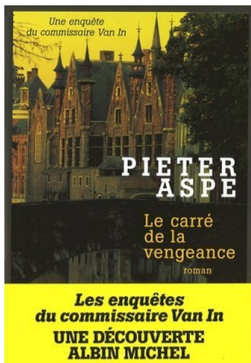


La statistique dans tous ses états



*Selon le profil que j'ai établi en me basant sur les deux fax, le plus vieux des ravisseurs est très certainement **universitaire**. Il doit être **ingénieur** ou **mathématicien**. Je ne serais pas étonné si on me disait que c'est un **statisticien** ou un spécialiste du **calcul des probabilités**.*

Source : *Le carré de la vengeance*, Pieter Aspe

Sébastien Déjean

www.math.univ-toulouse.fr/~sdejean

Ingénieur de recherche en calcul scientifique
Institut de Mathématiques de Toulouse - Université Toulouse III Paul Sabatier



Sylviane Gasquet-More

Plus vite que son nombre

DÉCHIFFRER L'INFORMATION



CE QUE VOUS CACHENT
LES CHIFFRES

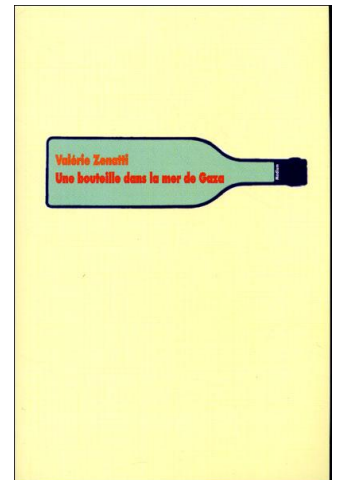
Seuil

Après nous avoir convaincu de leur objectivité fondamentale, il ne reste plus aux chiffres qu'à nous amener doucement à penser qu'ils en déterminent le monopole. Dès lors, une forme de hiérarchie gagne l'argumentation et le raisonnement : contenir quelques chiffres qualifie automatiquement votre discours, même si personne ne prend la peine de comprendre vraiment ce qu'ils signifient, voire même s'ils sont sans rapport avec le sujet traité ! A contrario, de ce fait, toute argumentation purement textuelle semble dépréciée [...] comme si le raisonnement et la rigueur ne pouvait exister hors des chiffres.

Sylviane Gasquet-More, *Plus vite que son nombre*

Mais les probabilités, les statistiques, c'est bon pour les maths, la biologie, ce sont des chiffres sur du papier.

Delphine Zenatti, *Une bouteille dans la mer de Gaza*



... face au déploiement de la société des calculs, il est nécessaire d'encourager la diffusion d'une culture statistique vers un public plus large que celui des seuls spécialistes.

Dominique Cardon, *A quoi rêvent les algorithmes. Nos vies à l'heure des big data.*

Dominique Cardon
À quoi rêvent
les algorithmes

Nos vies à l'heure des big data

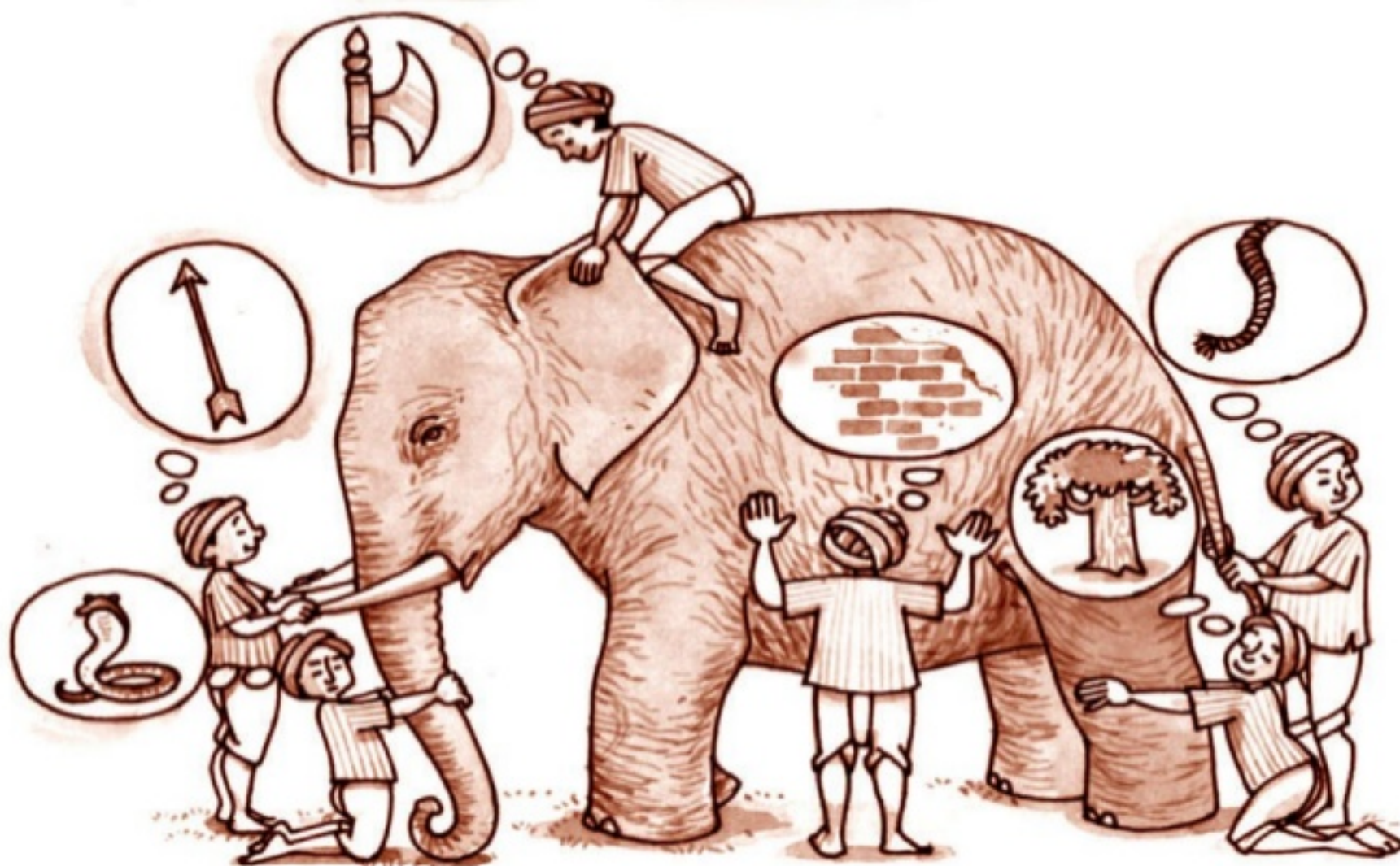
*T'échappes à la police,
pas aux statistiques.*
Jean-Jacques Goldman,
paroles de la chanson
Des vies

Statistique Descriptive

Résumer une série de chiffres par un ou plusieurs chiffres (mais pas beaucoup)

Six aveugles et un éléphant

en.wikipedia.org/wiki/Blind_men_and_an_elephant



Indicateurs statistiques

- Indicateurs pour 1 série de données
 - Position / Tendence centrale
 - Dispersion
- Indicateurs pour 2 séries de données

Indicateurs de tendance centrale

Moyenne : somme des observations divisées par le nombre d'observations

Médiane : Valeur qui sépare l'échantillon en 2 sous-ensembles de tailles égales.

Mode : Valeur la plus fréquente dans un échantillon.

Quartiles : **3** valeurs qui sépare l'échantillon en **4** sous-ensembles de tailles égales.

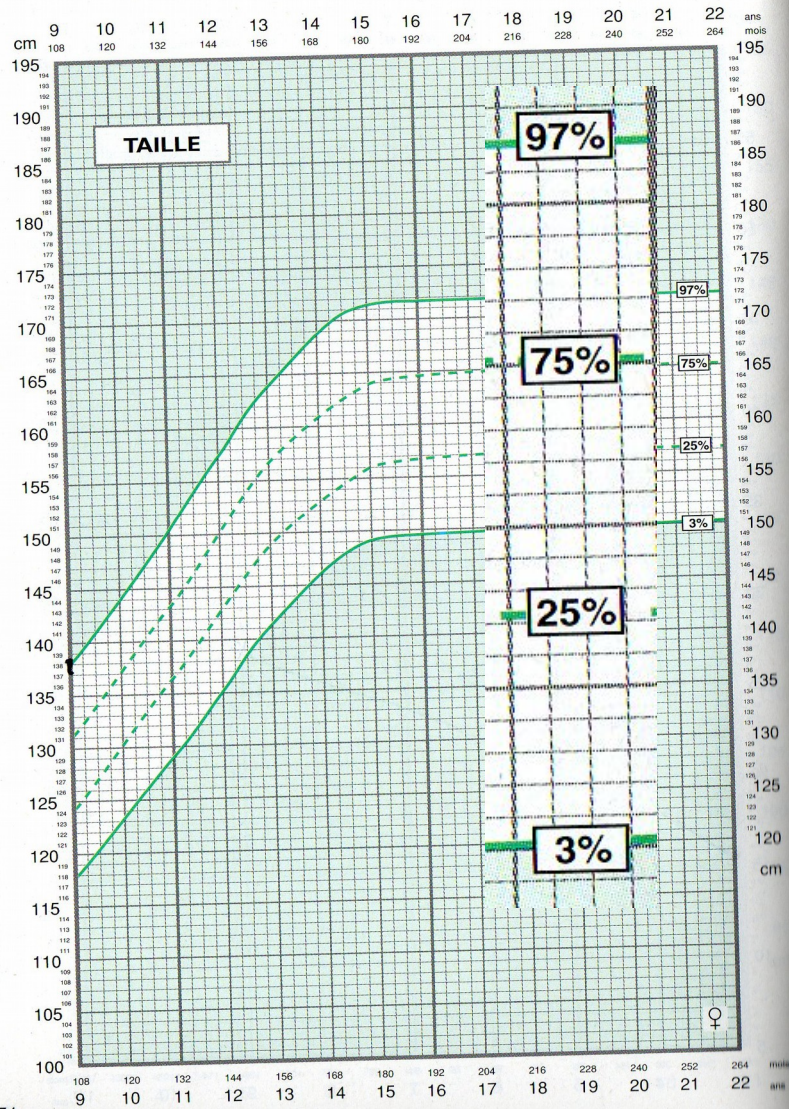
Déciles : **9** valeurs qui sépare l'échantillon en **10** sous-ensembles...

Percentiles : **99** valeurs qui sépare l'échantillon en **100** sous-ensembles...

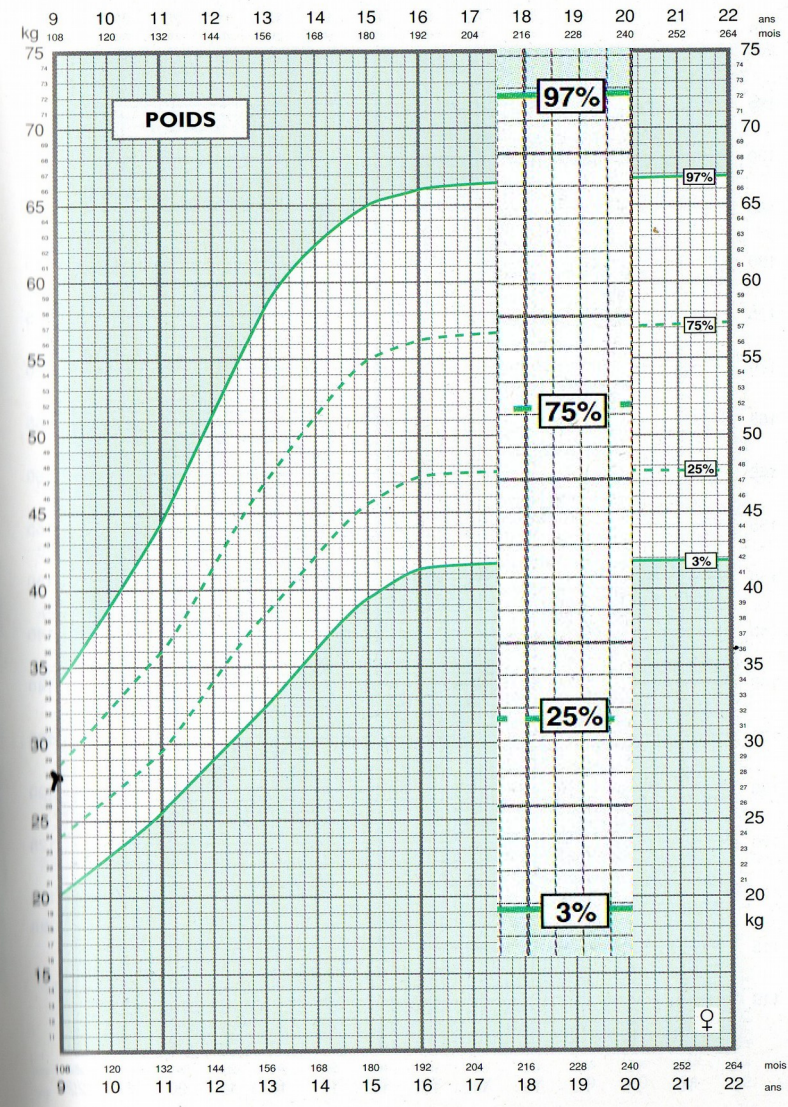
Quantiles : généralise les précédents

Percentiles

Croissance somatique des filles de 9 ans à 22 ans



Croissance somatique des filles de 9 ans à 22 ans



Tracés établis en CENTILES modélisés J.P.P.S. C.S. - Pr M. SEMPE 1995

Moyennes

Moyenne arithmétique : somme des observations divisées par le nombre d'observations

= *moyenne au sens de la somme comme dans une suite... arithmétique* $u_{n+1} = u_n + r$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Moyenne géométrique : racine n^{ème} du produit des observations

= *moyenne au sens du produit comme dans une suite... géométrique* $u_{n+1} = r \cdot u_n$

$$\bar{x}^G = \sqrt[n]{\prod_{i=1}^n x_i}$$

Exemple : nombre de participants à une manifestation

Les manifestants étaient 10000 selon les organisateurs et 100 selon les autorités.
Supposons que chaque camp a volontairement multiplié ou divisé le chiffre réel (=1000) par 10.

La moyenne arithmétique vaut $(10000+100)/2 = 5050$

La moyenne géométrique vaut $(10000 \cdot 100)^{1/2} = 1000$ (on retrouve bien la valeur réelle)

Et aussi : moyenne harmonique, moyenne quadratique...

Moyenne et médiane

Moyenne : somme des observations divisées par le nombre d'observations

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Médiane : valeur qui sépare l'échantillon en 2 sous-ensembles de tailles égales.

Un exemple

Exemple : supposons que l'entreprise compte 12 personnes :

- 8 ouvriers : ~ 1000 € / personne
- Un chef d'atelier : 2000 €
- Un directeur technique : 4000 €
- Un directeur des ressources humaines : 6000 €
- Un directeur général : 10 000 €

970
975
980
985
990
995
1005
1100
2000
4000
6000
10000

Moyenne : 2500
Médiane : 1000

Salaire moyen : **2500 €** - Salaire médian : **1000 €**

Comment augmenter le salaire moyen dans cette société ?

Solution 1 : augmenter tous les salaires de x %

Solution 2 : le dirigeant augmente son salaire

Solution 3 : « remercier » quelques postes de bas salaire

	1018.50		970		
	1023.75		975		
	1029		980		980
	1034.25		985		985
	1039.50		990		990
	1044.75		995		995
	1055.25		1005		1005
	1155		1100		1100
	2100		2000		2000
Solution 1	4200	Solution 2	4000	Solution 3	4000
	6300		6000		6000
	10500		12000		12000
Moyenne :	2625	Moyenne :	2666	Moyenne :	2805.50
Médiane :	1050	Médiane :	1000	Médiane :	1052.50



Tiré de Pour la Science,
Fév-mars 2018 ; Jean-
Paul Delahaye citant
*Statistiques Méfiez-
vous !* de Nicolas Gauvrit

Un autre exemple

- Les salaires des ouvriers et des cadres ont baissé de 10 % entre 2016 et 2017.
- Le salaire moyen dans l'entreprise a augmenté de 150 % !!!

Est-ce possible sachant que l'effectif global de l'entreprise n'a pas évolué ?

		Ouvriers	Cadres	Salaire moyen
2016	Salaire	200 €	2000 €	363,64 €
	Effectif	1000	100	
2017	Salaire	180 €	1800 €	916,34 €
	Effectif	600	500	

Le paradoxe de Simpson

Tiré de Pour la Science,
Fév-mars 2018 ; Jean-
Paul Delahaye citant
Statistiques Méfiez-
vous ! de Nicolas Gauvrit

	Physique		Biologie		Cumul	
	G	F	G	F	G	F
Réussite	80	10	4	50	84	60
Echec	10	0	6	40	16	40
Total	90	10	10	90	100	100

Que constate-t-on de surprenant dans ces chiffres ?

- Les filles ont un meilleur pourcentage de réussite que les garçons en physique et en biologie.
- Les filles ont globalement un pourcentage de réussite inférieur à celui des garçons.

Revue de presse

L'an dernier, **le salaire moyen brut a progressé de 2,1%**, soit près d'un point de plus que l'inflation (+1,2%). Mais cette hausse ralentit.

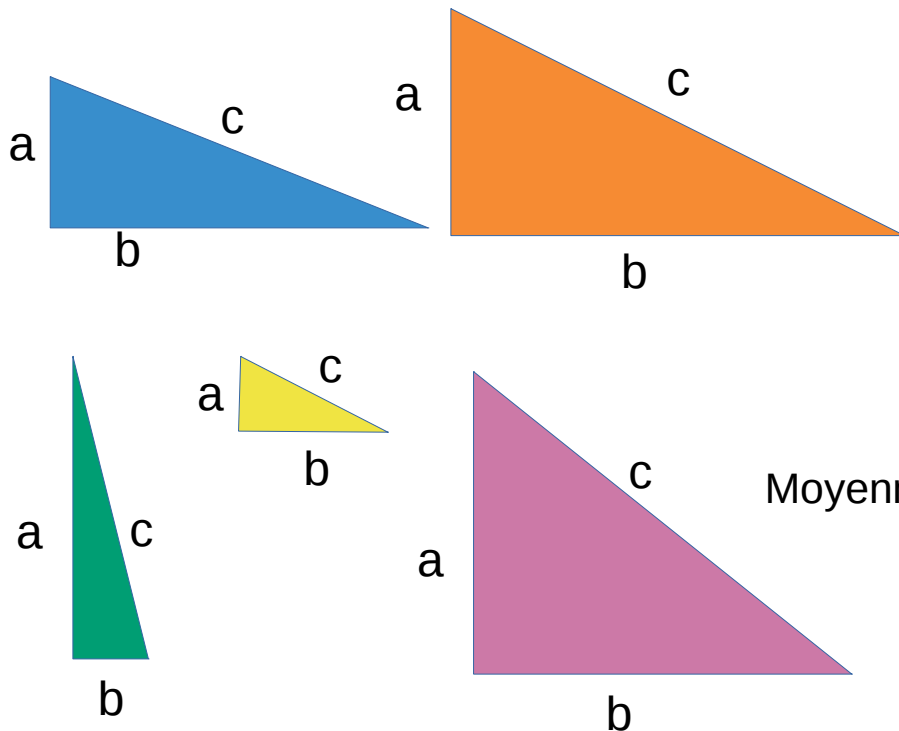
Petite éclaircie pour le pouvoir d'achat des Français. L'Agence centrale des organismes de Sécurité sociale (Acos) a annoncé les chiffres sur l'évolution du **salaire moyen** brut en France. En 2012, il a augmenté de 2,1% pour s'établir à 2.410 euros brut, soit 0,9 point de plus que l'inflation qui s'est tenu à 1,2%. Mais il apparaît que derrière le chiffre de cette progression, se cache en réalité, un ralentissement tendanciel. Au premier trimestre, la hausse était de 0,6%, puis de 0,5% aux deux trimestres suivants, pour terminer sur une progression de 0,4% au dernier trimestre.

Les effectifs salariés en légère baisse

La progression du salaire moyen s'explique en grande partie par la hausse de la masse salariale (+1,7%) alors que le nombre de salariés du secteur privé s'est légèrement contracté (-0,5%, soit 94.000 pertes nettes d'emploi). L'Acos note que la hausse de la masse salariale s'inscrit en net ralentissement puisqu'elle avait été de 3,5% l'année précédente. La baisse des effectifs est davantage marquée dans le secteur immobilier (-1,1%). En revanche, l'organisme note une hausse des effectifs dans l'action sociale (+0,4%), et dans la restauration (+0,2%). En termes géographiques, "les régions du nord-est" enregistrent les plus fortes baisses d'effectifs salariés. L'emploi a toutefois progressé dans les Dom, en Corse, en Midi-Pyrénées et en Ile-de-France.

La moyenne, encore...

- *Whenever I read statistical reports, I try to imagine my unfortunate contemporary, the Average Person, who, according to these reports, has 0.66 children, 0.032 cars, and 0.046 Tvs. Kato Lomb*
- Travaux d'Adolphe Jacques Quetelet. Pour une population donnée, comment trouver un « bon » représentant ? « Un homme moyen » peut-il représenter un groupe d'hommes ?



a	b	c
2	5	5.4
3	6	6.7
4	1	4.1
1	2	2.2
4	5	6.4
2.8	3.8	4.97

Moyenne

$$\sqrt{2.8^2 + 3.8^2} \neq 4.97$$



Le triangle
« moyen »
n'est pas
rectangle.

Humour...



« La majorité des français sont plus cons que la moyenne »

Gustave Parking



Durant les élections de 1956, un électeur apostropha le candidat Adlai E. Stevenson :

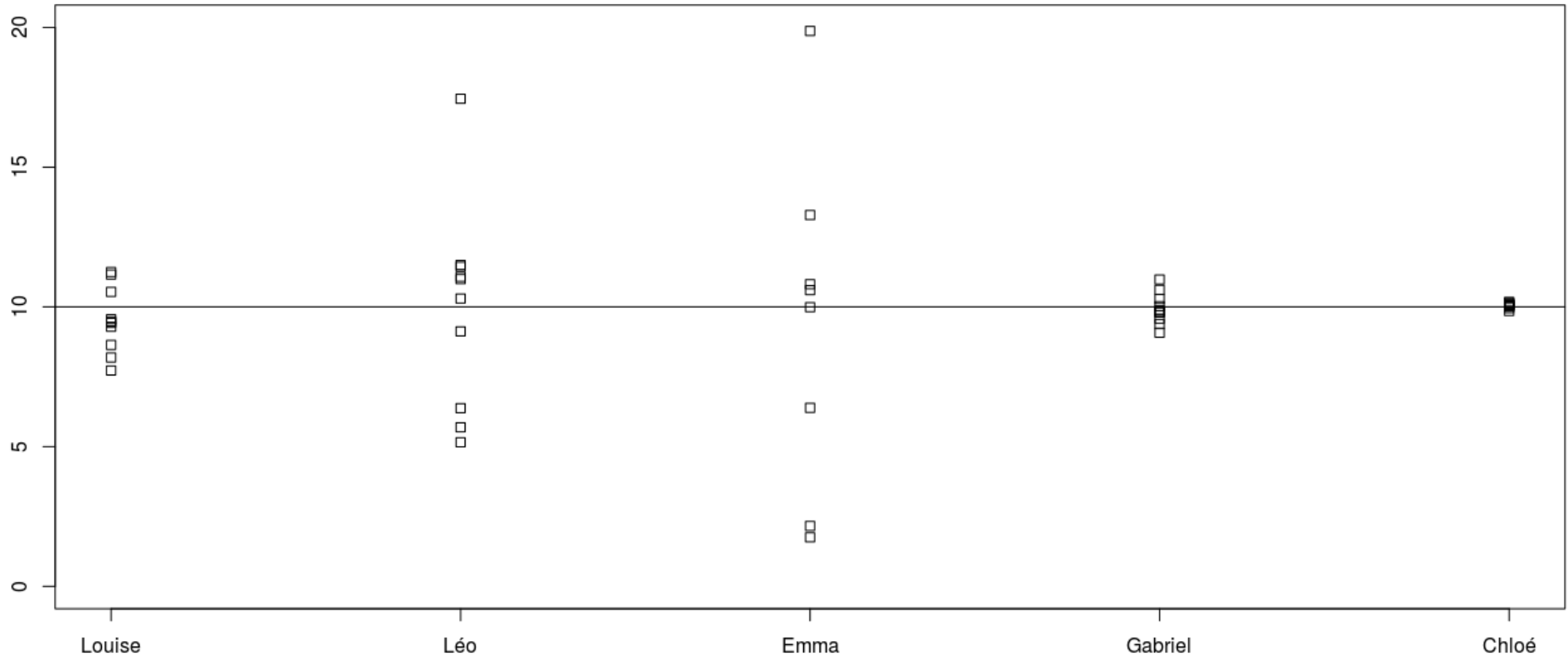
« Sénateur, vous avez pour vous toutes les personnes intelligentes ». Stevenson répondit : « Ce n'est pas suffisant, nous avons besoin d'une majorité ! »

Que se passe-t-il lorsqu'une blonde française passe la frontière pour aller en Belgique ?

Le QI moyen des 2 pays augmente !



Limites des indicateurs de position



Ce graphique représente les notes obtenues par 5 élèves dans 10 matières différentes. Les 5 élèves ont une moyenne (et une note médiane) à peu près égale à 10. Donneriez-vous la même appréciation (*Élève moyen*) à chacun de ces élèves ?

[illegible]

Indicateurs de dispersion

- **Étendue** : différence entre la plus grande et la plus petite valeur d'un échantillon
- **Variance** : Moyenne des carrés des écarts à la moyenne

$$\text{var}(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

- **Écart-type** : racine carrée de la variance

www.variance.fr

Variance



e-shop

Lingerie
P'tites culottes
Lingerie de nuit

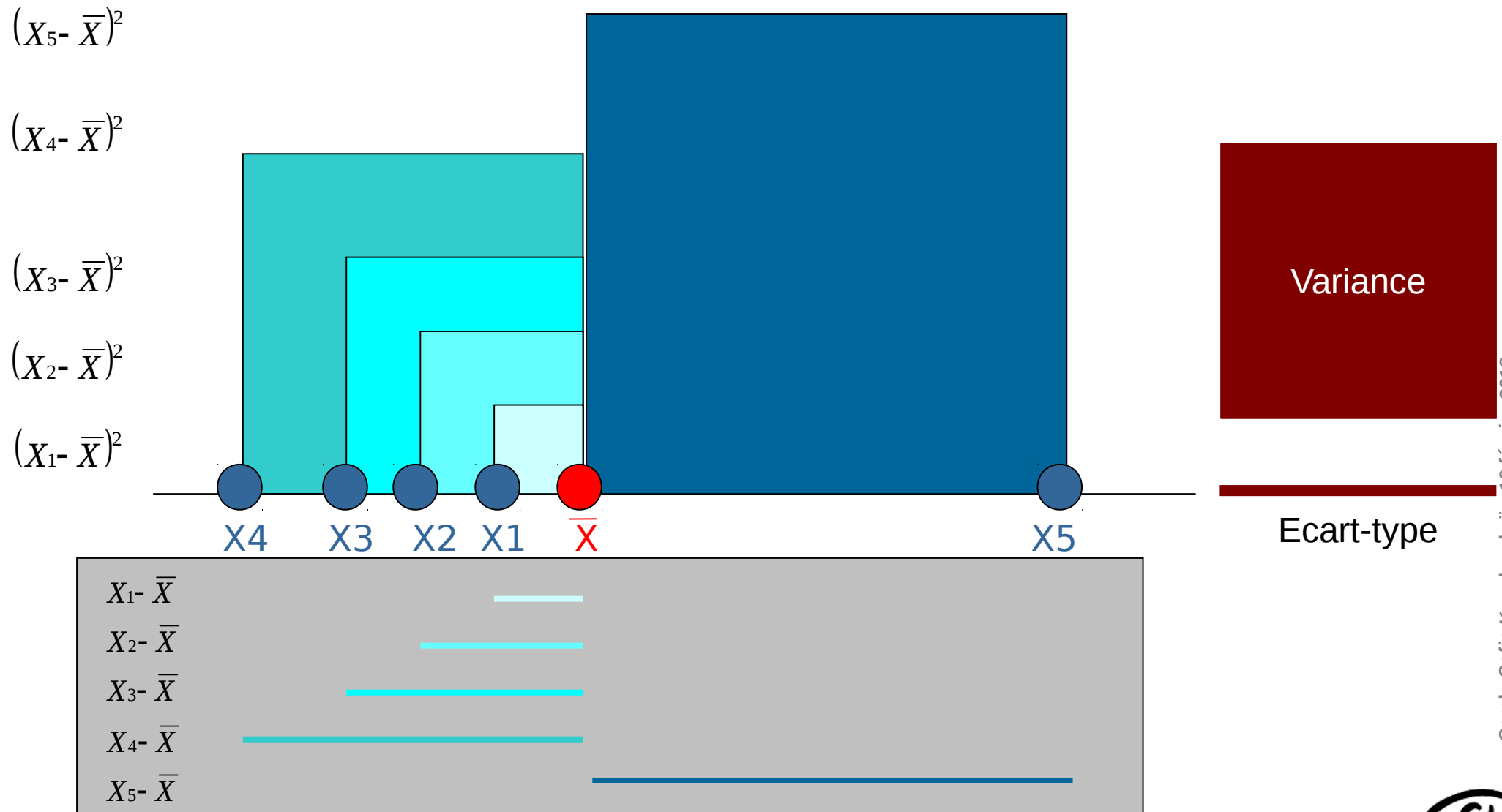
La
marque

Contact | Mentions légales | © 2013 Variance

managed by 

Dessiner la variance et l'écart-type

Racine carrée de la moyenne des carrés des écarts à la moyenne



Revue de presse

Les filles brillent en classe, les garçons aux concours

LE MONDE | 07.09.09 - Article paru dans l'édition du 08.09.09. Philippe Jacqué

Elles obtiennent de meilleurs résultats en cours de scolarité, mais réussissent moins bien les concours des meilleures grandes écoles que les hommes. Raison : les femmes souffriraient plus dans un *"environnement concurrentiel"*.

[...]

Les conclusions de cette étude sont accablantes. Les candidates aux concours de l'école de Jouy-en-Josas (Yvelines) ont beau avoir de meilleurs dossiers que leurs concurrents masculins (mentions au bac supérieures, meilleure représentation dans les bonnes classes préparatoires), elles y réussissent moins bien. Alors que le pourcentage d'hommes et de femmes candidats est équilibré sur les trois années étudiées (50,84 % d'hommes, 49,16 % de femmes), le pourcentage de femmes admissibles tombe à 46,32 %, et celui d'admissibles à 45,92 %... Pis, après le concours, *"celles qui l'ont réussi obtiennent en première année en moyenne des notes d'examen supérieures à celles de leurs congénères masculins."*

[...]

[...]

"Autour de la moyenne"

Pour expliquer la moindre réussite des femmes, une rumeur court depuis de nombreuses années : les femmes seraient discriminées aux oraux. *"Si un des jurys d'oral peut avoir des biais, aucune consigne n'est donnée en ce sens", assure Frédéric Palomino. "Nous avons mené cette enquête statistique pour tordre le cou à ce fantasme", explique Eloïc Peyrache. "De plus, quand on regarde les pourcentages de réussite, on voit que c'est à l'écrit que la part des candidates chute le plus."*

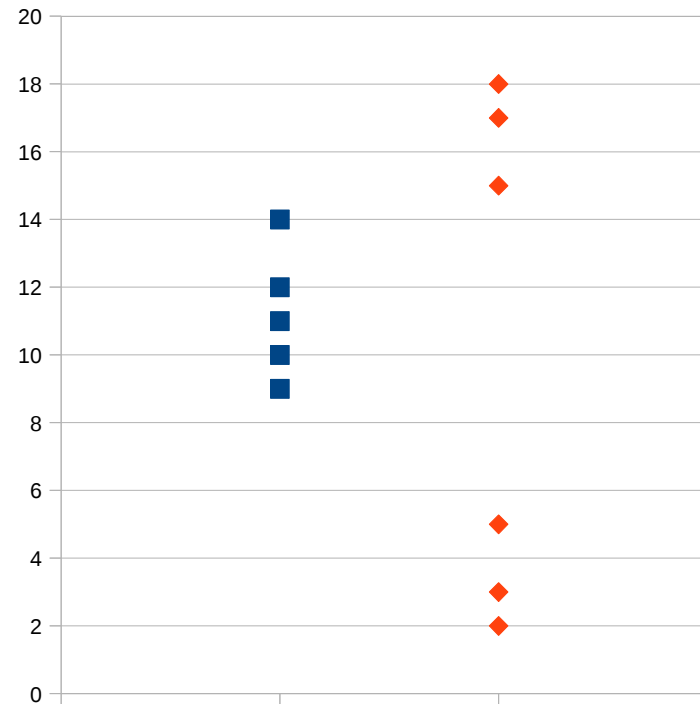
Alors comment comprendre ce déséquilibre ? ***"D'un point de vue technique, il semble que la structure du concours HEC crée d'avantage d'hétérogénéité chez les hommes que chez les femmes", estime M. Peyrache. Si, "en moyenne", les performances des hommes et des femmes sont similaires, "les notes des femmes sont concentrées autour de la moyenne, tandis que celles des hommes sont très dispersées avec beaucoup de très bonnes notes et de très mauvaises. Mécaniquement, quand on sélectionne les 380 premiers résultats, on a un peu plus d'hommes".***

Circle Koffi Kouamevissita, février 2018

Illustration

	Femme	Homme
	9	2
	11	18
	10	15
	12	3
	12	17
	14	5

Moyenne	11.33	10
Etendue	5	16
Ecart-type	1.75	7.43
Variance	3.07	55.2

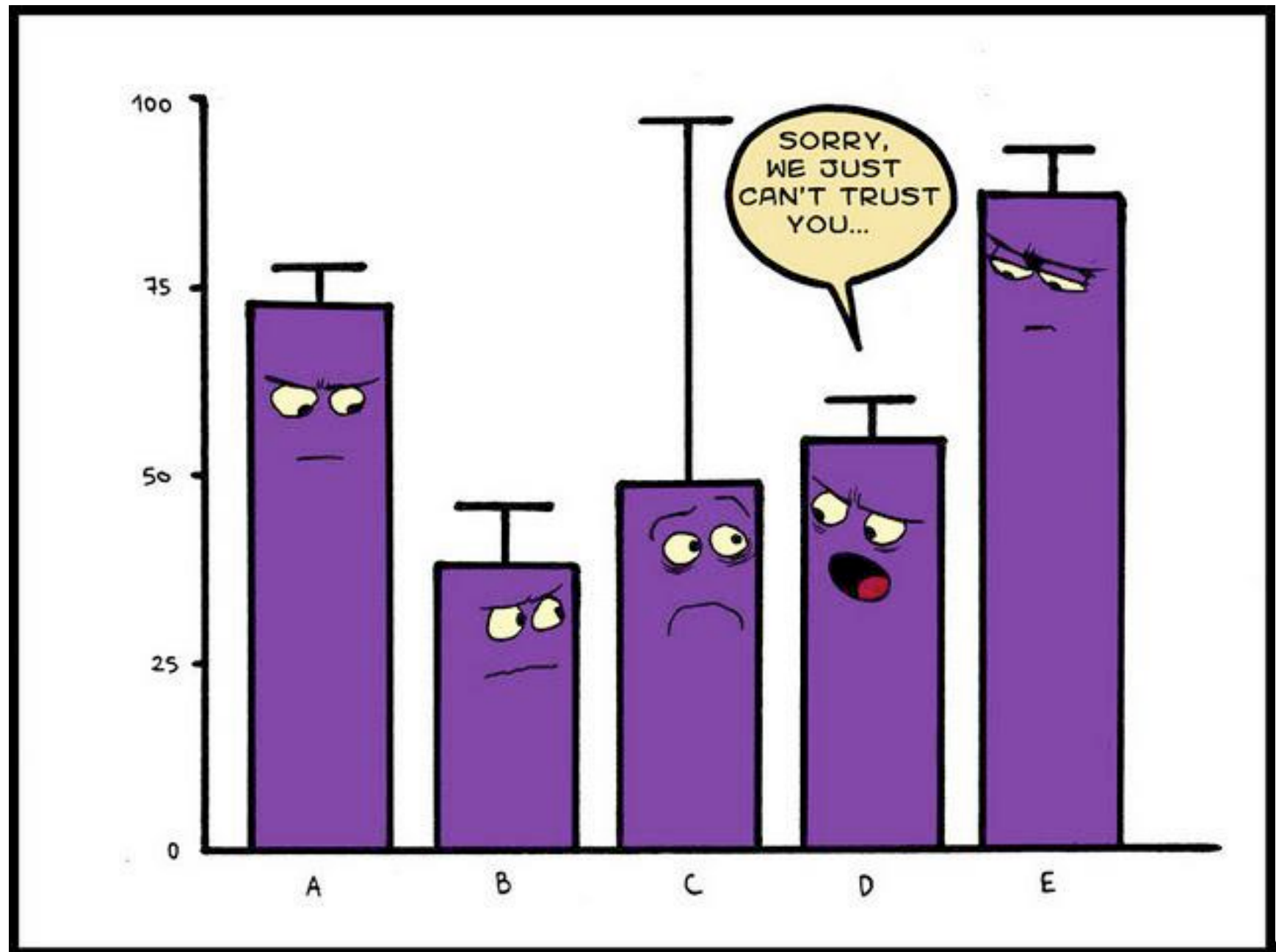


Classement

18
17
15
14
12
12
12
11
10
9
5
3
2

Humour

Source : xkcd.com



Indicateurs pour 2 séries de données

Deux séries de données acquises pour les mêmes individus varient-elles de la même façon ?

Covariance

Covariance

$$\text{cov}(X,Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

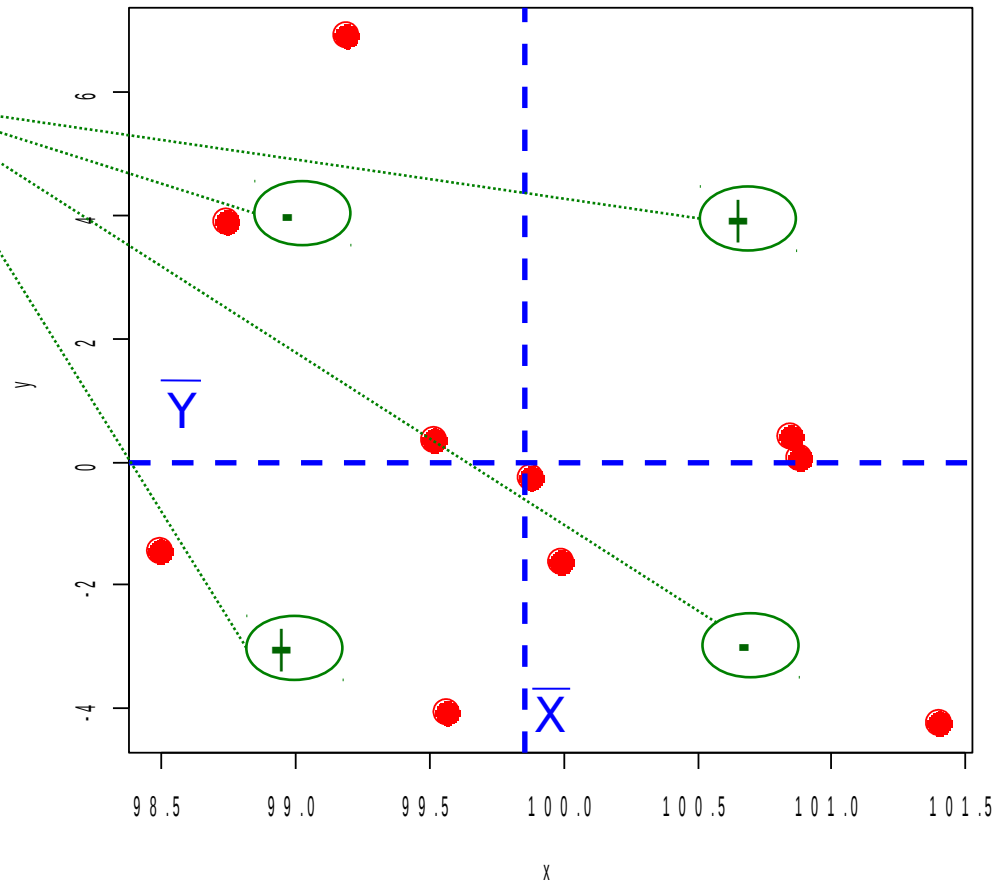
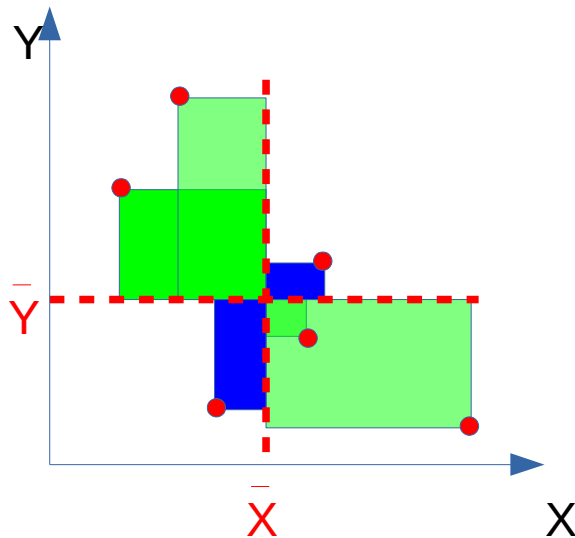
$$\text{cov}(X,X) = \text{var}(X)$$

Signe du produit $(X_i - \bar{X})(Y_i - \bar{Y})$

Intuitivement :

- Si les **+** l'emportent
→ liaison linéaire positive
- Si les **-** l'emportent
→ liaison linéaire négative

Sur cet exemple : $\text{cov}(X,Y) = -1.36$



La covariance dépend des unités de mesure →
coefficient de corrélation

Corrélation

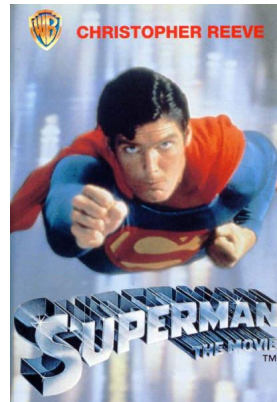
- Coefficient de corrélation **linéaire** de Pearson

$$\rho(X, Y) = \text{cov}(X, Y) / (\sigma_X \sigma_Y)$$

- Coefficient de corrélation de Spearman
→ Robustesse due au travail sur les rangs



$$\rho_s(X, Y) = \rho(RX, RY) = 1 - 6 \frac{\sum_{i=1}^n d^2}{n(n^2 - 1)}$$



(1) calcul des rangs

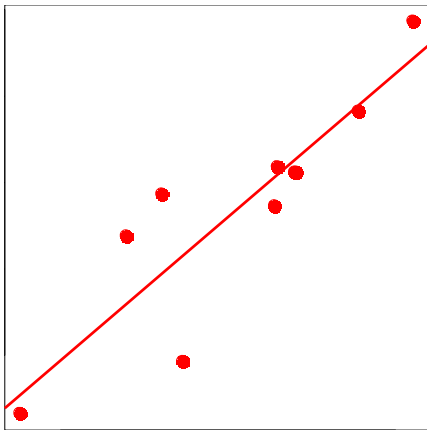
(2) différence des rangs

(3) carrés des différences des rangs

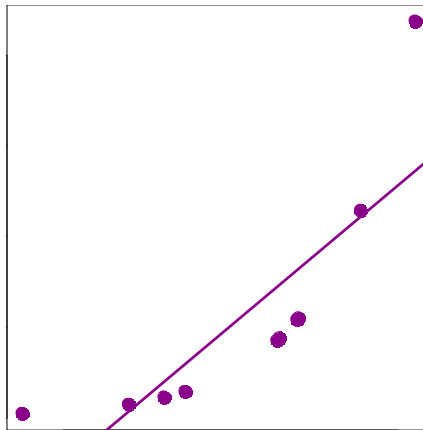
(4) somme des carré des différences des rangs

			(1)	(2)	(3)	
X	Y		RX	RY	d = RX-RY	d²
20,6	20,7		6	7	-1	1
21,6	21,8		2	2	0	0
18,8	20,4		9	8	1	1
20,8	21,1		3	4	-1	1
17,5	18,3		10	10	0	0
19,5	18,9		7	9	-2	4
20,8	21,1		4	4	0	0
20,6	21,2		5	3	2	4
19,2	20,9		8	6	2	4
22,2	22,9		1	1	0	0
					Somme	15

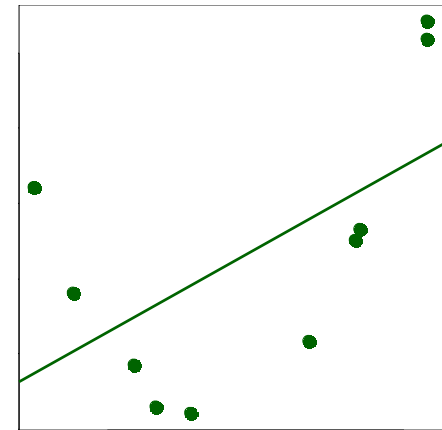
Exemples



$$\rho = 0.884 - \rho_s = 0.9$$



$$\rho = 0.822 - \rho_s = 1$$



$$\rho = 0.584 - \rho_s = 0.491$$

Eduardo
Mendoza
Les égarements
de mademoiselle Baxter

Un exemple de corrélation négative : *Ma vie et celle du commissaire Flores* avaient suivi des lignes à la fois divergentes et concomitantes : il montait et je descendais dans une corrélation non fortuite, attendu que ses mérites se fondaient généralement sur mes échecs.

Eduardo Mendoza (traduction de Delphine Valentin), *Les égarements de mademoiselle Baxter*

Corrélation $\neq$ causalité

La fréquence des attaques de requins est fortement corrélée à la vente de glaces sur les plages ! Manger des glaces rend-il plus appétissant pour les requins ?

Faut-il manger du chocolat pour avoir des prix Nobel ?

©REUTERS/Denis Balibouse ©REUTERS/Denis Balibouse

La revue médicale New England Journal of Medicine vient de publier une étude qui fait le lien entre une forte consommation de chocolat et l'attribution des Nobel.

New England Journal of Medicine, hebdomadaire américain, publié depuis 1812, est considéré comme la revue médicale la plus prestigieuse.

Selon l'équipe de chercheurs, les flavonoïdes, de puissants antioxydants qu'on trouve en grande quantité dans les fèves de cacao, le thé vert et le vin rouge, ont montré qu'ils réduisent le risque de démence et améliorent les fonctions mentales chez les personnes âgées.

Le docteur Franz Messerli, de l'université Columbia à New York et auteur de l'étude, explique "qu'il y a une **corrélation significative surprenante** entre la consommation de chocolat per capita et le nombre de lauréats du Nobel pour dix millions d'habitants dans un total de 23 pays".

La Suisse arrive en tête à la fois en nombre de Nobel (rapporté à sa population) et en consommation de chocolat... Suivent ensuite, la France, l'Allemagne et les États-Unis. Inversement, la Chine, le Japon et le Brésil se montrent moins gourmands en chocolat et aussi en Nobel.

Seule exception de l'étude : la Suède. Les habitants ne consomment "que" 6,4 kilos de chocolat par an et par personne pour un total de 32 Nobel. Qu'à cela ne tienne, pour les chercheurs il s'agirait d'un simple favoritisme du comité Nobel.

Si une corrélation est montrée pour les pays, l'étude ne dit rien en revanche sur le niveau de consommation individuel de chocolat des lauréats du Nobel.

Le chocolat engendre des tueurs en série

<http://plus.lefigaro.fr/lien/le-chocolat-engendre-des-tueurs-en-serie-20121123-1589103>

23/11/2012 | Mise à jour: 11:12 Réactions (24)

Sélectionné par la rédaction

Par Hayat Gazzane

Des chercheurs britanniques se sont amusés à démonter l'étude de Franz Messerli qui établit une corrélation forte entre consommation de chocolat et prix Nobel. En utilisant la même méthodologie, ils arrivent à prouver que les pays où l'on mange beaucoup de chocolat sont aussi ceux qui engendrent le plus de sérial killer et d'accidents de la route (étude en anglais).

Merci pour le chocolat ?

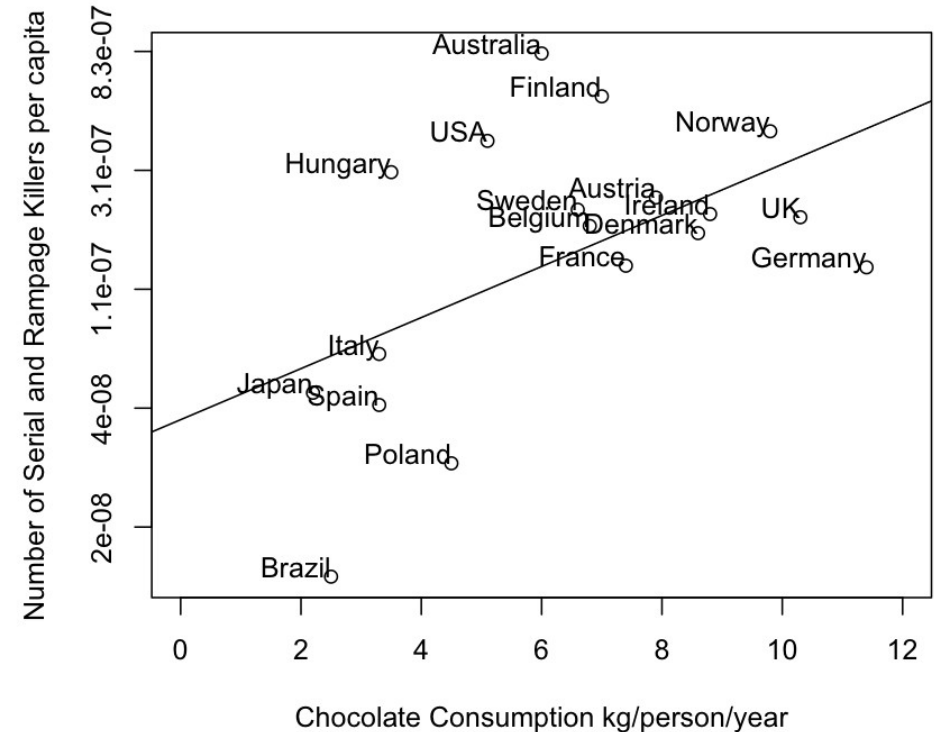
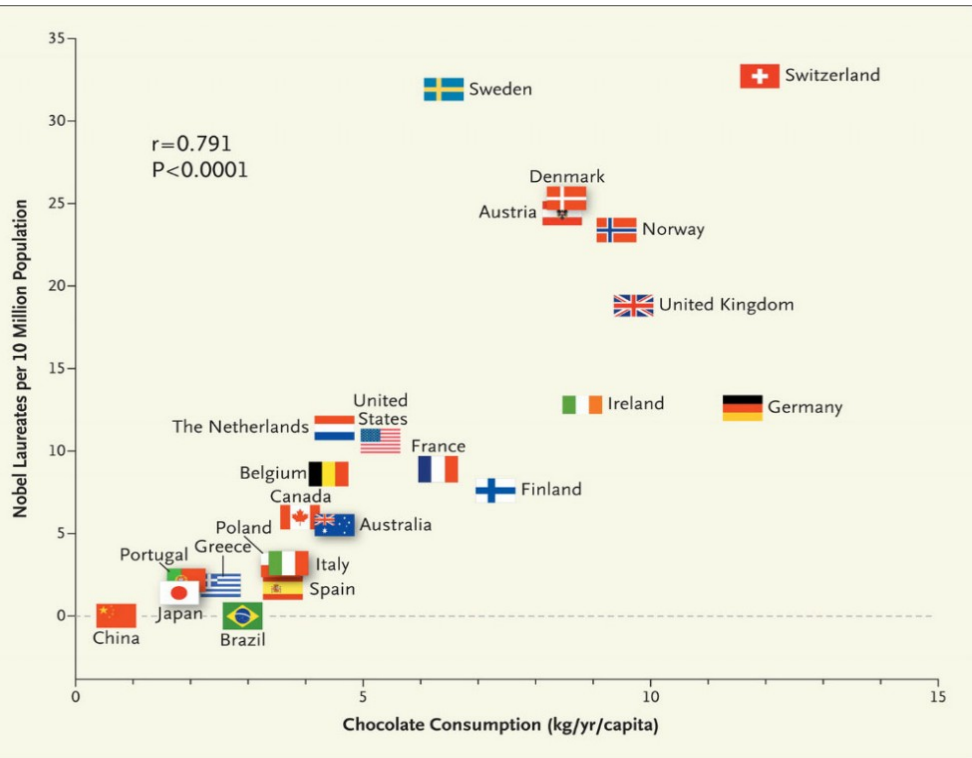
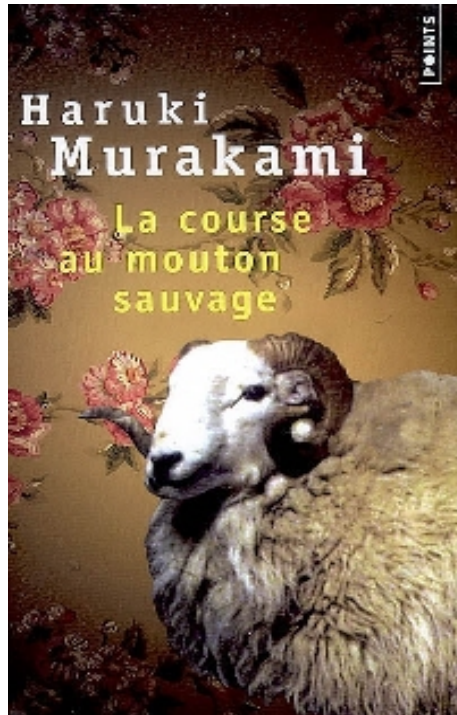


Figure 1: Correlation between countries' annual per capita chocolate consumption and the serial and rampage killers per capita since 1900.

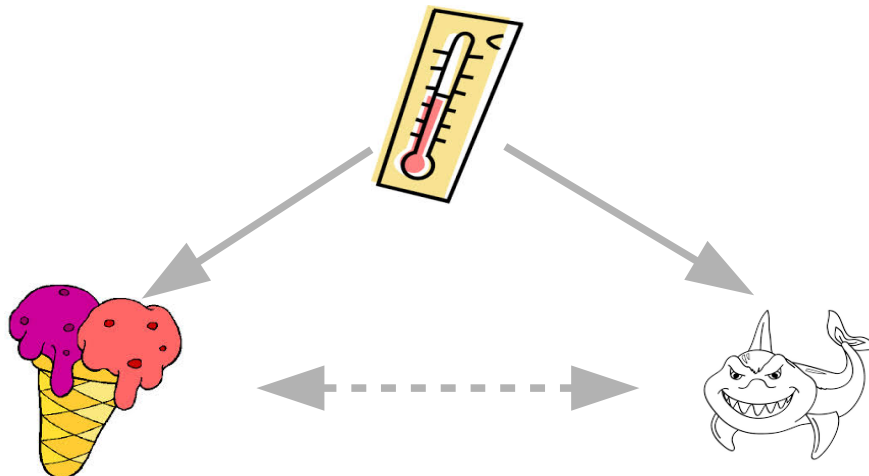
Revue de presse



Corrélation $\neq$ causalité



Votre hypothèse à vous, dis-je, serait qu'entre ces deux phénomènes il n'y aurait pas de relation de cause à effet mais une simple situation de parallélisme derrière laquelle il y aurait un autre et mystérieux facteur.

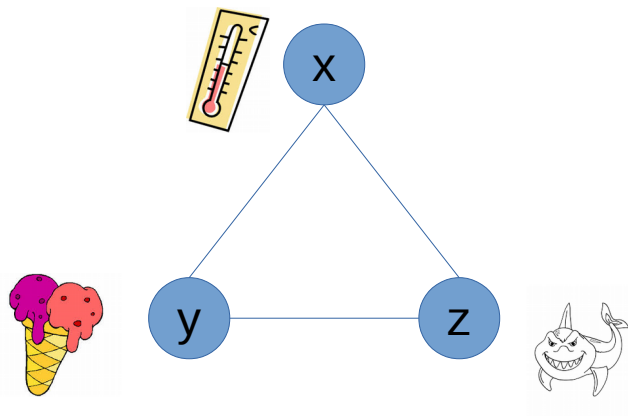


Corrélation partielle

Formellement, la corrélation partielle entre X et Y conditionnellement à Z est la corrélation entre les résidus RX et RY des régressions linéaires de X sur Z et de Y sur Z .

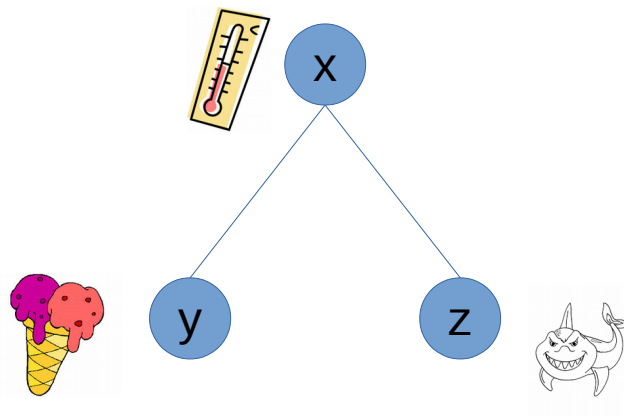
Matrice de corrélation

	x	y	z
x	1.00	0.95	0.93
y	.	1.00	0.88
z	.	.	1.00

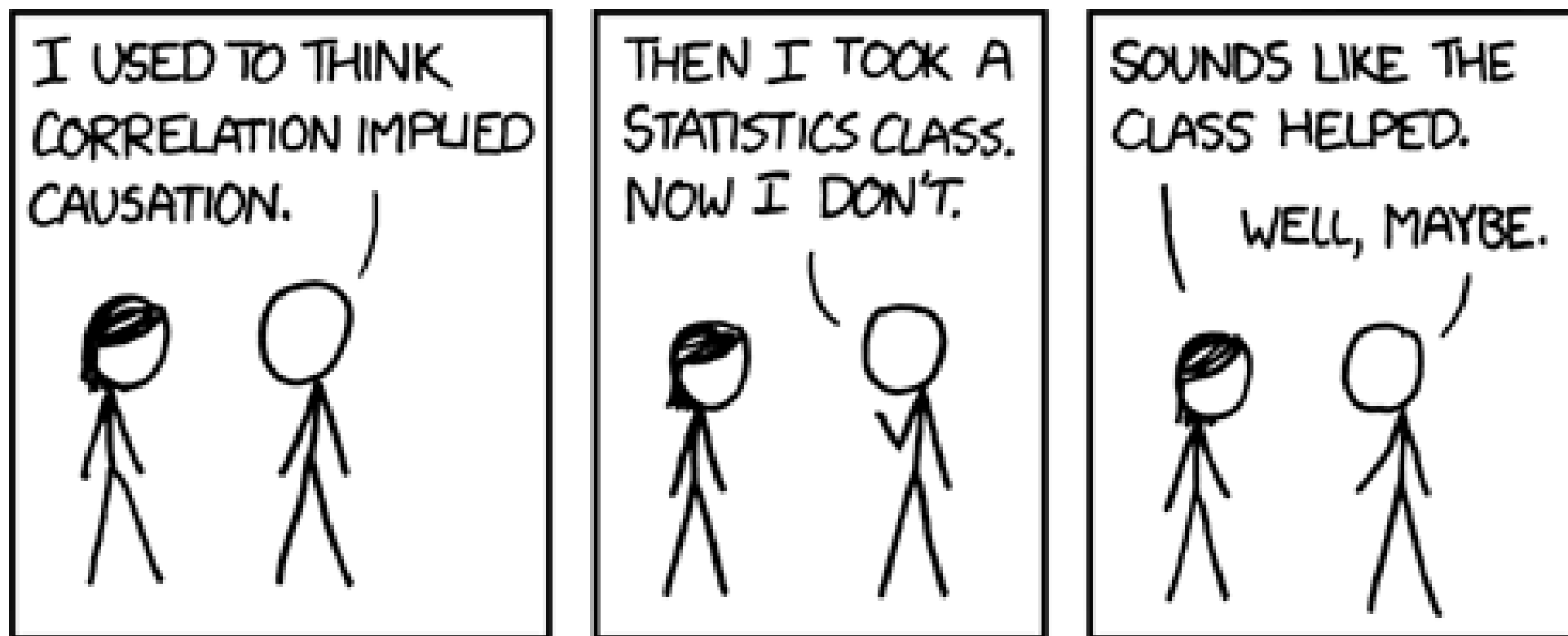


Matrice des corrélations partielles

	x	y	z
x	1.00	0.74	0.66
y	.	1.00	0.04
z	.	.	1.00



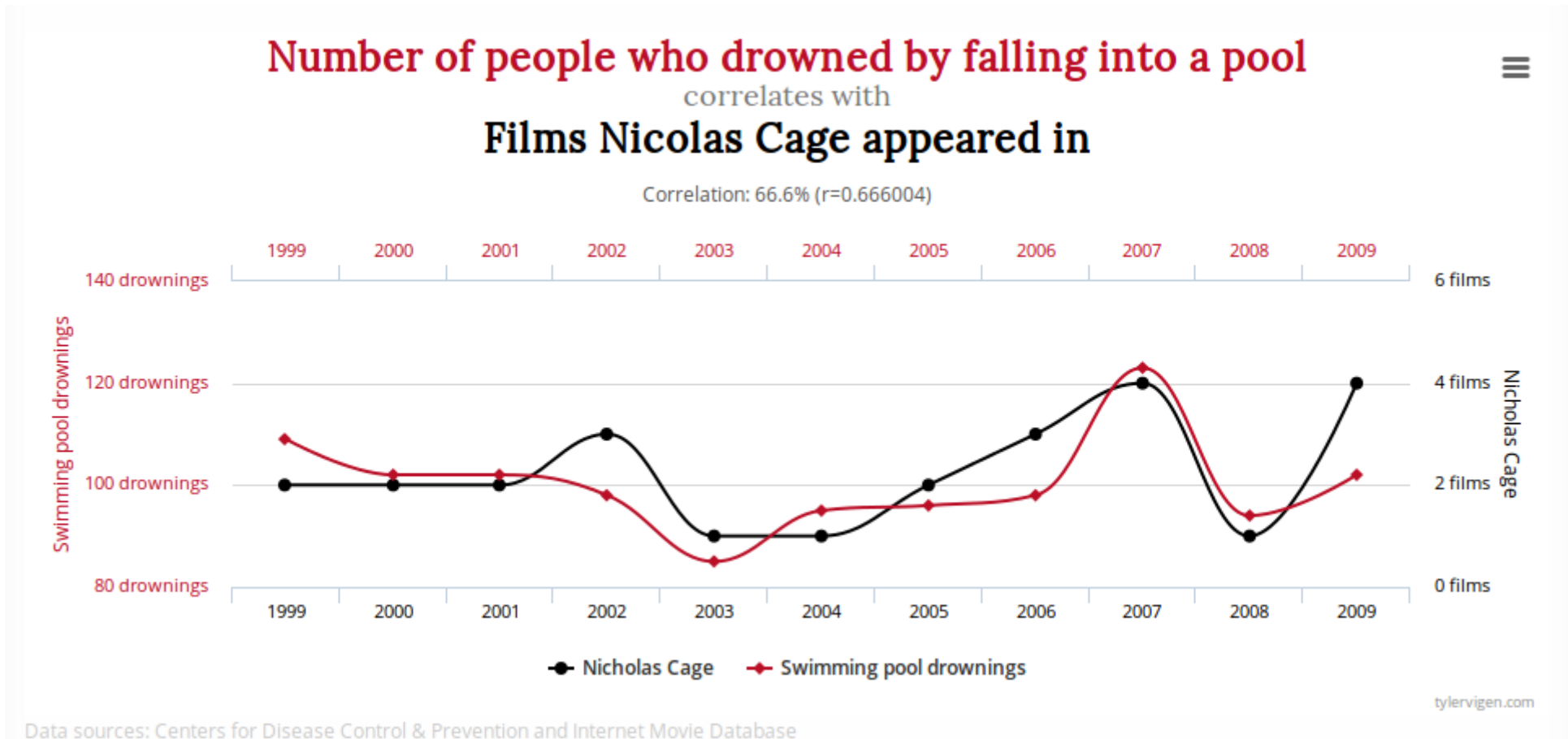
Humour de statisticien...



Source : xkcd.com

Spurious correlation

<http://www.tylervigen.com/spurious-correlations>



Divorce rate in Maine / Consumption of margarine ($p= 0.99$)

Consumption of Mozzarella cheese / Civil engineering doctorates awarded ($p= 0.96$)

...

Représentations graphiques

Représenter une ou plusieurs séries de chiffres
par un graphique

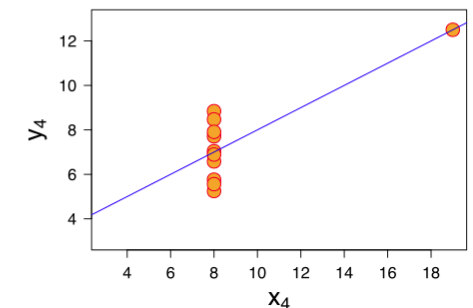
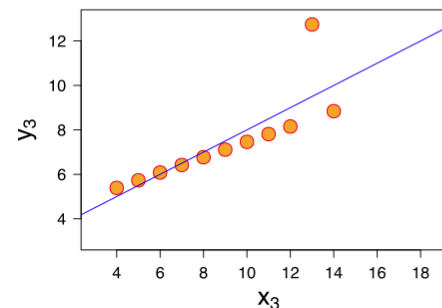
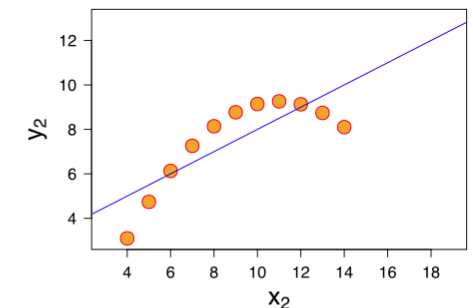
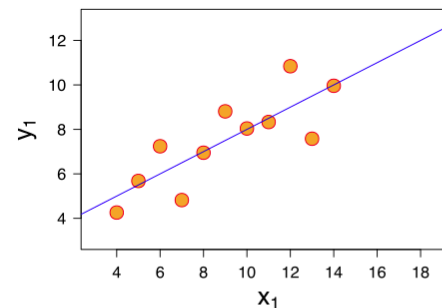
Pourquoi visualiser ?

...make both calculations and graphs. Both sorts of output should be studied; each will contribute to understanding.

F. J. Anscombe, 1973

Anscombe's quartet

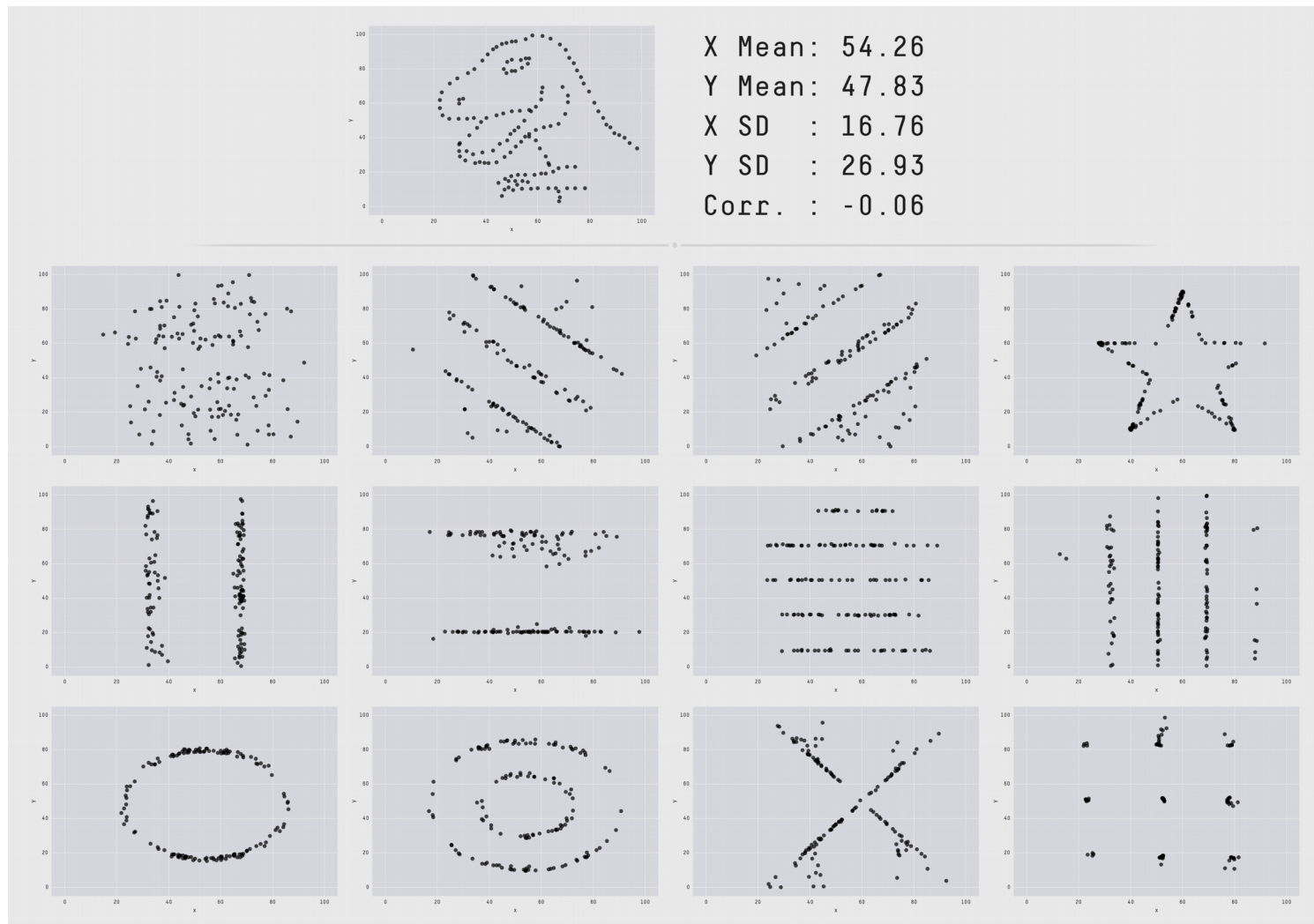
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89



Dans les 4 cas :

	X	Y
Moyenne	9	7.5
Variance	11	4.125
Corrélation		0.816

Same stats, different graphs...

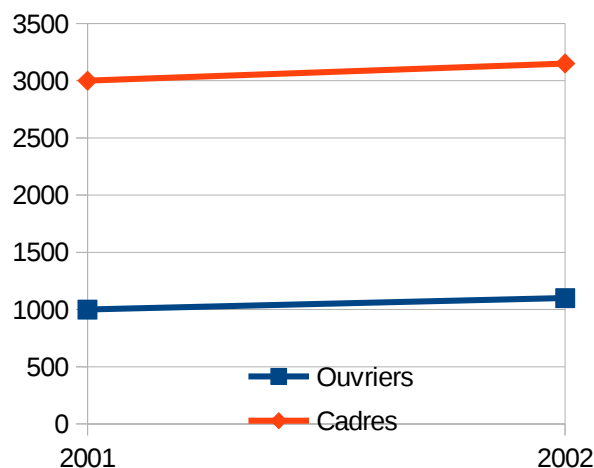


Matejka, J., & Fitzmaurice, G. (2017, May). **Same stats, different graphs**: Generating datasets with varied appearance and identical statistics through simulated annealing. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 1290-1294). ACM.

Un exemple

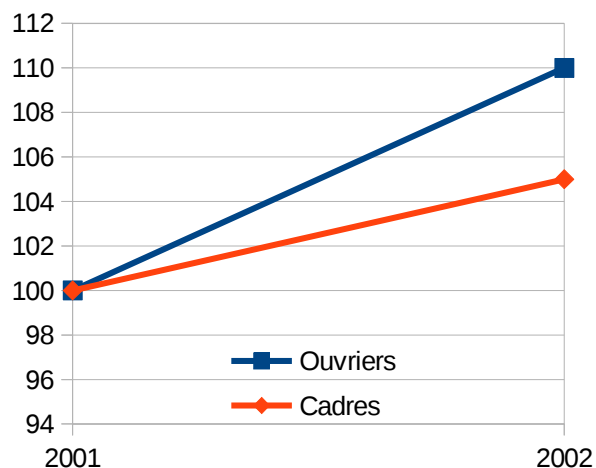
	2001	2002
Ouvriers	1000	1100
Cadres	3000	3150

Données brutes



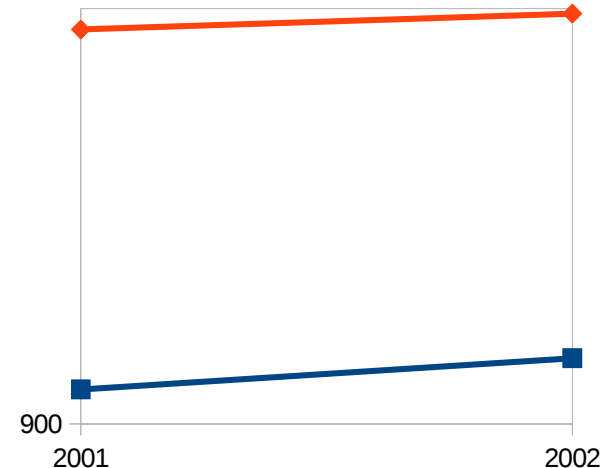
Les ouvriers et les cadres ont été augmentés. Les cadres recevaient plus et reçoivent encore plus que les ouvriers. L'écart entre les salaires empire.

Augmentations relatives



Les cadres ont été relativement moins augmentés que les ouvriers.

Échelle logarithmique



Les cadres recevaient et reçoivent encore plus que les ouvriers mais les inégalités se réduisent.

Un peu de théorie

Un graphique devrait :

- Montrer les données
- Inciter celui qui regarde à penser
- Éviter de distordre ce que les données ont à dire
- Présenter beaucoup de données sur une petite surface
- Révéler les données à des niveaux différents : d'un aperçu global à des structures plus fines
- Servir un objectif clair et raisonnable
- Être étroitement intégré à une description statistique du jeu de données

Edward Tufte, *The Visual Display of Quantitative Information*, Cheshire, CT, Graphics Press, 2001, 2e éd. (1re éd. 1983)

Représentations graphiques

Données de type « effectif »

A	30
B	15
C	30
D	20
E	25

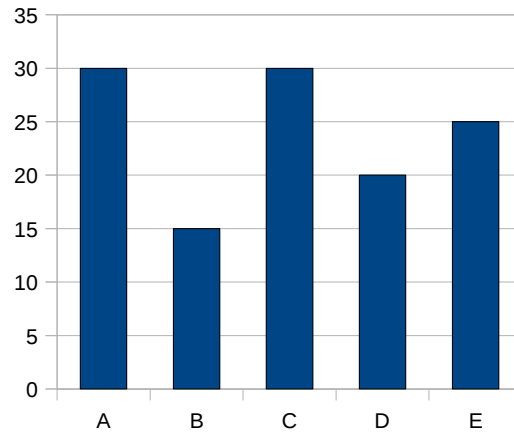
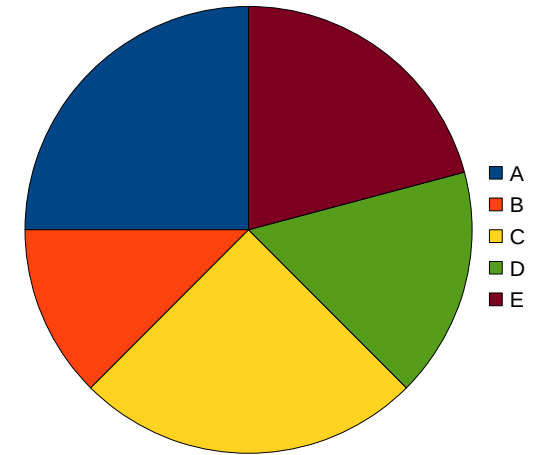
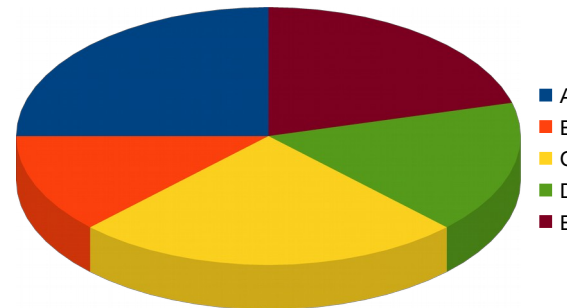
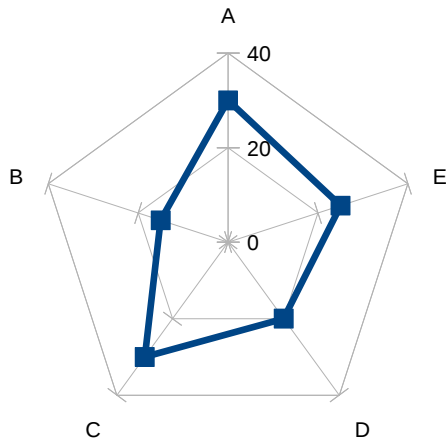


Diagramme en bâtons

Diagramme circulaire



Spider, radar plot

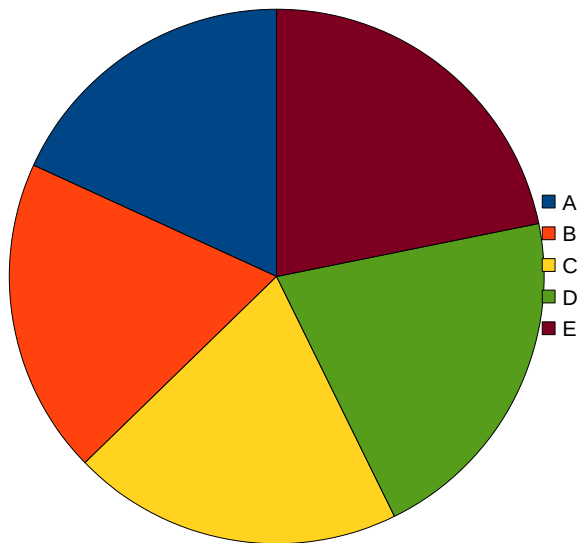


Attention aux représentations 3D. Est-ce si évident que A et C ont la même valeur ?

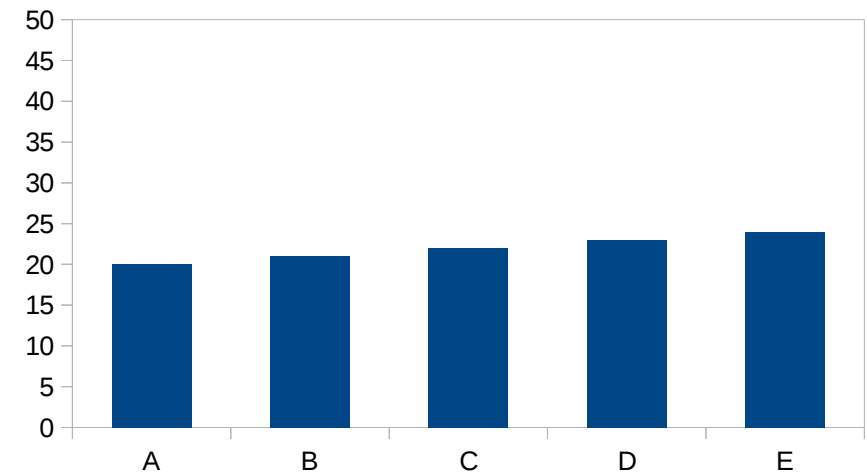
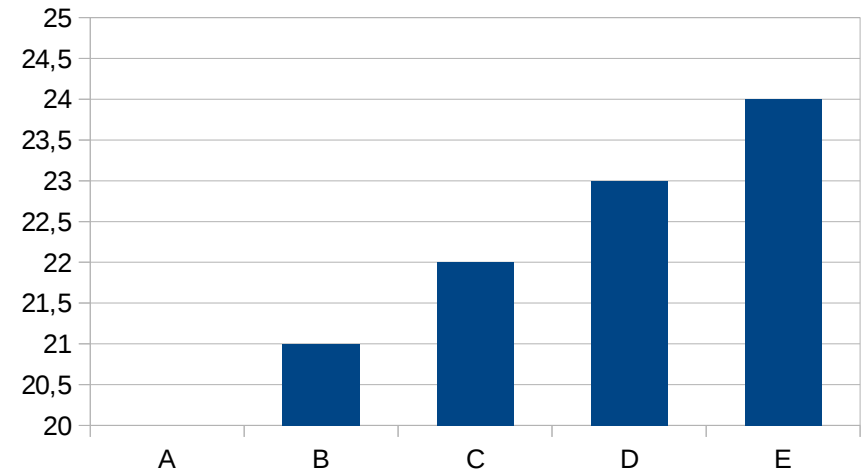


Représentations graphiques

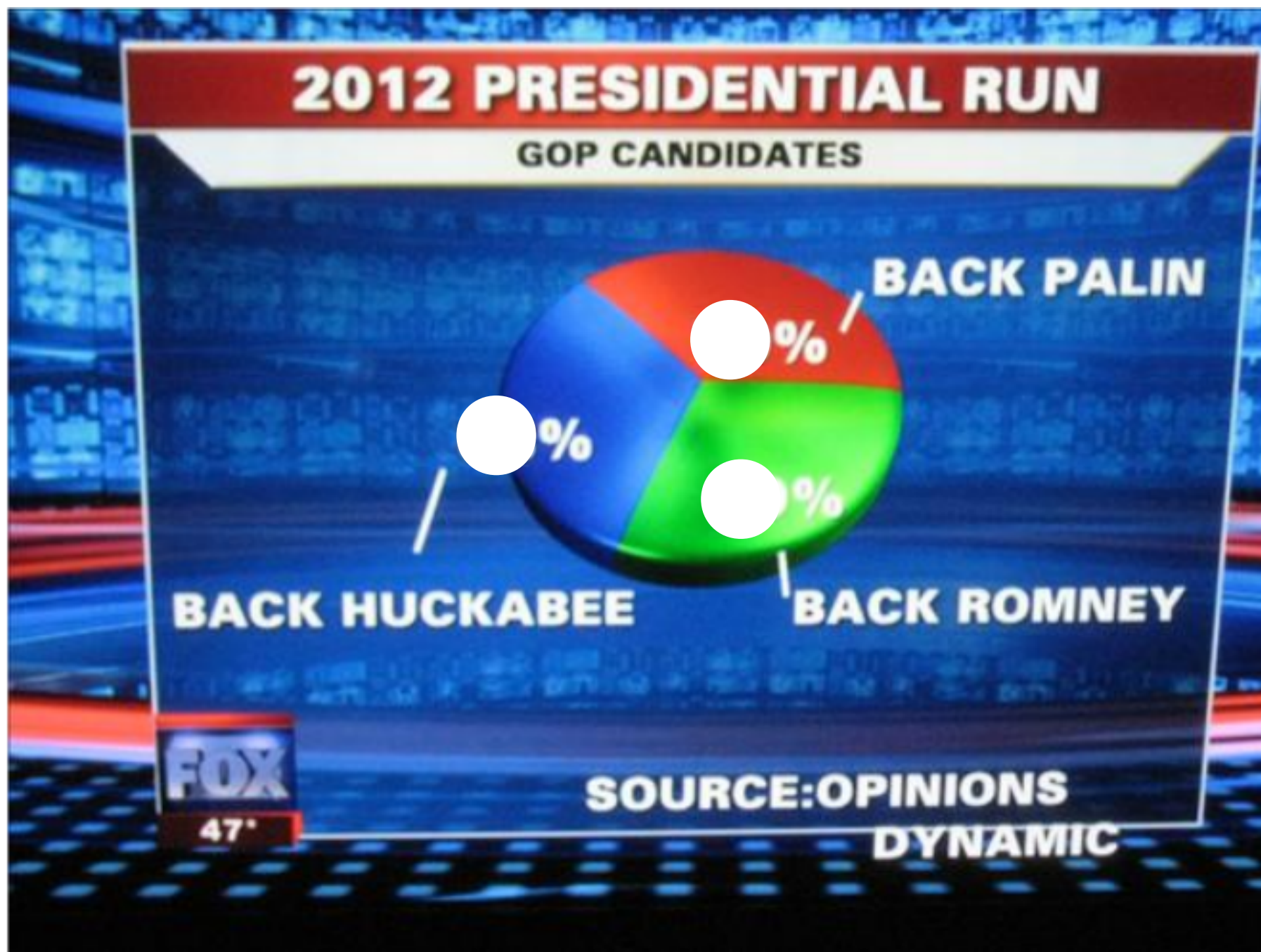
Données de type « effectif »



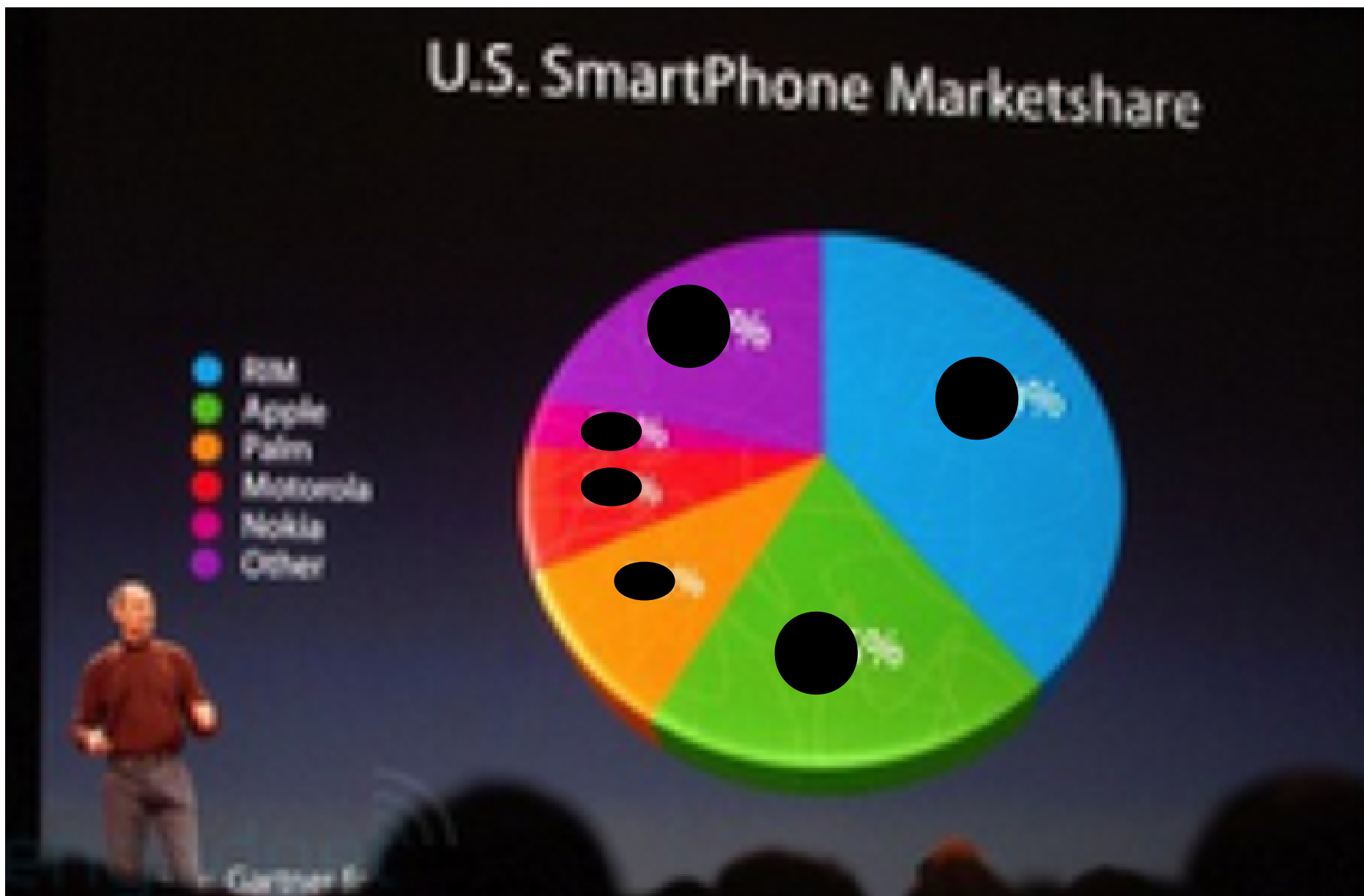
A	20
B	21
C	22
D	23
E	24



Hum, hum...

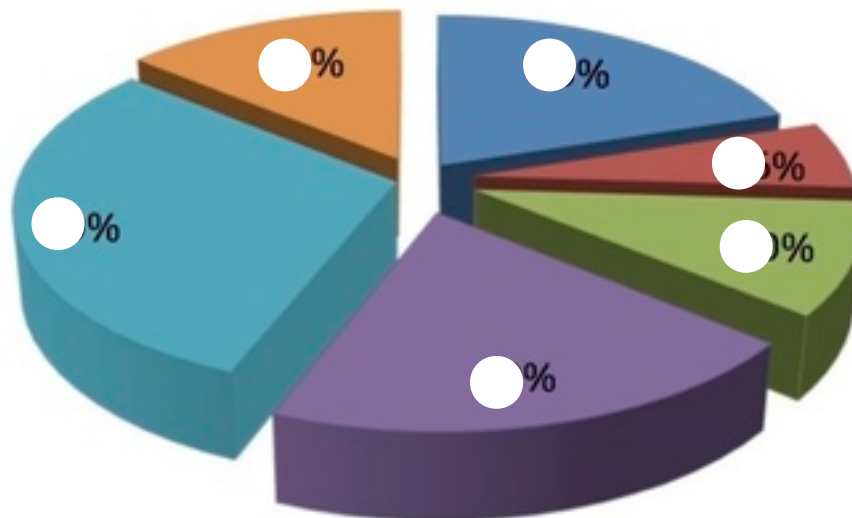


Hum, hum...



Hum, hum...

Budget Breakdown

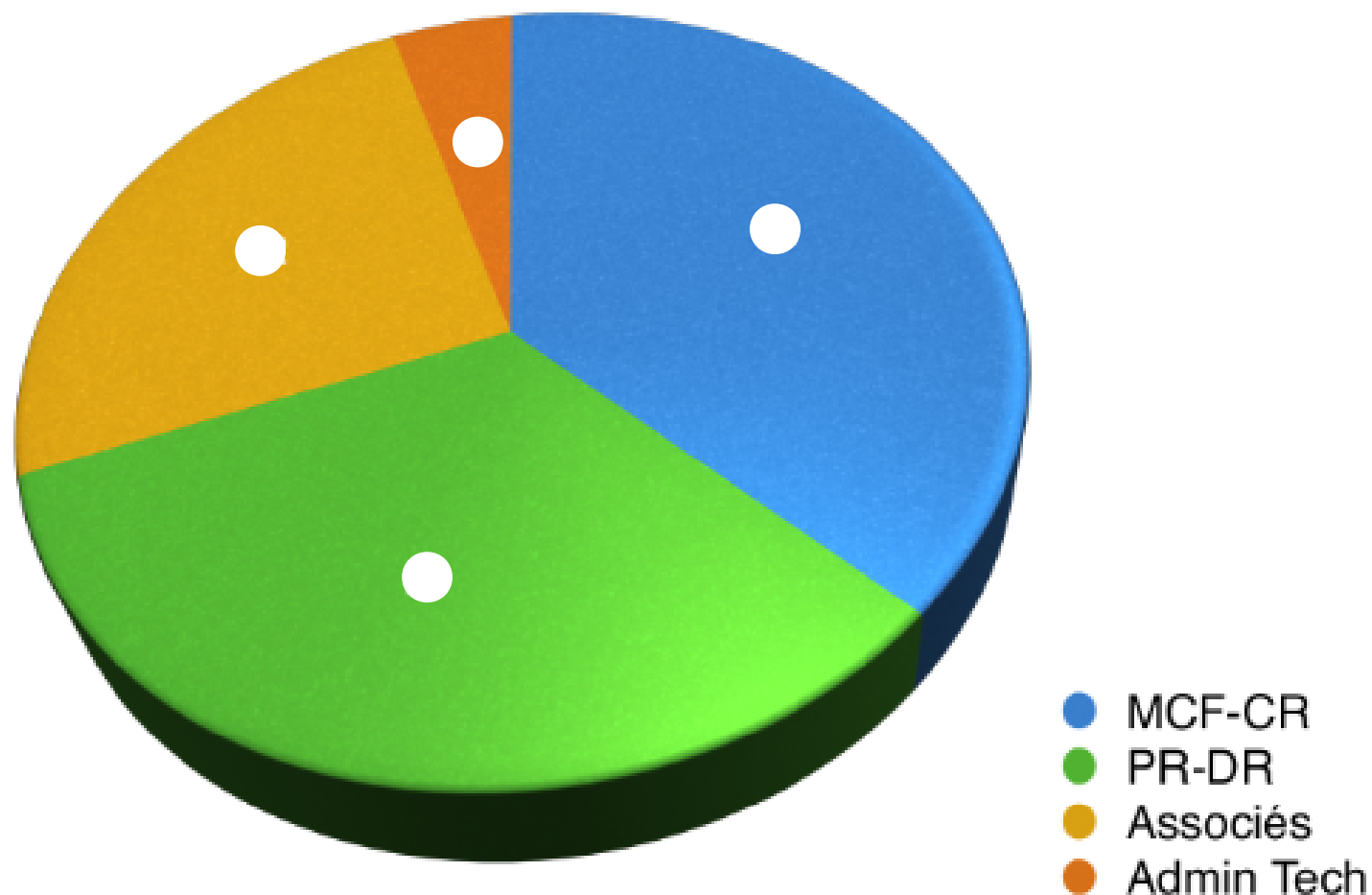


- Production Design (Locations, sets, props)
- Talent
- Misc. Location expenses (Permits, etc.)
- Post/Editorial (Editing, color correction, sound mixing)
- Equipment (cameras, lights, dollies, sound equipment, rental fees, etc.)
- Operations (catering, necessary expenses)

Hum, hum...

ESP in few facts : 88 Permanent people

Permanent STAFF



Revue de presse

VENDREDI 5 JUIN 2009

EUROPE 11

SONDAGE Ipsos dévoile les dernières intentions de vote pour « 20 Minutes »-SFR-« Le Point »

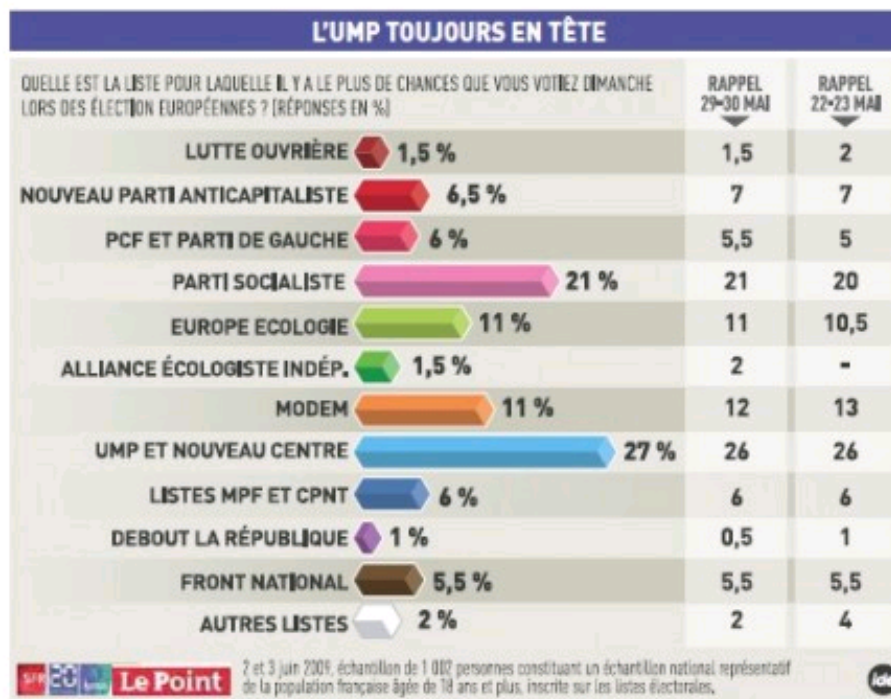
LE SCRUTIN DE DIMANCHE, UN TOUR DE CHAUFFE ?

BASTIEN BONNEFOUS

Qui sortira en tête dimanche soir ? Le PS finira-t-il 2^e ? A qui la 3^e place : au centre ou aux Verts ? Derniers points d'étape, d'après les intentions de vote Ipsos pour 20 Minutes-SFR-Le Point.

► **UMP-Majorité parlementaire** Plus vraiment de doute, le parti présidentiel devrait finir en tête du scrutin, avec quelque 27 % prévus. L'écart avec le PS se creuse de six points, mais, surtout, 85 % de l'électorat UMP dit avoir fait définitivement son choix. Un socle solide, qui permettrait ainsi à l'UMP de passer la barre des 25 % qu'il s'est lui-même fixée. Serait-ce pour autant une victoire ? « Elle serait à l'image de la popularité de Nicolas Sarkozy : solide chez ses partisans, mais qui a du mal à s'élargir. C'est suffisant pour un scrutin à un tour comme les européennes, mais cela pose des problèmes de réserves de voix dans des votes à deux tours », estime Jean-François Doridot, directeur général d'Ipsos. Message reçu pour les régionales 2010 et la présidentielle 2012.

► **PS** Avec 21 % d'intentions de vote, les socialistes dépassent à peine le niveau



clé des 20 %. Pas sûr même, selon les sondeurs, que le PS y parvienne. Un mauvais score aurait des conséquences lourdes dès dimanche soir : une relance illico des divisions internes jamais éteintes dans le parti, un allié Verts de plus en plus gourmand, et le risque que

le PS ne soit plus majoritaire à terme au sein de la gauche.

► **La 3^e place** Impossible de départager le MoDem et Europe Ecologie (11 % chacun). « La dynamique de fin de campagne profite aux Verts, qui bénéficient d'une image de vote protestataire plus utile que

■ EXTRÊME DROITE

Les listes FN et Libertas de Philippe de Villiers gravitent autour de 6 %, alors que ce camp a toujours profité par le passé du scrutin européen. « Leur thème classique du protectionnisme a été repris par la plupart des autres partis, de gauche comme de droite », explique Ipsos.

le NPA », estime Jean-François Doridot. Par sa campagne très antisarkozyste, le MoDem pourrait avoir crispé son électorat de droite et celui, europhile, de l'ex-UDF, au profit de l'UMP et des Verts.

► **La gauche radicale** Même situation du côté du NPA et du Front de gauche, au coude à coude autour de 6 %. Un score en dessous des espoirs de début de campagne portés par le discours sur la crise. « Le NPA, qui touche un électorat jeune, donc volatil, n'a pas réussi à convaincre d'un vote utile ; le Front de gauche a marqué des points, profitant de l'électorat communiste, structuré, âgé et mobilisé », estime Jean-François Doridot. ■

Revue de presse

Développement

PE/PME Ardan réveille les projets des petites entreprises

ans les petites entreprises, le dirigeant cumule souvent plusieurs casquettes : responsable financier, de la production, des ressources humaines, responsable commercial, marketing et juridique, etc. Pour faire grandir l'entreprise - qu'il s'agisse de lancer un nouveau produit, de démarcher de nouveaux clients... il lui faut souvent se résoudre à confier ce projet à une personne dédiée. Une étape difficile à franchir, qui fait souvent hésiter les patrons de petites entreprises : ce nouveau salarié va-t-il me permettre de faire augmenter le chiffre d'affaires de l'entreprise ? Comment s'intégrera-t-il ?

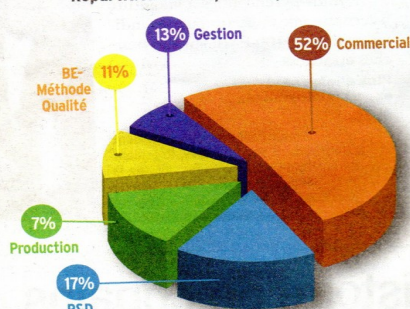
Pour aider les chefs d'entreprises à sauter le pas sans mettre en péril leur entreprise, il existe le dispositif ARDAN Développeur, mis en place par le CNAM (Centre National des Arts et Métiers) et soutenu financièrement par la Région Midi-Pyrénées. Le principe est simple. Pendant une période de 6 mois, le chef d'entreprise accueille un demandeur d'emploi ayant les compétences requises pour mener à bien son projet. Cela permet à la fois à celui-ci d'avoir le temps de faire ses preuves et au chef d'entreprise de s'assurer que son projet est viable.

325 entreprises de la région ont bénéficié du dispositif depuis 2007.

Pendant ce stage, le demandeur d'emploi bénéficie également d'une formation qualifiante tout en conservant ses droits et ses indemnités chômage.

La mise en place d'ARDAN Développeur coûte au total 5 000 euros à l'entreprise, la Région prenant en charge le solde (5 500 euros). Selon une étude réalisée par le CNAM Midi-Pyrénées en février dernier, 80 % des demandeurs d'emploi sont embauchés par l'entreprise à l'issue de ce « stage » de 6 mois dont 64 % en CDI. Ces embauches entraînent également d'autres, en

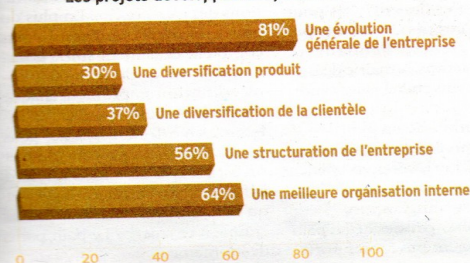
Répartition des dispositifs par fonction



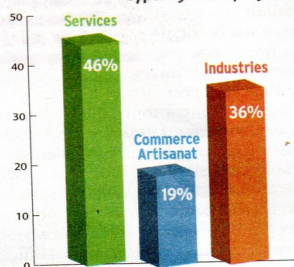
moyenne 1,3 emploi est créé en plus de celui du pilote du projet. Autre particularité du dispositif, il permet de valoriser l'expérience des demandeurs d'emploi : un tiers des bénéficiaires du dispositif ont plus de 50 ans. Du côté de l'entreprise, les bénéfices sont également visibles. 40 % des chefs d'entreprise déclarent avoir vu leur chiffre d'affaires augmenter et 43 % d'entre eux en ont profité pour investir dans de nouveaux projets. Depuis janvier 2007, 325 entreprises (86 % d'entreprises de moins de 20 salariés) de Midi-Pyrénées ont bénéficié du dispositif Ardan.

Plus d'infos sur www.ipst-cnam.fr/Dipostif-Ardan-Developpeur

Les projets développés ont permis...



Typologie des projets



CONCOURS

34^e édition des Inn'Ovations

Votre entreprise innove en Midi-Pyrénées ? Faites-le savoir et participez à la 34^e édition des Inn'Ovations ! Vous avez jusqu'au 18 octobre prochain pour déposer votre candidature.

Inn'Ovations
LE CONCOURS DE L'INNOVATION
EN MIDI-PYRÉNÉES

Organisé par MPI, l'agence régionale de l'innovation et financé par la Région Midi-Pyrénées en partenariat avec Airbus, EDF et BNP Paribas, ce concours a pour objectif de récompenser et de soutenir les entreprises régionales innovantes. Histoire de ne pas passer à côté d'une pépite et de saluer tous les talents, le concours doté de 190 000 euros de prix comporte six catégories* dont trois nouvelles ainsi que deux prix spéciaux. Pour chacun de ces prix, le jury délibérera en tenant compte du caractère de l'innovation, de son originalité ou encore du champ d'application qu'elle couvre. Il prendra en considération l'existence d'une première démonstration, de la stratégie d'accès au marché mise en place, des retombées économiques envisagées, des passerelles de coopération induites entre les milieux dits « académiques » et économiques et enfin de la pertinence de la candidature par rapport à la catégorie dans laquelle l'entreprise s'est inscrite. La sélection aura lieu du 17 au 25 novembre 2014. Quant à la cérémonie de remise des trophées, celle-ci se déroulera le 29 janvier 2015, devant un large public, en clôture du salon Midinnov organisé chaque année par MPI pour la Région Midi-Pyrénées.

*Innovation, produits et services du futur/ Innovation, croissance et développement territorial/ Innovation et société/ Innovation et formation/ Innovation et international/ Innovation et jeunes entreprises/ Grand prix/ Prix coup de cœur.

Pour plus d'info, rendez-vous sur : www.mp-i.fr

Représentations graphiques

Données de type « effectif »

	X	Y
A	30	25
B	15	15
C	30	20
D	20	30
E	25	35

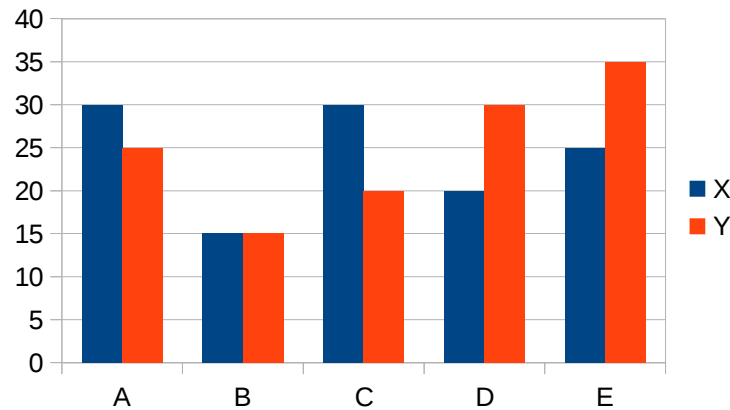
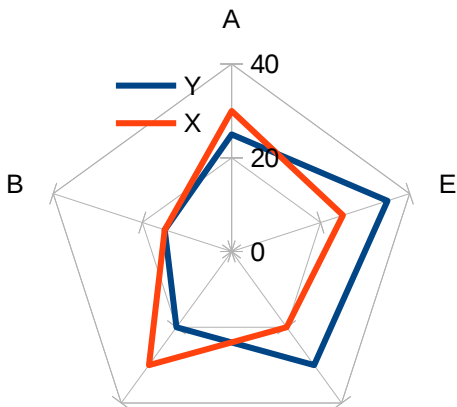
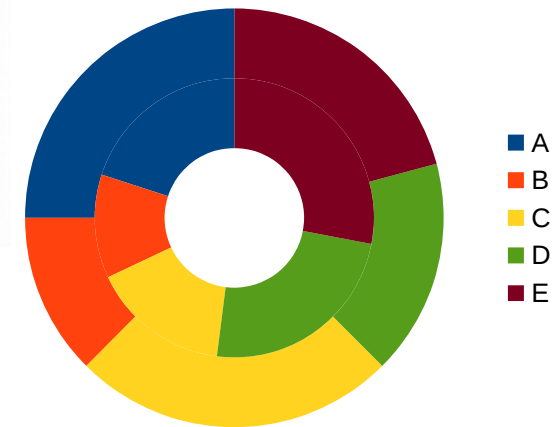


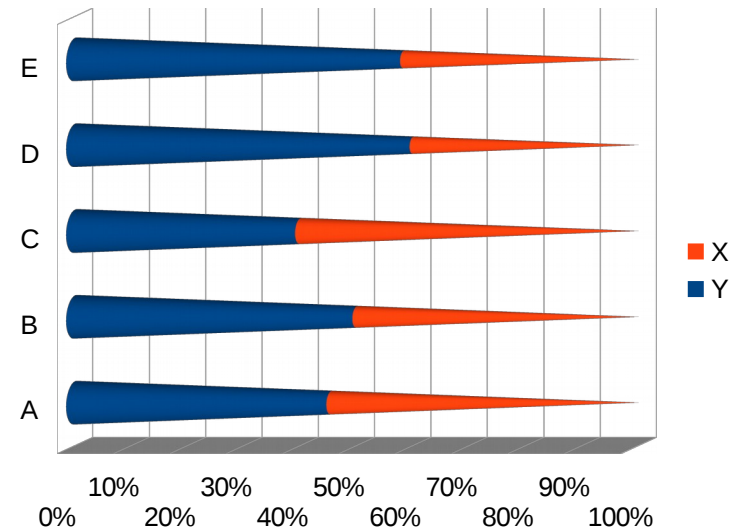
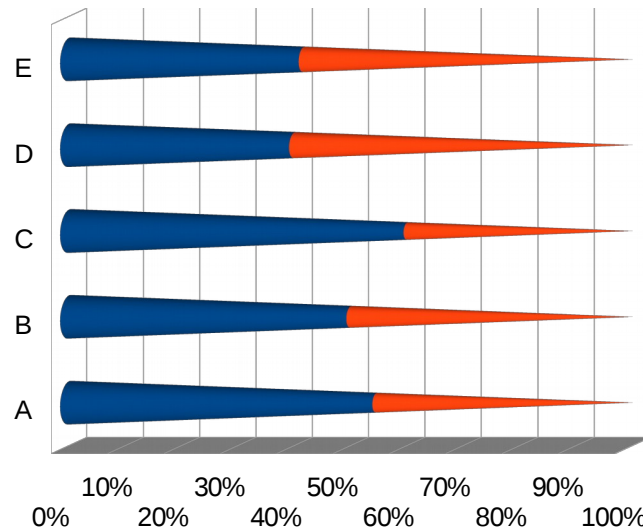
Diagramme en bâtons



Diagramme « tore »



Cônes 3D (!?)



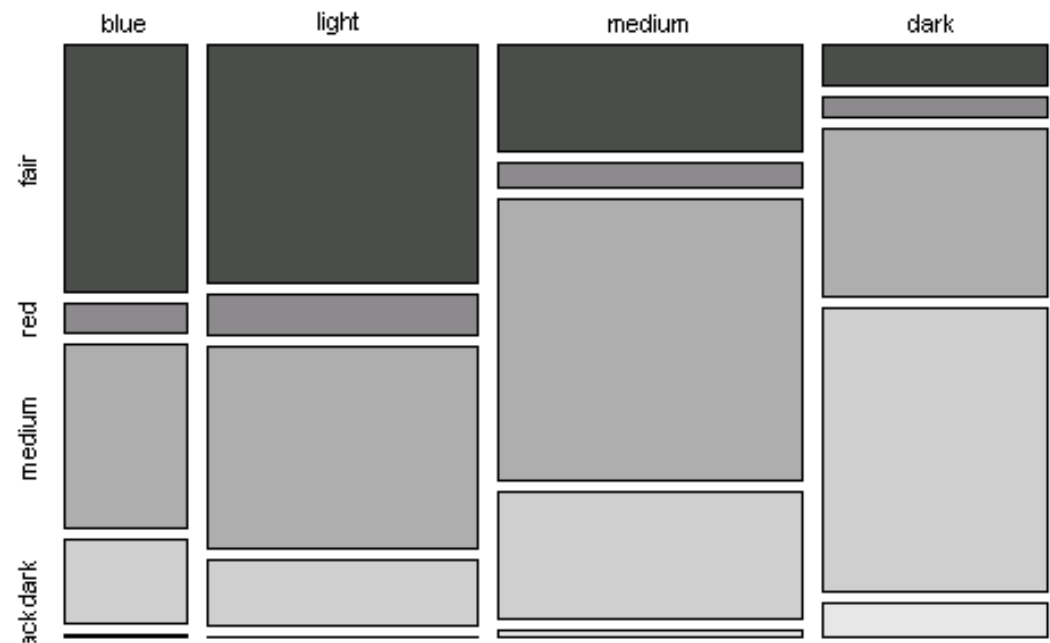
Représentations graphiques

Données de type « effectif » - Table de contingence

(données *caith*, R)

	fair	red	medium	dark	black
blue	326	38	241	110	3
light	688	116	584	188	4
medium	343	84	909	412	26
dark	98	48	403	681	85

Diagramme mosaïque



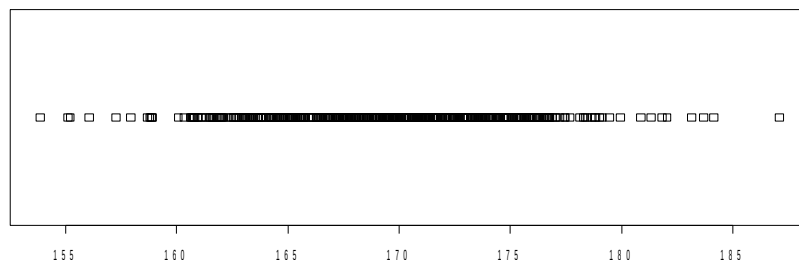
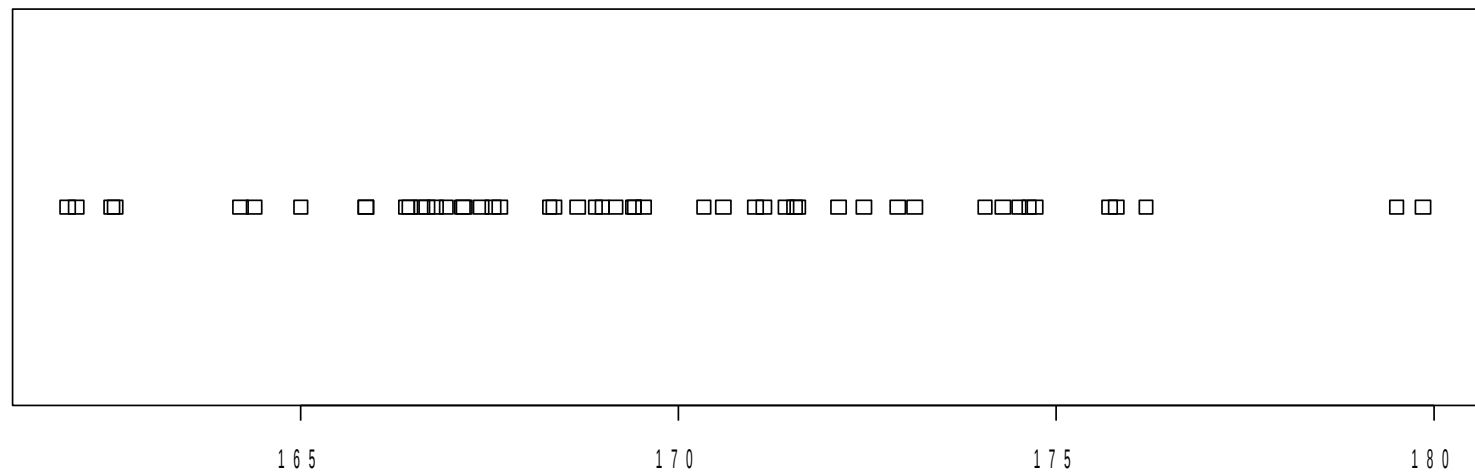
Représentations graphiques

Données quantitatives

50 observations : taille en cm de 50 individus

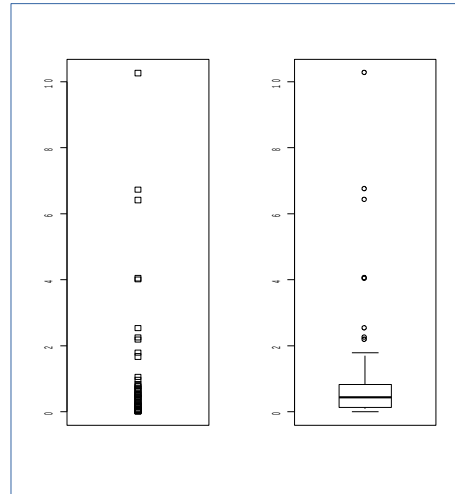
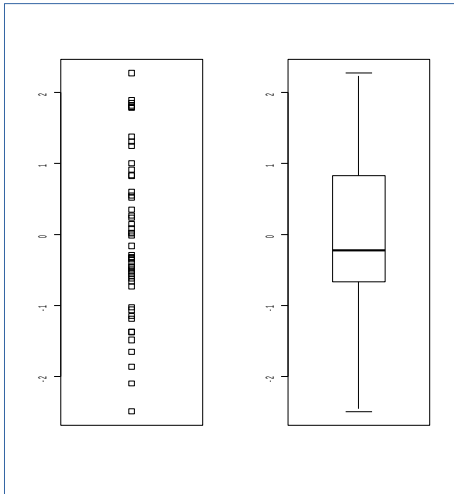
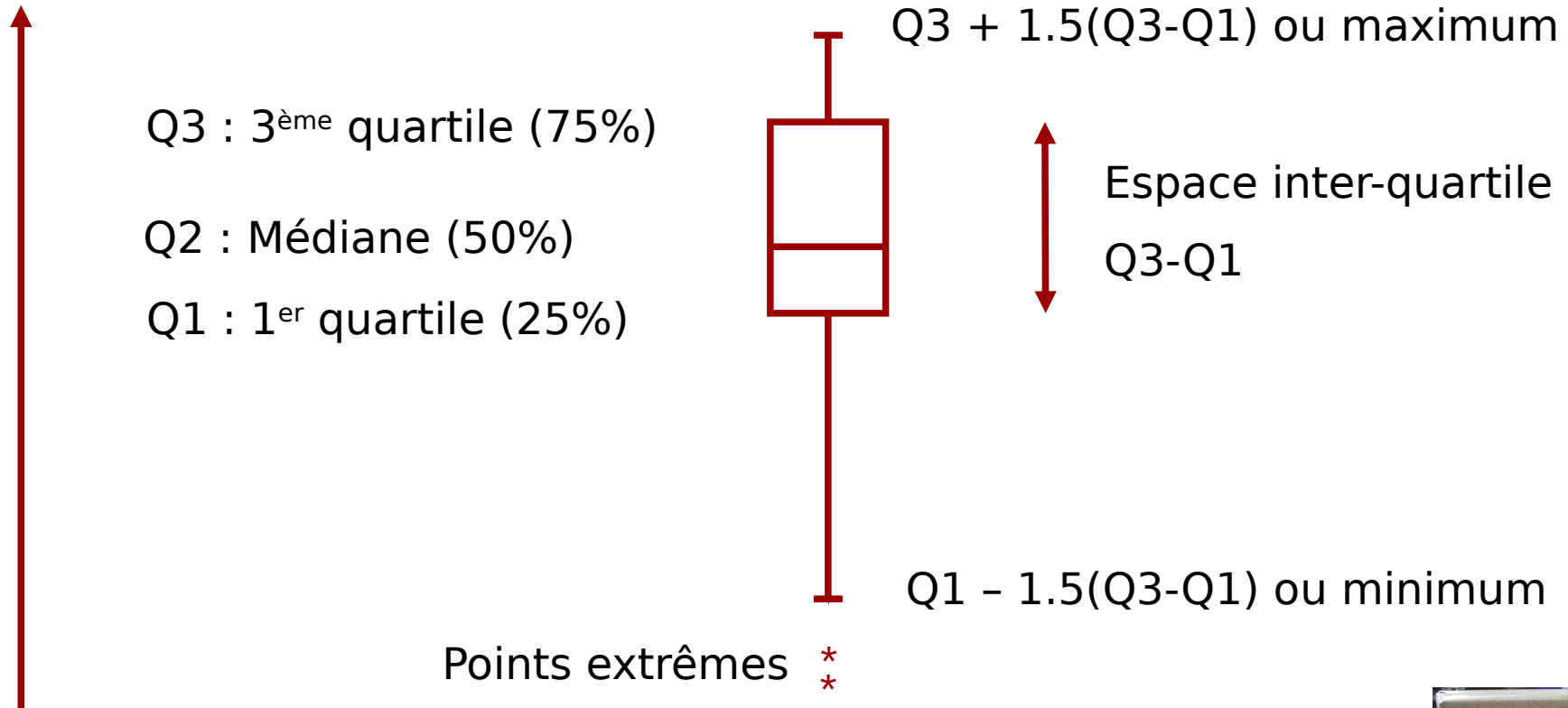
```
171.4 175.8 168.9 166.8 174.3 169.2 165.9 171.6 167.1 179.8
174.5 169.0 170.6 173.1 169.4 167.4 162.5 172.9 174.7 164.2
175.7 167.2 169.5 174.1 162.5 172.4 166.4 162.0 167.5 168.7
166.9 171.0 176.2 172.1 166.4 168.3 170.3 164.4 167.6 168.3
165.9 174.6 171.5 169.4 166.6 179.5 166.7 171.1 161.9 165.0
```

Nuage de points 1D
(*stripchart*)



Avec 500 observations, la lisibilité devient délicate.

Boxplot



Boxplot : on
dit aussi
boîte à
moustaches

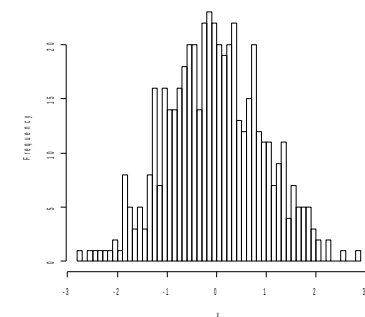
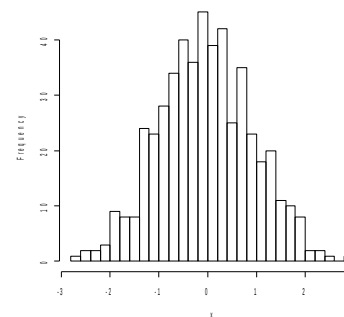
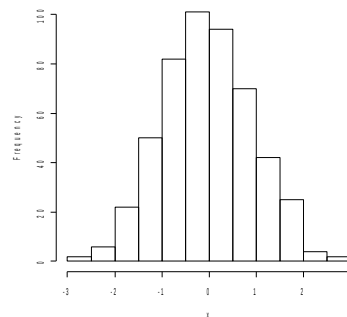
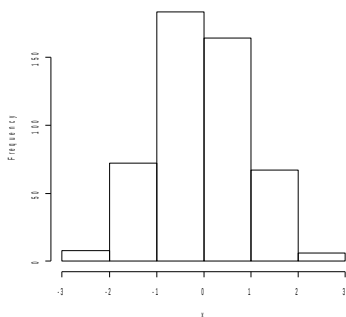
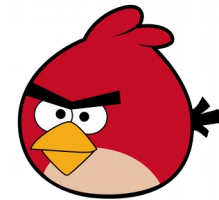


Source : ML 

Histogramme

≠

diagramme en bâtons !



Histogramme

~

Densité de
probabilité

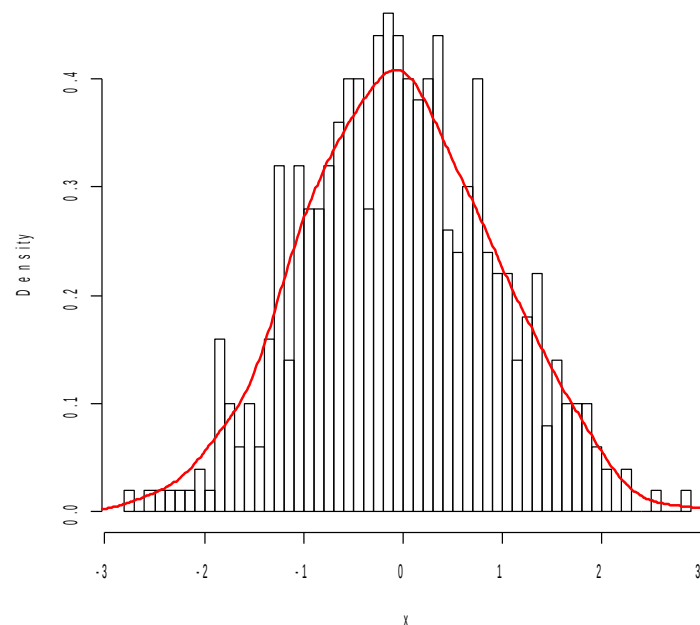
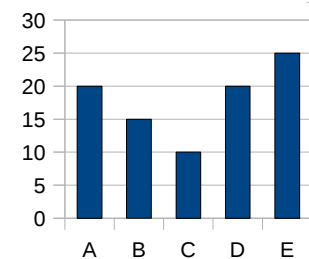
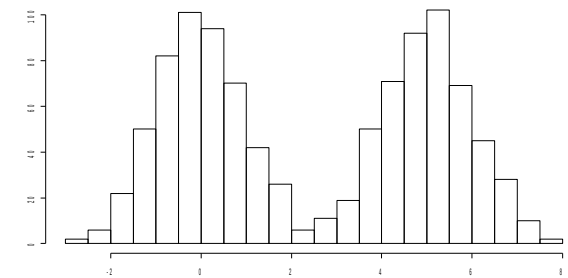
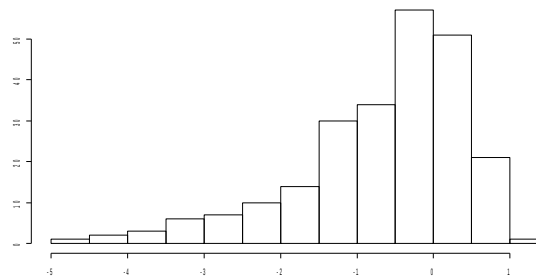
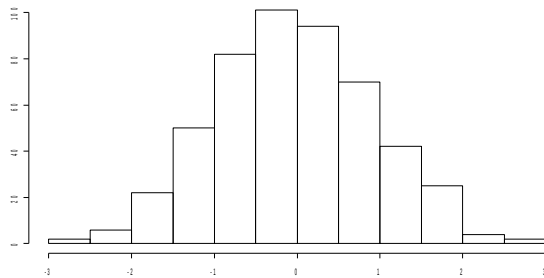
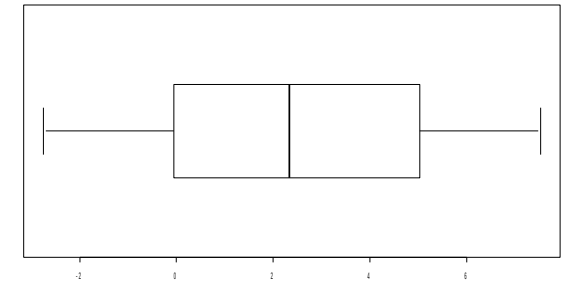
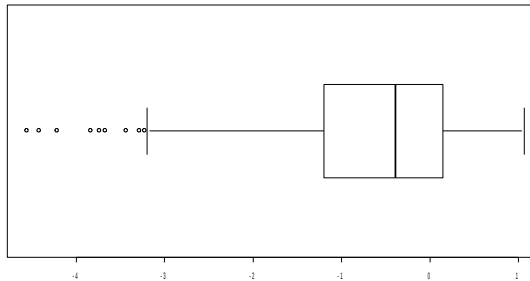
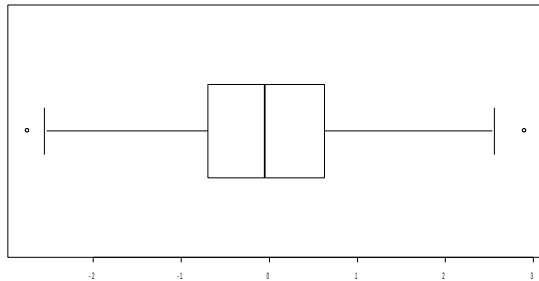


Diagramme
en bâtons



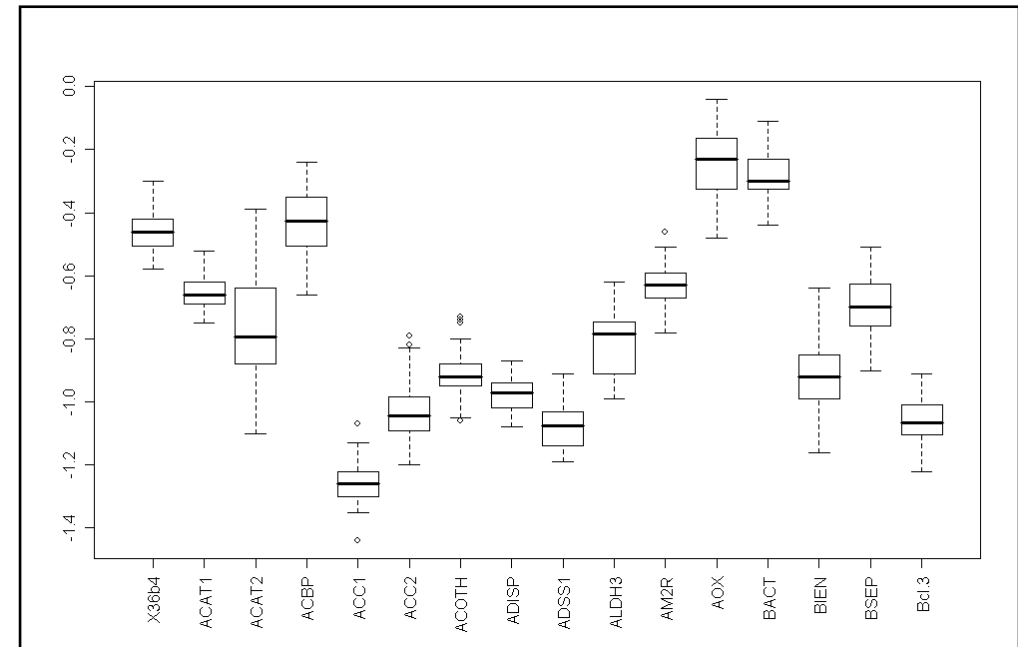
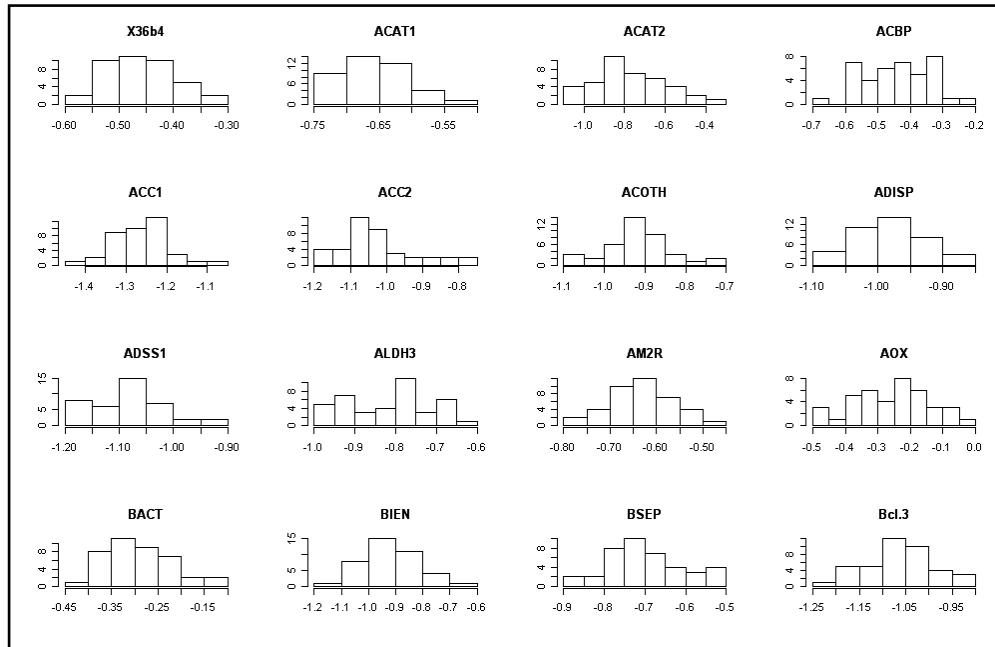
Boxplot et histogramme



Individuellement, ces graphiques apportent la même information sur la position et la répartition des données. Attention cependant aux cas de multi-modalités que le boxplot ne peut pas capter.

⇒ *avantage histogramme* 

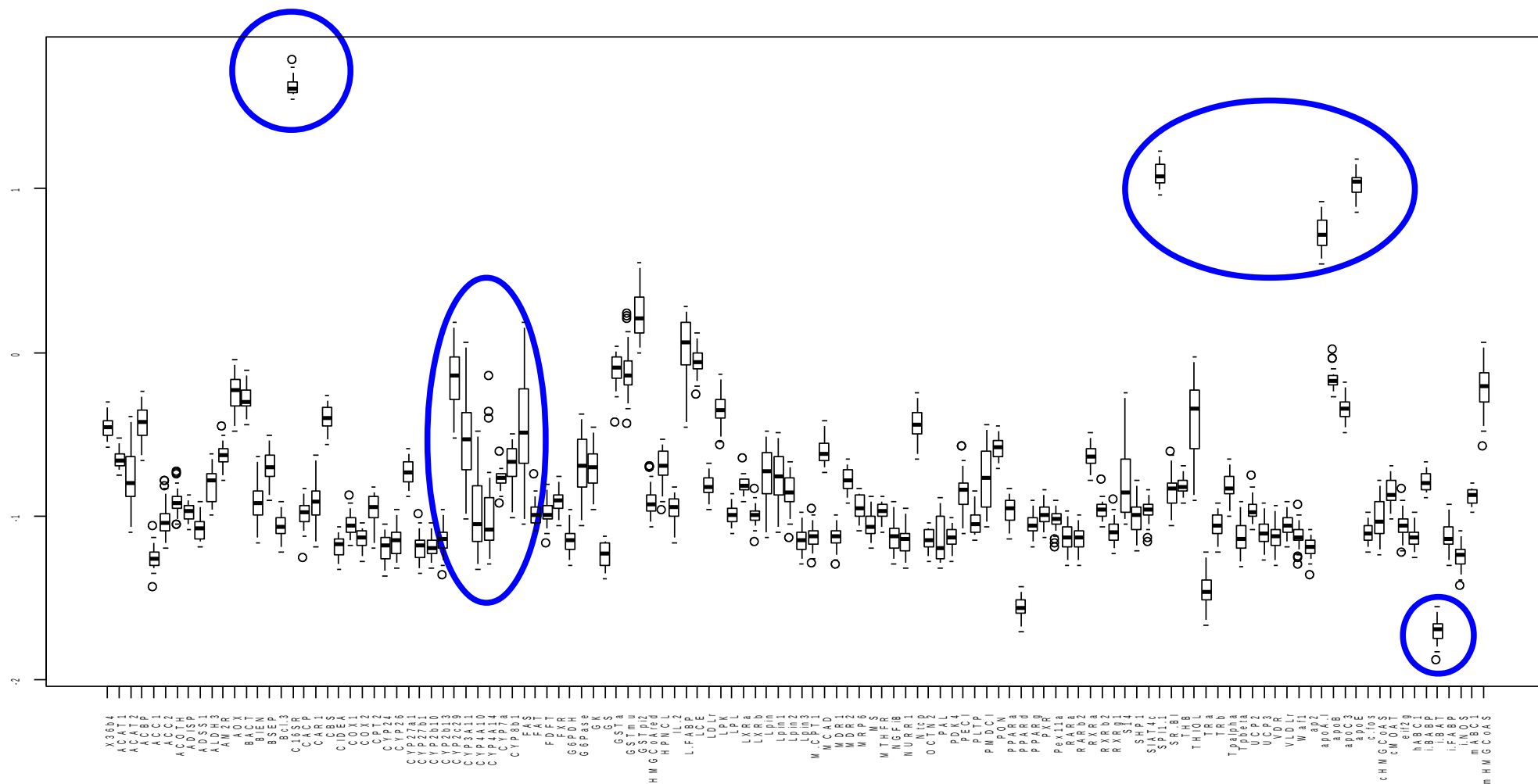
Boxplot et histogramme



⇒ égalité ?

Interpréter des *boxplot*

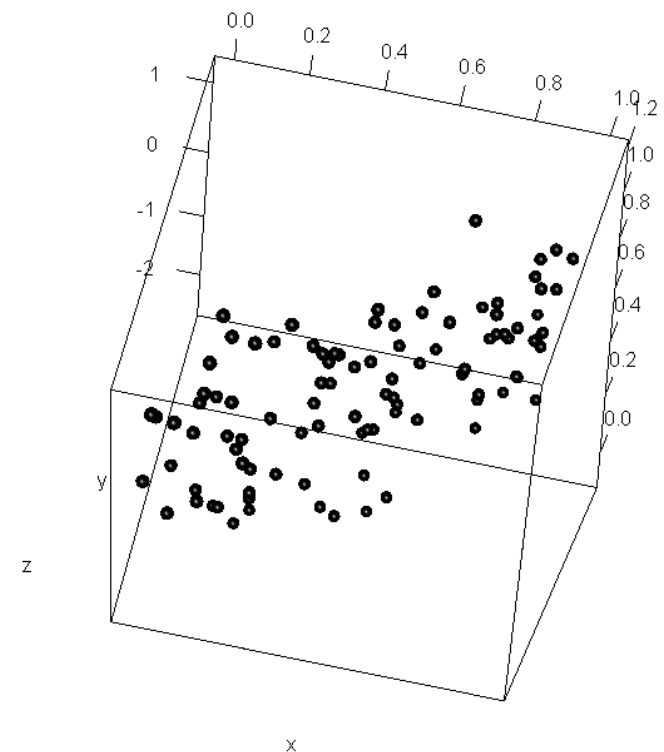
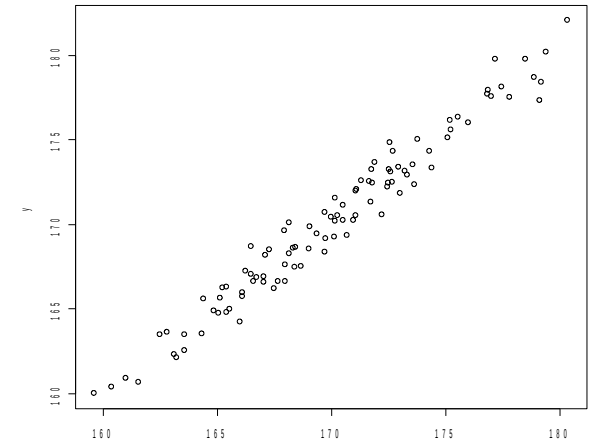
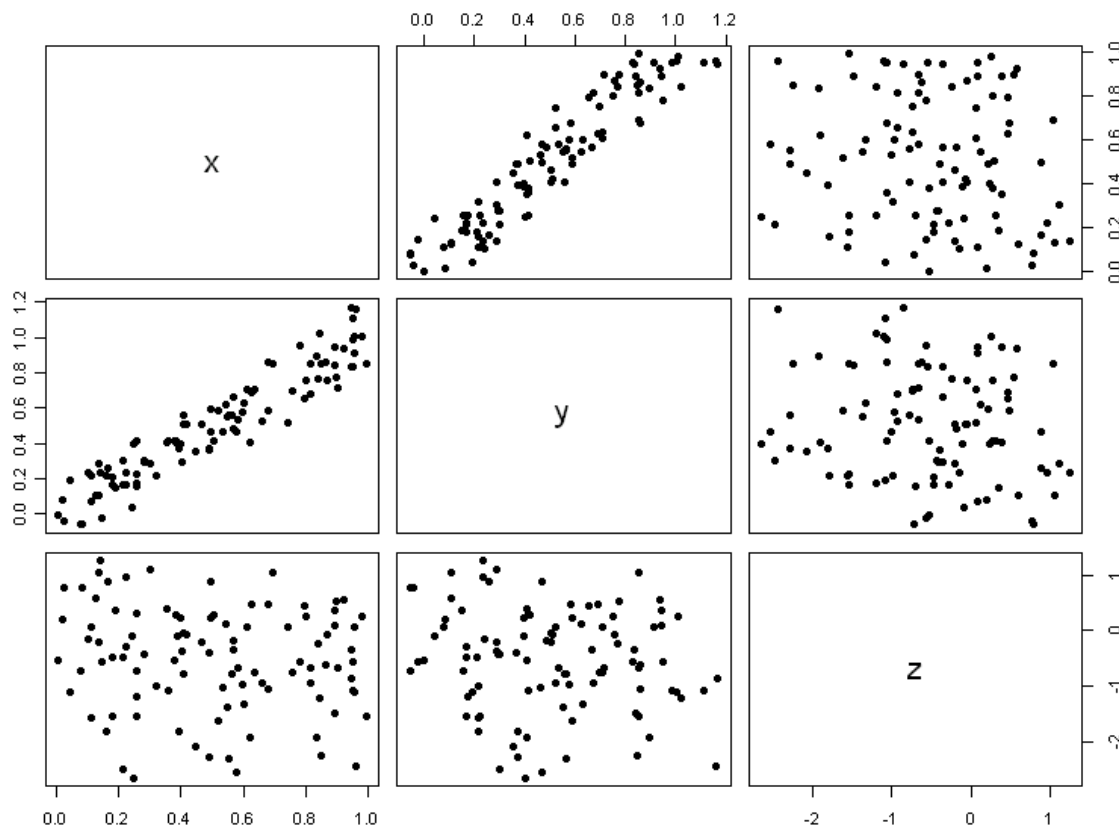
120 boxplots correspondant à l'expression de gènes mesurée sur **40** souris (**2** génotypes $\times$ **5** régimes alimentaires $\times$ **4** réplicats).



Représentations graphiques

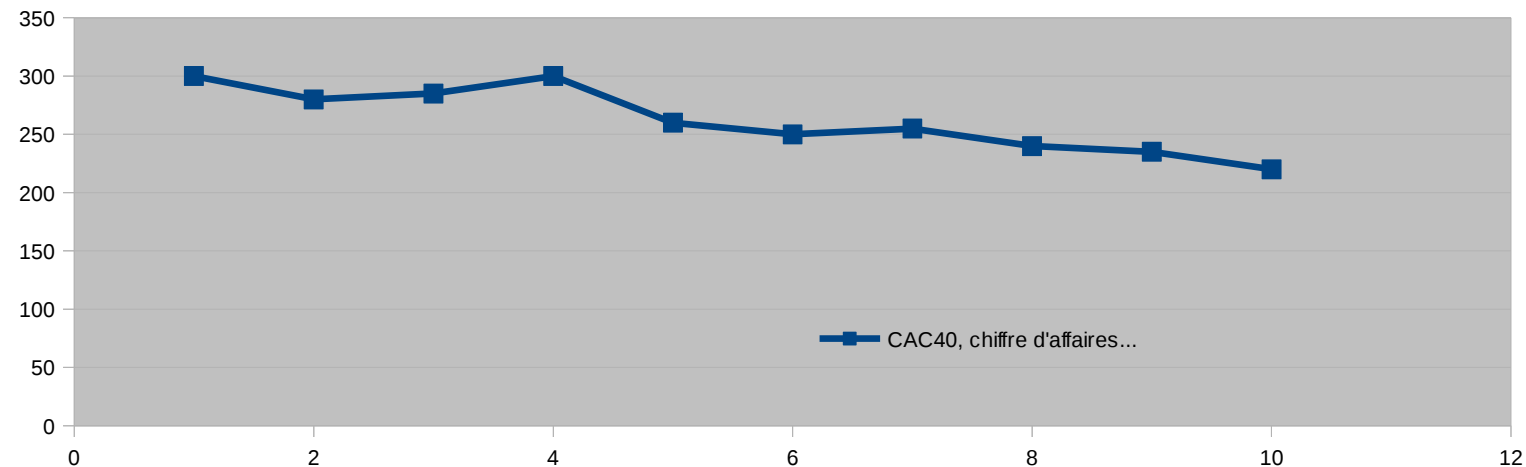
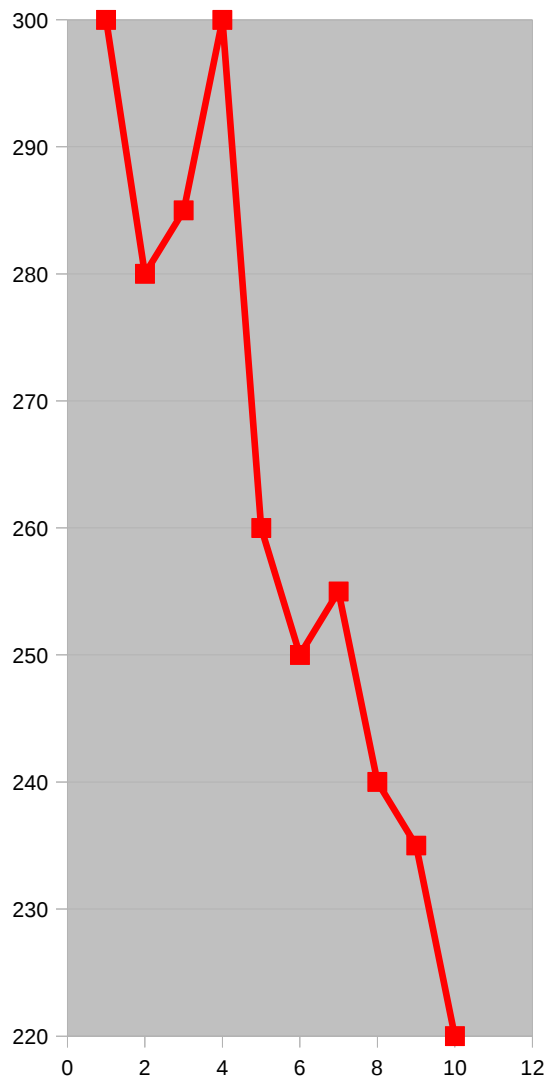
2 variables ou plus

Nuage de points, diagramme de dispersion (*scatterplot*)



Échelles des axes (1)

Temps	1	2	3	4	5	6	7	8	9	10
Chômage, CAC40, chiffre d'affaires...	300	280	285	300	260	250	255	240	235	220

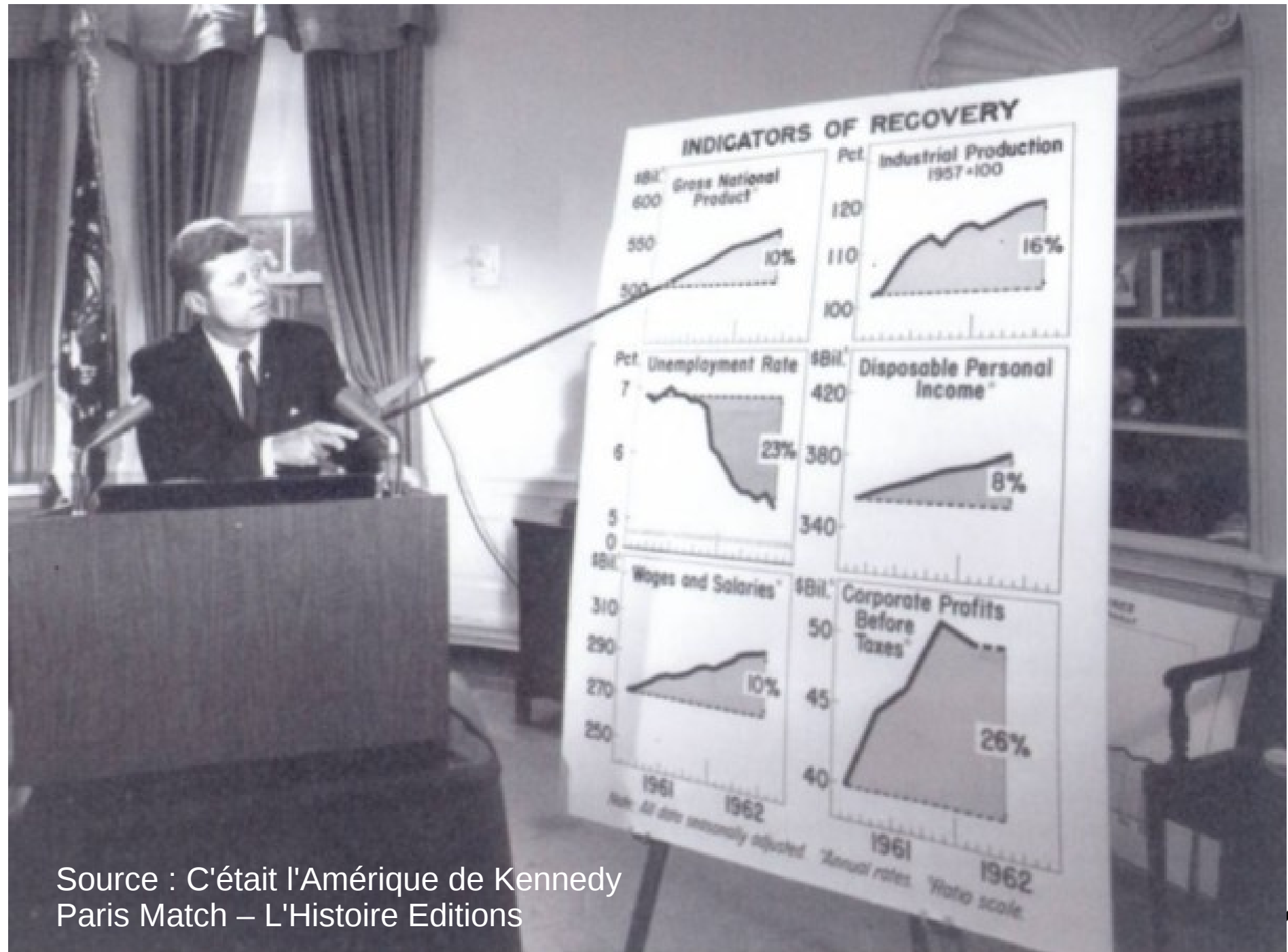


Pas de panique,
léger fléchissement...

C'est la crise !!!
effondrement...



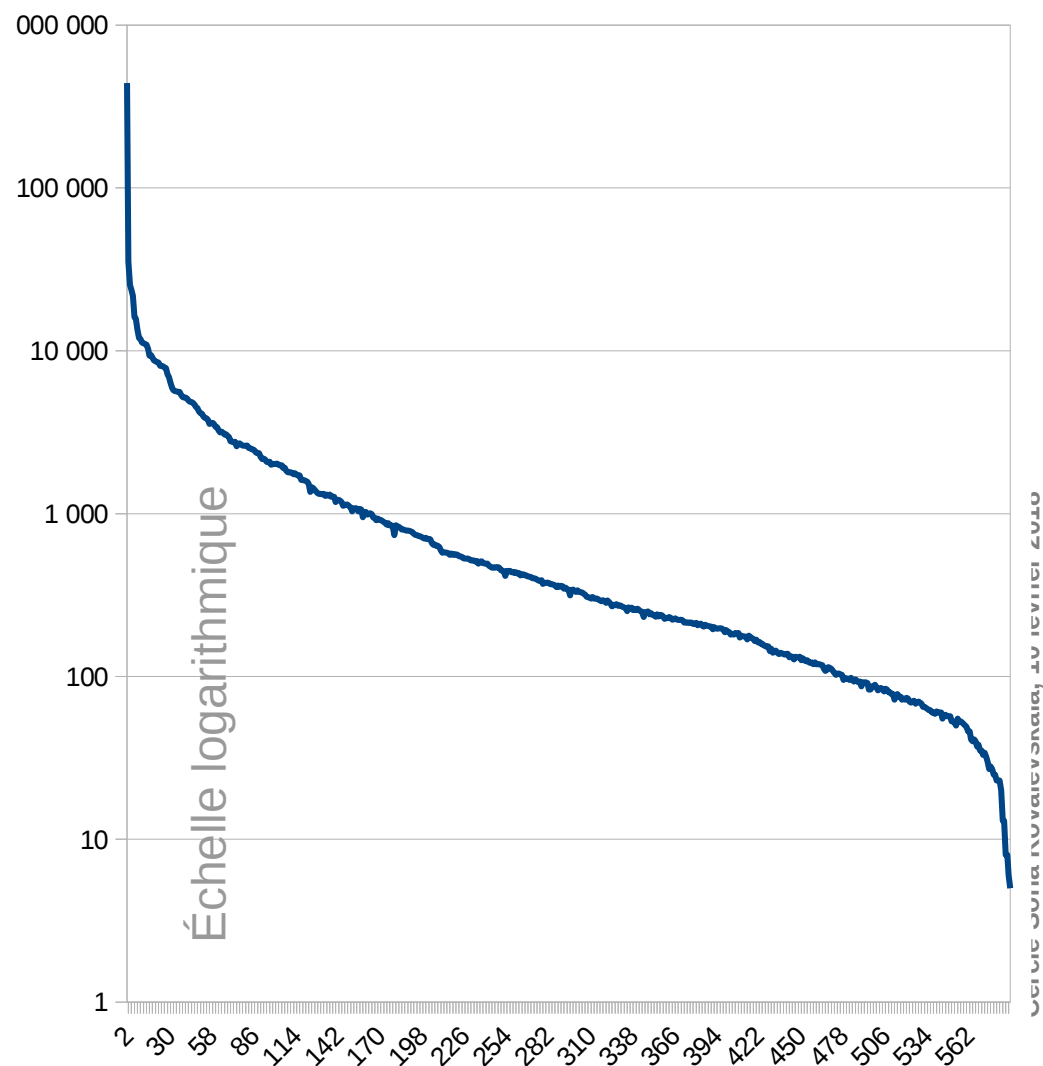
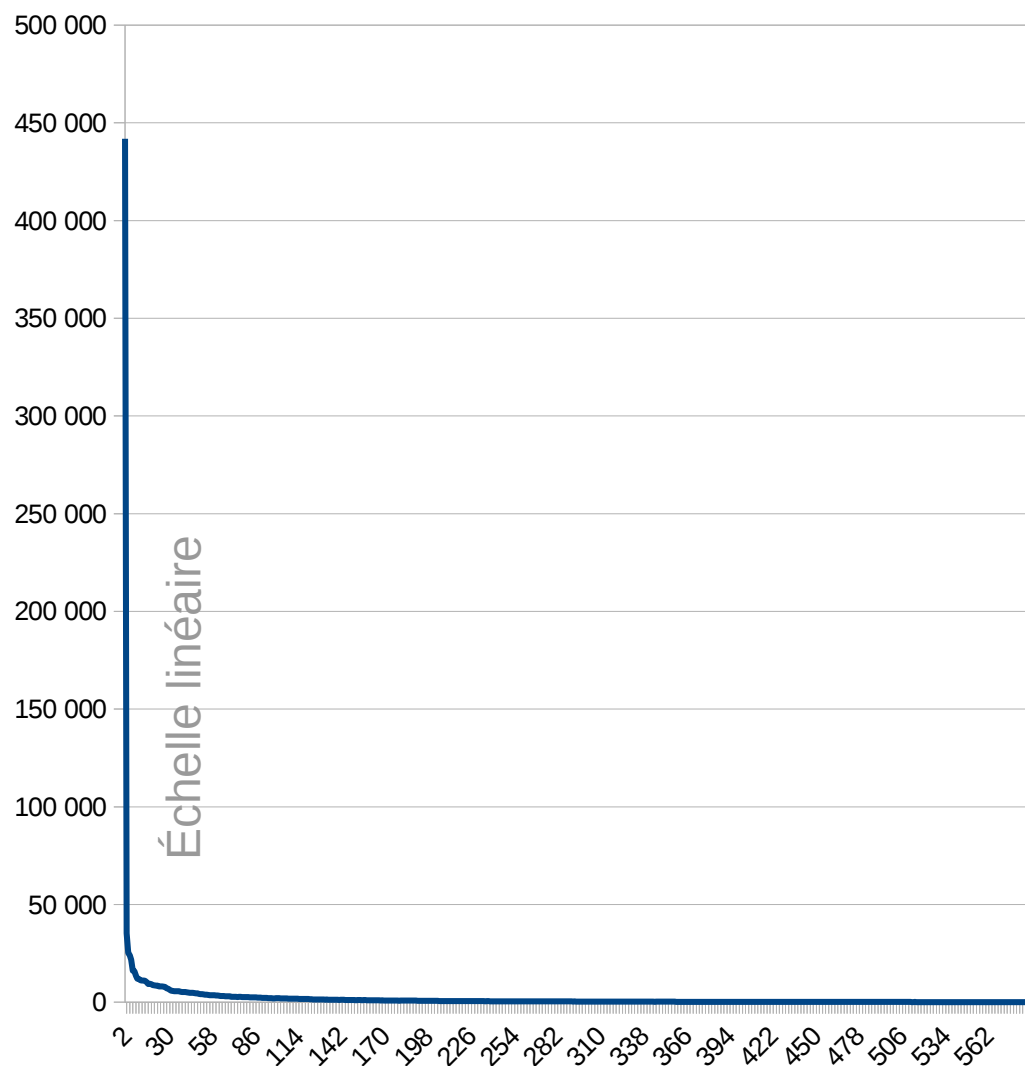
Échelle des axes



Source : C'était l'Amérique de Kennedy
Paris Match – L'Histoire Editions

Échelles des axes (2)

Données : Population des 589 communes de Haute-Garonne au 1^{er} janvier 2010, source INSEE, *Recensement de la population 2010*



Lie factor

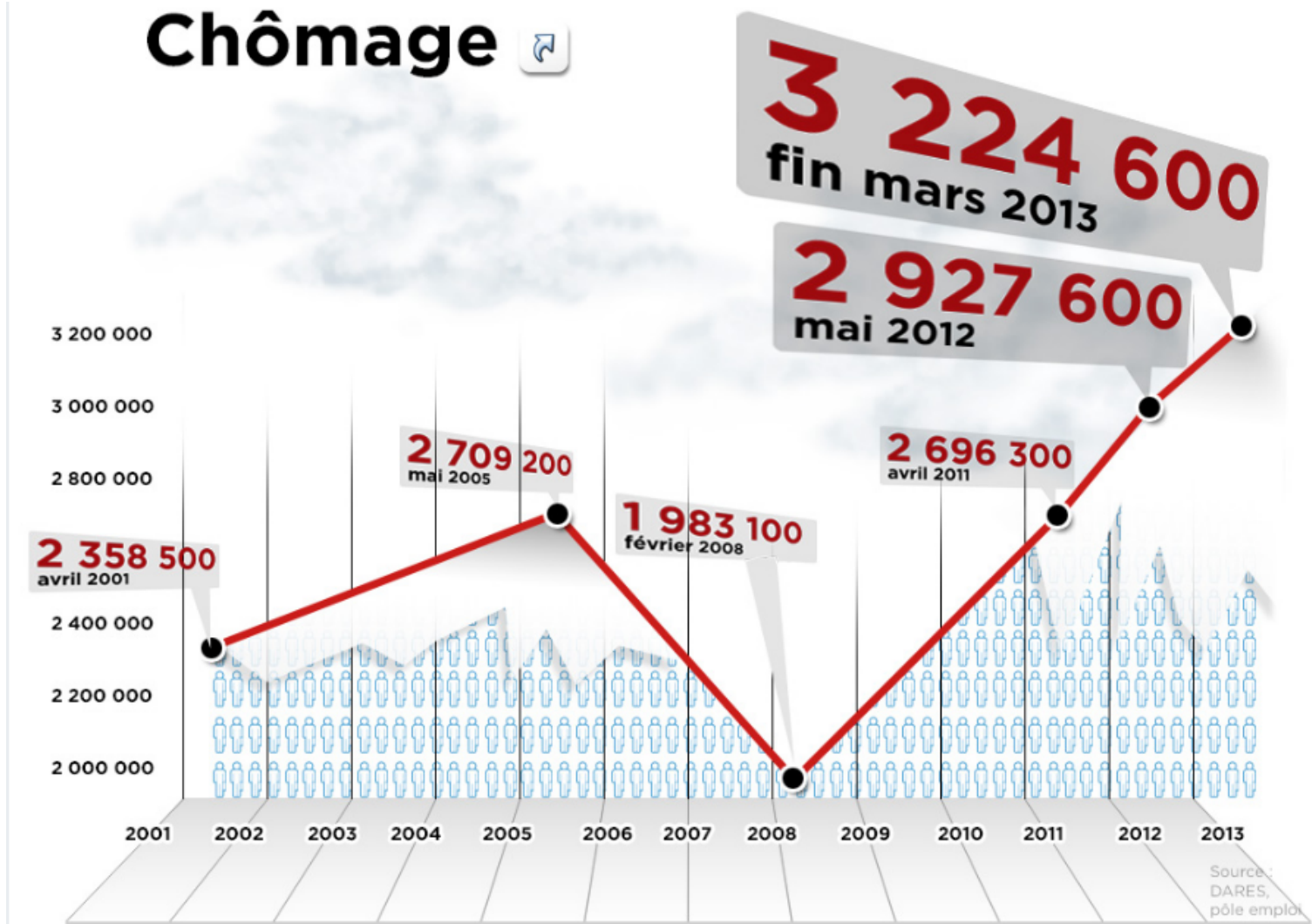
Taille de l'effet montré sur le graphique

Taille de l'effet dans les données

Réactions de lecteurs

- vous êtes vous interrogé sur le contenu de votre infographie ? sur les polices de caractères utilisées ? Sur les zooms ? Où est l'objectivité et le factuel ?

- Merci de corriger votre 1er graphique, qui fait croire à une augmentation du chômage plus lente sur la dernière période, après une augmentation plus rapide sur la période précédente, ce qui est l'EXACT CONTRAIRE de la réalité : Février 08 à Avril 11 (38 mois) : $2696300 - 1983100 = 713200 / 38 \text{ mois} = 18768$ par mois. Avril 11 à Mai 12 (13 mois) : $2927600 - 2696300 = 231300 / 13 \text{ mois} = 17792$ par mois (976 de -). Mai 12 à Mars 13 (10 mois) : $3224600 - 2927600 = 297000 / 10 \text{ mois} = 29700$ par mois (11908 de +).



Cercle Sofia Kovalevskaja, 10 février 2018

Une année de présidence Hollande en chiffres

http://www.lemonde.fr/politique/visuel/2013/05/06/une-annee-de-presidence-hollande-en-chiffres_3170215_823448.html

