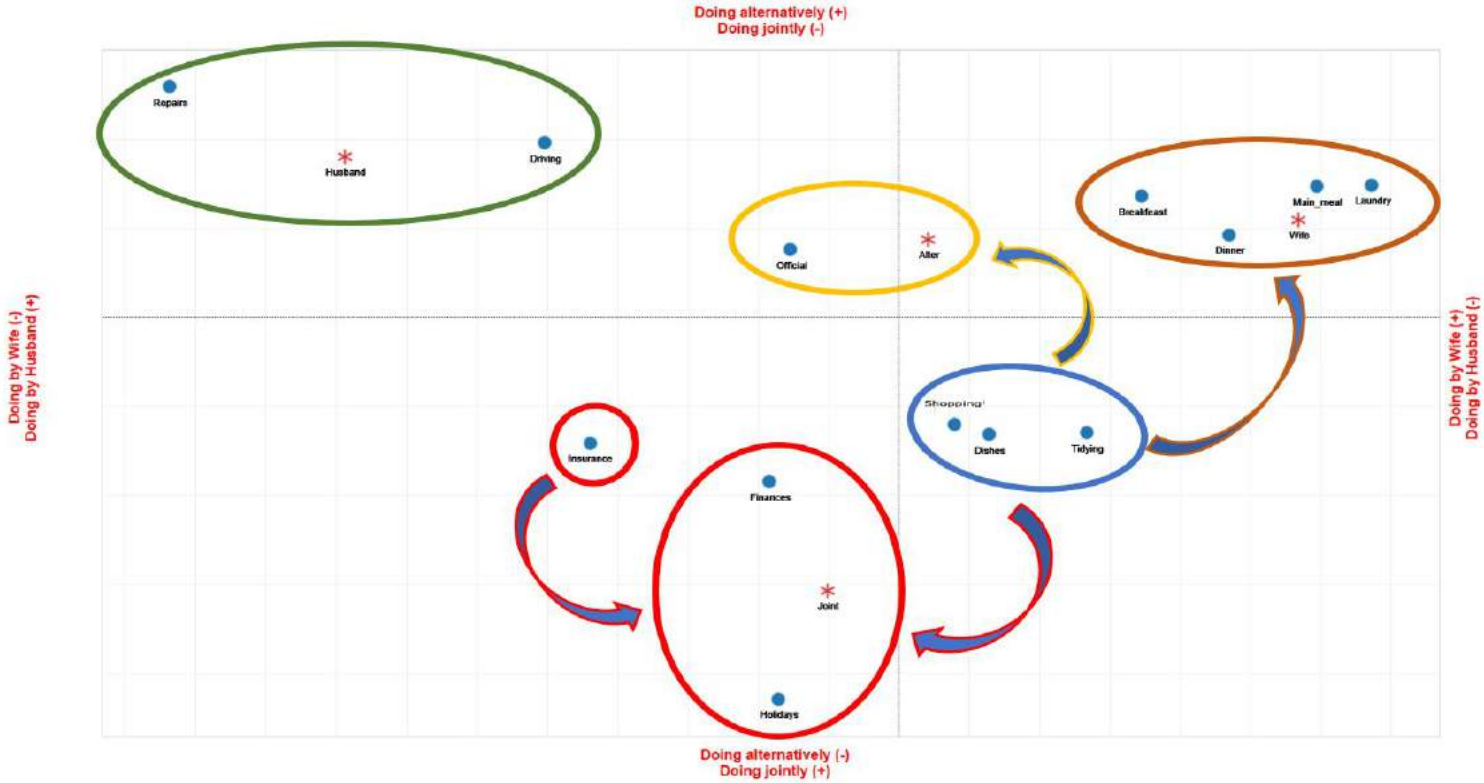
Synaltic

Tableau de Contingence

	Alterné	Ensemble	Femme	Homme
Assurance	1	77	8	53
Conduite	51	3	10	75
Dejeuner	20	4	124	5
Diner	11	13	77	7
Finances	13	66	13	21
Lessive	14	4	156	2
Official	46	15	12	23
Petit-Dejeuner	36	7	82	15
Rangement	11	57	53	1
Réparations	3	2	0	160
Shopping	23	55	33	9
Vacances	1	153	0	6
Vaisselle	24	53	32	4

https://www.fun-mooc.fr/asset-v1:grenoblealpes+92001+session01+type@asset+block/mod5-cap2.pdf

Méthodes multivariées : démo AFC



Méthodes multivariées : démo AFC

- **Simplicité :** l'analyse de correspondances permet d'appréhender une grande partie du résultat de l'analyse d'un simple coup d'œil.

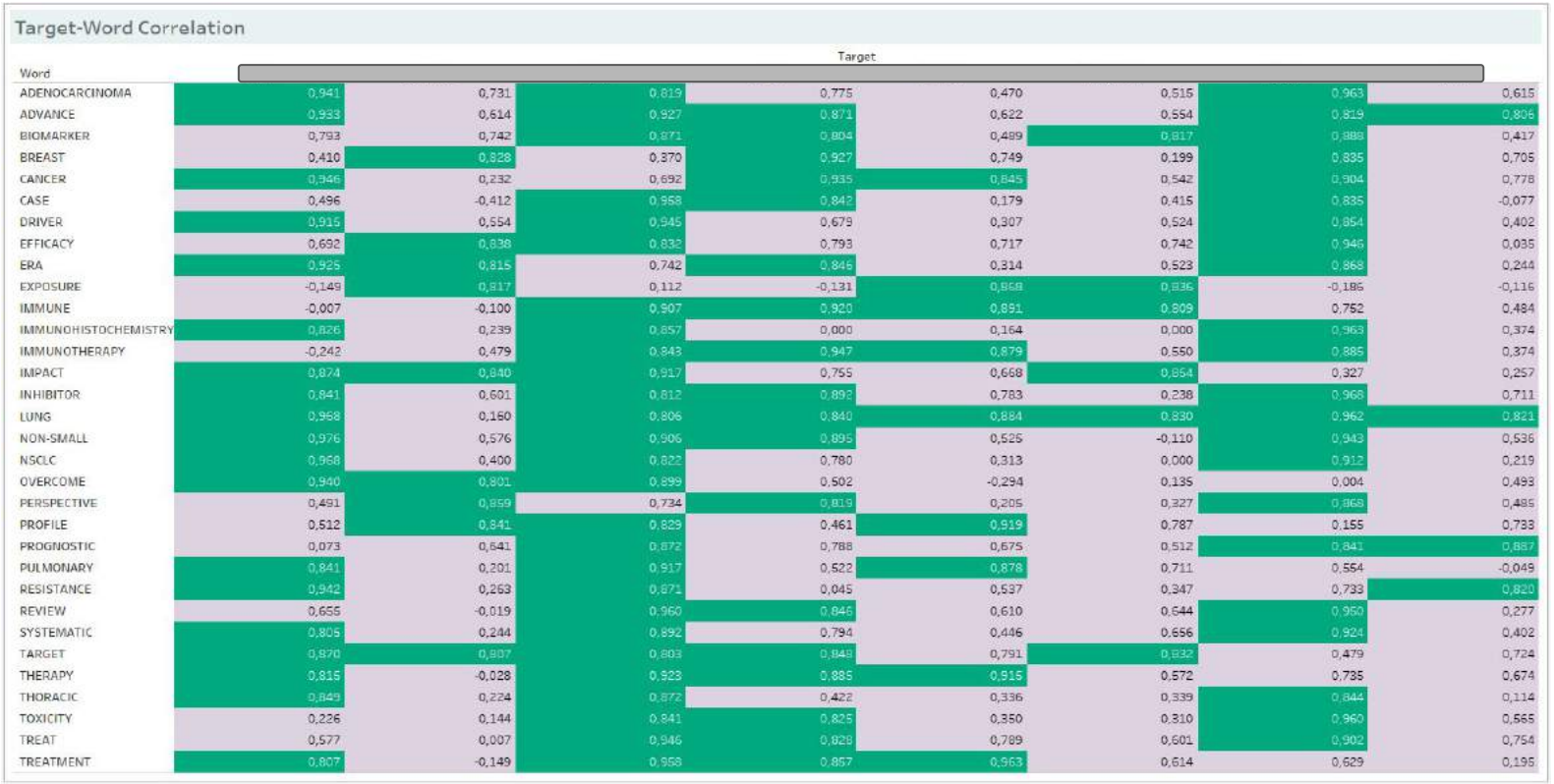
- **Puissance :** l'analyse de correspondances offre un résumé et une vue complète des relations existant dans une population étudiée.

- **Flexibilité :** cette technique s'appuie sur un ensemble de données de tout type et de taille quelconque. Cette souplesse traduit ainsi la diversité de ses applications.

Exemple de cas d'usage

Exemple de cas d'usage

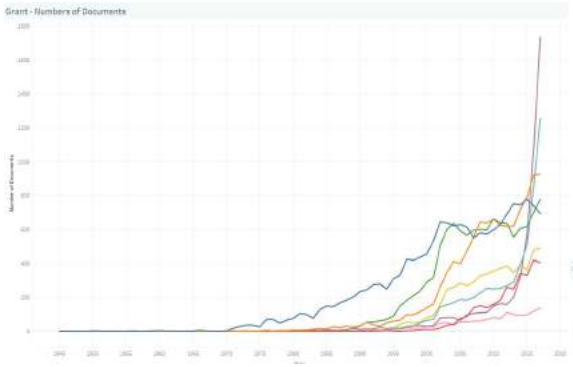
Synaltic

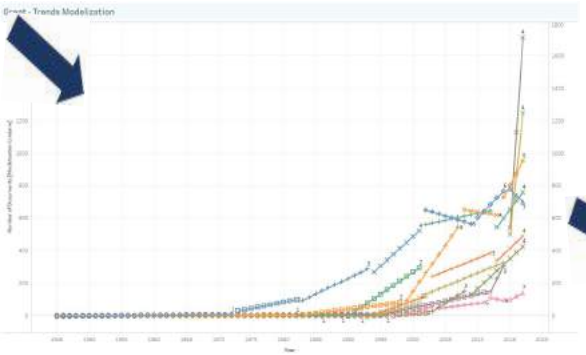


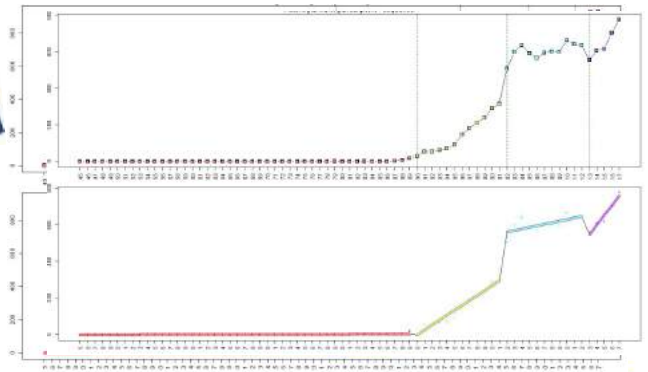
Target-Word Correlation

Word	Target							
ADENOCARCINOMA	0,941	0,731	0,819	0,775	0,470	0,515	0,963	0,615
ADVANCE	0,933	0,614	0,927	0,871	0,622	0,554	0,819	0,806
BIOMARKER	0,793	0,742	0,871	0,804	0,489	0,817	0,888	0,417
BREAST	0,410	0,828	0,870	0,927	0,749	0,199	0,835	0,705
CANCER	0,946	0,232	0,692	0,935	0,845	0,542	0,904	0,778
CASE	0,496	-0,412	0,958	0,842	0,179	0,415	0,835	-0,077
DRIVER	0,915	0,554	0,945	0,679	0,307	0,524	0,854	0,402
EFFICACY	0,692	0,838	0,832	0,793	0,717	0,742	0,946	0,035
ERA	0,925	0,815	0,742	0,846	0,314	0,523	0,868	0,244
EXPOSURE	-0,149	0,817	0,112	-0,131	0,868	0,836	-0,186	-0,116
IMMUNE	-0,007	-0,100	0,907	0,920	0,891	0,809	0,752	0,484
IMMUNOHISTOCHEMISTRY	0,826	0,239	0,857	0,000	0,164	0,000	0,963	0,374
IMMUNOTHERAPY	-0,242	0,479	0,843	0,947	0,879	0,550	0,885	0,374
IMPACT	0,874	0,840	0,917	0,755	0,668	0,854	0,327	0,257
INHIBITOR	0,841	0,601	0,812	0,892	0,783	0,238	0,968	0,711
LUNG	0,968	0,160	0,806	0,840	0,884	0,830	0,962	0,821
NON-SMALL	0,976	0,576	0,906	0,895	0,525	-0,110	0,943	0,536
NSCLC	0,968	0,400	0,822	0,780	0,313	0,000	0,912	0,219
OVERCOME	0,940	0,801	0,899	0,502	-0,294	0,135	0,004	0,493
PERSPECTIVE	0,491	0,859	0,734	0,819	0,205	0,327	0,868	0,485
PROFILE	0,512	0,841	0,829	0,461	0,919	0,787	0,155	0,733
PROGNOSTIC	0,073	0,641	0,872	0,788	0,675	0,512	0,841	0,887
PULMONARY	0,841	0,201	0,917	0,522	0,878	0,711	0,554	-0,049
RESISTANCE	0,942	0,263	0,871	0,045	0,537	0,347	0,733	0,820
REVIEW	0,655	-0,019	0,960	0,848	0,610	0,644	0,960	0,277
SYSTEMATIC	0,805	0,244	0,892	0,794	0,446	0,656	0,924	0,402
TARGET	0,870	0,807	0,803	0,848	0,791	0,832	0,479	0,724
THERAPY	0,815	-0,028	0,923	0,885	0,915	0,572	0,735	0,674
THORACIC	0,849	0,224	0,872	0,422	0,336	0,339	0,844	0,114
TOXICITY	0,226	0,144	0,841	0,825	0,350	0,310	0,960	0,565
TREAT	0,577	0,007	0,946	0,828	0,789	0,601	0,902	0,754
TREATMENT	0,807	-0,149	0,958	0,857	0,963	0,614	0,629	0,195

Exemple de cas d'usage







Merci !

Festival Online de la Data

Synaltic

PROCHAINES DATES

Jeudi 04 Mars de 15h à 15h30
De Tableau Prep à la visualisation sur Tableau
Aya MHADHBI, Data Analyst, Synaltic

Jeudi 11 Mars de 15h à 15h30
Data-Asso
Marc SALLIERES, SYNALTIC

Jeudi 18 Mars de 15h à 15h30
ELASTICSEARCH - KIBANA SANS EFFORT
Galla TOPALIAN, Cheffe de projet Analytics, Synaltic
Alexandre NASRY, Solution Architect DevOps, Synaltic

Jeudi 25 Mars de 15h à 15h30
La Data Gouvernance par la pratique
Charly CLAIRMONT, CTO, Synaltic