Chapitre 5
Diagrammes et graphiques

Dernière mise à jour le 27 septembre 2009

« *L'information graphique est à l'œil ce que la poésie est au langage* » Théodore MITCHELL

« *Il existe trois catégories de personnes : celles qui voient, celles qui voient quand on leur montre, celles qui ne voient jamais*. » Léonardo da VINCI (1452-1519)

Introduction

1 – Un beau graphique vaut mieux qu'un long discours
- A – Utilité des graphiques et des diagrammes
  1) Renforcer l'efficacité de la communication d'un message
  2) Faciliter la prise de décision
- B – Qualités d'un bon graphique
- C – Quartet d'ANSCOMBE

2 – Les échelles graphiques
- A – Echelles numériques
- B – Echelles de catégories
  1) catégories numériques
  2) catégories nominales
- C – Echelles ordinales
- D – Echelles verticales doubles
- E – Echelles logarithmiques
  1) Définition
  2) Calcul pratique du log décimal d'un nombre
  3) Rappels sur le log décimal
  4) Exemples
     a) L'échelle log pour mieux voir les différences de progression
     b) L'échelle log linéarise les évolutions à taux constant

3 – Diagrammes statistiques
- A - Pictogramme
- B - Cartogramme
- C – Diagramme de GANTT

4 – Graphiques usuels
- A - Graphique en barres
  - 1) Barres verticales
    - a) Simples
    - b) Multiples
    - c) Tronçonnées
  - 2) Barres horizontales
    - a) Simples
    - b) Multiples
    - c) Tronçonnées
  - 3) Graphique en bâtons
- B – Courbes et aires
  - 1) Courbe simple
  - 2) Courbes multiples
  - 3) Aires délimitées par des courbes
- C - Graphique de dispersion ou nuage de points

D - Secteurs
- 1) Secteurs à 360 degrés
- 2) Secteurs à 180 degrés
- 3) Anneaux
  - a) Simples
  - b) concentriques

5 – Autres graphiques
- A – Graphiques en radar et toiles d'araignée
  - 1) Radar
  - 2) Toile d'araignée
- B – Graphique à bulles
- C – Graphiques boursiers
- D – Graphiques de TUKEY
  - 1) Les éléments constitutifs du graphique original
  - 2) Exemple
  - 3) Graphique de TUKEY et forme de la distribution sous-jacente
    - a) Aplatissement
    - b) Etalement
  - 4) Le graphique de TUKEY simplifié
- E – Graphiques panachés
  - 1) Secteur complété par une barre tronçonnée
  - 2) Graphique de PARETO
    - a) De la loi de PARETO au graphique de PARETO
    - b) Définition, construction, exemple et interprétation
    - c) Interprétation

F – Histogramme
1) amplitudes de classes identiques
a) Histogramme d'effectifs
b) Histogramme de fréquences
2) Amplitudes de classes différentes
a) Histogramme d'effectifs
b) Histogramme de fréquences
G – Pyramide des âges
H – Graphique en cascade
I – Autres graphiques en forme d'aires
J – Graphique à trois dimensions
1) Graphique en 2D avec ajout de « profondeur »
2) Graphique en barres avec 3 dimensions réelles

6 - Résumé

Exercices corrigés (PDF) :

TD5 2008 : Sujets - Corrigés
TD5 2009 : Sujets - Corrigés

Voir aussi les sujets d'examens corrigés (lien vers page HTML)

Exercices spécifiques (PDF) :

Diagramme de GANTT : énoncé - corrigé
Graphique de PARETO : énoncé - corrigé
Graphique de TUKEY: énoncé - corrigé
Histogramme : énoncé - corrigé
Graphique à bulles: énoncé - corrigé
Graphique panaché (secteur puis barre) : énoncé - corrigé
Graphique avec échelle semi-logarithmique: énoncé - corrigé

0 – Introduction

Les **diagrammes** et les **graphiques** sont des représentations visuelles simplifiées, mais néanmoins rigoureuses, de chiffres, de concepts, de catégories, d'images, etc. Bien que certains diagrammes représentent aussi des chiffres (voir la section 3 de ce chapitre), c'est principalement au graphique que cette fonction incombe. Plus précisément, en statistique descriptive, le graphique sert à représenter visuellement des informations sur les dimensions d'une population d'unités statistiques.

Le graphique statistique combine dans des proportions variées des points, des lignes, un système de coordonnées généralement cartésiennes, des chiffres, des symboles, des mots, des formes et des couleurs.

L'idée de représenter visuellement des nombres traduisant eux-mêmes des propriétés quantitatives ou qualitatives de populations statistiques[1] et les relations éventuelles entre ces nombres et entre ces propriétés, est assez récente dans l'histoire de l'humanité. Ainsi, on ne trouve pas à proprement parler de graphiques statistiques dans les cavernes préhistoriques. Selon Edward TUFTE[2], un des meilleurs spécialistes contemporains des graphiques, c'est vers les années 1750-1800 que l'idée de représenter visuellement des données statistiques ou des séries temporelles dans un repère de coordonnées cartésiennes est apparue de façon pérenne[3]. C'est l'anglais William PLAYFAIR (1759-1823) qui fut l'un des pionniers de la représentation graphique de données statistiques jusque-là proposées sous forme de tableaux[4]. On doit aussi mentionner le rôle joué par Florence NIGHTINGALE (1820-1910), infirmière anglaise, qui fut la première à réaliser des graphiques, suivant les travaux de son père, lui-même pionnier de l'épidémiologie.

[1] Nous avons vu dans le chapitre 1 que l'expression « population statistique » recouvrait aussi bien des entités humaines que des collections d'objets matériels, des entités abstraites, des territoires, etc.

[2] TUFTE, Edward (2001), The Visual Display of Quantitative Information, Graphics Press, page 13. Voir aussi le site internet de Edward TUFTE. Les premières représentations graphiques sont cependant plus anciennes. Edward TUFTE montre dans son livre un graphique datant de 950 à 1000 après J.-C., d'un auteur inconnu, qui illustre l'inclinaison de l'orbite des planètes en fonction du temps. Ce graphique, qui fait partie d'un ensemble de parchemins destiné à la formation des élites religieuses de l'Occident est un précurseur isolé de la représentation graphique des séries chronologiques qui n'apparaîtront en fait qu'au 18ème siècle. Pour voir ce graphique, se rapporter à Gray FUNKHOUSER, « A Note on a Tenth Century Graph, » Osiris, 1 (January 1936), 260-262 ou charger l'URL suivante : http://www.fi.uu.nl/wiskrant/artikelen/hist_grafieken/begin/images/planeten.gif

[3] C'est cependant semble-t-il à Nicole ORESME (1323-1382), évêque de Lisieux, que l'on doit l'idée de représenter une variable en fonction d'une autre sous forme de graphiques en barres verticales. Voir ORESME, Nicole (1482), Tractatus de latitudinibus formarum. Padova. BL: IA 3Q024. Et aussi ORESME, Nicole (1968). *Nicole ORESME and the Medieval Geometry of Qualities and Motions: A Treatise on the Uniformity and Difformity Known as Tractatus de Configrationibus Qualitatum et Motuum*. Madison WI: University of Wisconsin Press. Tr.: M. Clagget. Ou, pour voir le scan de ces premiers graphiques : http://www.math.yorku.ca/SCS/Gallery/images/oresme6.gif. Plus tard, il convient de mentionner l'un des premiers graphiques, décrivant le lien entre l'âge d'une personne et le nombre médian d'années lui restant à vivre, proposé par le hollandais Christian HUYGENS (1629 - 1695). L'histoire des graphiques est retracée sur le site de Michael FRIENDLY et Daniel J. DENIS, http://www.math.yorku.ca/SCS/Gallery/milestone/index.html . En fait, l'histoire des graphiques se confond à ses origines avec l'invention de l'écriture, du calcul, du dessin et de la cartographie.

[4] Voir par exemple ce graphique qui montre l'évolution dans le temps du prix du blé et des salaires, datant de 1786 : http://www.math.yorku.ca/SCS/Gallery/images/playfair-wheat1.gif

Ses graphiques des différentes causes de mortalité des soldats anglais durant la guerre de Crimée (1853-1856) firent sensation dans l'Angleterre victorienne et obligèrent l'administration militaire à améliorer le système de soins aux soldats (c'est un exemple spectaculaire du rôle des graphiques comme outils d'aide à la décision). Il semble cependant établi aujourd'hui que certains de ses graphiques avaient déjà été réalisés par le français André-Michel GUERRY (1802-1866).

1 – Un beau graphique vaut mieux qu'un long discours

A – Utilité des graphiques et des diagrammes

Les graphiques sont des outils qui servent notamment à :

- Renforcer l'efficacité de la communication d'un message
- Faciliter la prise de décision

1) Renforcer l'efficacité de la communication d'un message

Nous sommes quotidiennement bombardés d'informations visuelles, que ce soit par l'intermédiaire des journaux (gratuits ou pas), d'internet, des téléphones portables, des panneaux publicitaires géants ou non, numériques ou en papier. Or nous ne pouvons en absorber qu'une partie limitée. De plus, une partie de cette information est complexe et requiert un degré non négligeable d'attention et de concentration pour être comprise et exploitée utilement.

Les annonceurs et autres diffuseurs d'informations sont conscients de ce défi et c'est pourquoi ils redoublent de créativité et d'imagination pour forcer notre attention et tenter de la captiver suffisamment longtemps pour qu'elle se transforme « magiquement », à notre insu, en concentration, cela pour qu'au bout du compte nous retenions cette information et l'utilisions dans le sens souhaité (achat de produits, certes, mais aussi adoption de certains comportements dans le cas des politiques de santé publique ou de développement durable, entres autres).

Pour ce faire, quand l'information doit prendre la forme de données chiffrées, ils font appel à des **infographistes**. En effet, l'utilisation des graphiques pour transmettre un message s'avère très efficace. Mais le statisticien se doit de rappeler qu'il faut néanmoins respecter certaines règles et ne pas laisser l'esthétique empiéter sur l'exactitude, fut-ce au nom de l'efficacité du message.

2) Faciliter la prise de décision

Dans ce chapitre, nous donnons plusieurs exemples de l'utilisation des graphiques et des diagrammes en tant qu'outil facilitant la prise de décision. Ainsi :

- Le diagramme de GANTT permet d'avoir une représentation visuelle de la planification des différentes tâches d'un projet dans le temps.
- Le graphique à échelle verticale double permet d'apprécier l'existence d'une covariation entre deux phénomènes.
- Le graphique à échelle logarithmique permet d'avoir une idée plus juste des phénomènes en évolution.
- Le graphique de PARETO est utilisé dans les processus de contrôle de qualité.
- Etc.

B – Qualités d'un bon graphique

Selon Edward TUFTE, l'excellence en matière de graphiques statistiques consiste à communiquer avec clarté, précision et efficacité des idées complexes. Ensuite, il énumère neuf caractéristiques d'un « excellent graphique ».

Ainsi, un excellent graphique devrait-il avoir tout ou partie des qualités suivantes :

- Montrer les données.
- Attirer l'attention du lecteur ou de l'auditoire sur l'idée essentielle que le graphique vise à mettre en évidence plutôt que sur les qualités esthétiques du graphique lui-même.
- Eviter de déformer le message contenu dans les chiffres
- Présenter un grand nombre de chiffres dans un espace restreint
- Donner de la cohérence à de vastes ensembles de données
- Faciliter les comparaisons visuelles entre différents chiffres
- Révéler les chiffres à différents niveaux de détails, allant de la vision d'ensemble à une structure plus fine.
- Servir un objectif clair : décrire des données, explorer des données, ou simplement les tabuler ou leur donner un aspect esthétique.
- Etre étroitement corrélé avec la description purement statistique ou verbale des données.

B – Quartet d'ANSCOMBE

Les graphiques révèlent des informations sur la forme des séries et/ou des distributions que les tableaux et les statistiques résumées ne peuvent pas toujours montrer. Une illustration extrême en est le quartet d'ANSCOMBE, nommé ainsi après la publication d'un article sur ce thème par le statisticien anglais Francis ANSCOMBE (1918-2001)[5] .

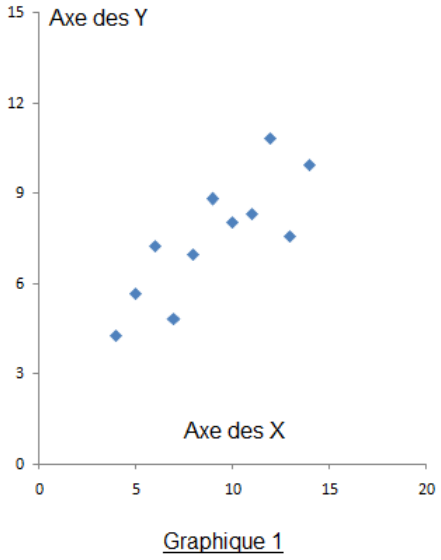
Dans le tableau ci-après les quatre paires de séries différentes de 11 chiffres ont toutes les mêmes statistiques résumées. La moyenne de la variable X est égale à 9 et son écart-type est 1,94 pour les 4 séries. La moyenne des Y est égale à 7,5 et son écart-type à 1,94 pour les 4 séries. Sans les quatre graphiques ci-après, on pourrait déduire de façon erronée que comme les 4 paires de séries ont la même moyenne et la même dispersion (en outre, elles ont le même coefficient de corrélation et la même droite de régression Y = 3 + 0,5 X [sur le calcul de la droite de régression voir le chapitre 6]), elles sont très semblables. Or, les 4 graphiques dits « en nuages de points » qui leurs sont associés montrent qu'elles ont des formes très différentes. Et ceci confirme l'adage qui dit que « *un beau graphique vaut mieux qu'un long discours* » !

[5] F.J. ANSCOMBE, « Graphs in Statistical Analysis, » American Statistician, 27 février 1973, pages 17-21.

Quartet d'ANSCOMBE

Chiffres du graphique I

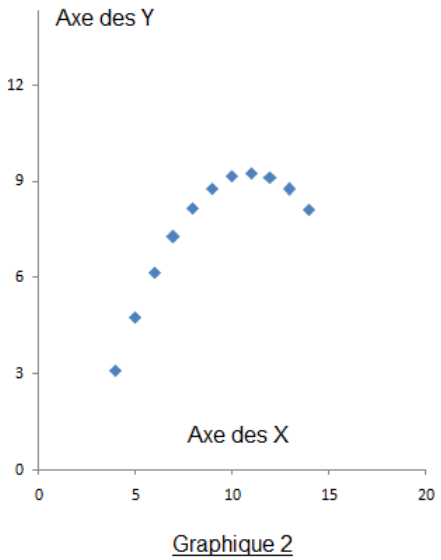
X	Y
10	8,04
8	6,95
13	7,58
9	8,81
11	8,33
14	9,96
6	7,24
4	4,26
12	10,84
7	4,82
5	5,68
Moyenne simple	9	7,5
Ecart-type	3,16	1,94



Graphique 1

Chiffres du graphique II

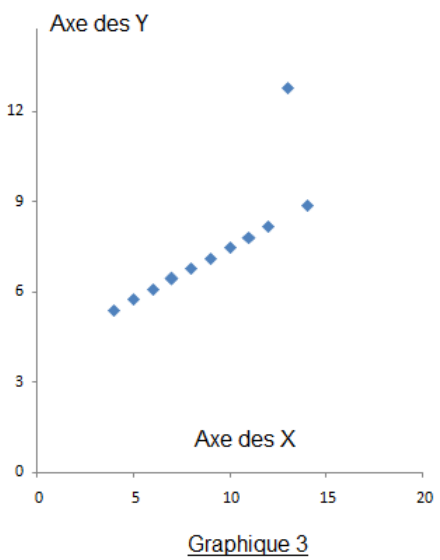
X	Y
10	9,14
8	8,14
13	8,74
9	8,77
11	9,26
14	8,1
6	6,13
4	3,1
12	9,13
7	7,26
5	4,74
Moyenne simple	9	7,5
Ecart-type	3,16	1,94



Graphique 2

Chiffres du graphique III

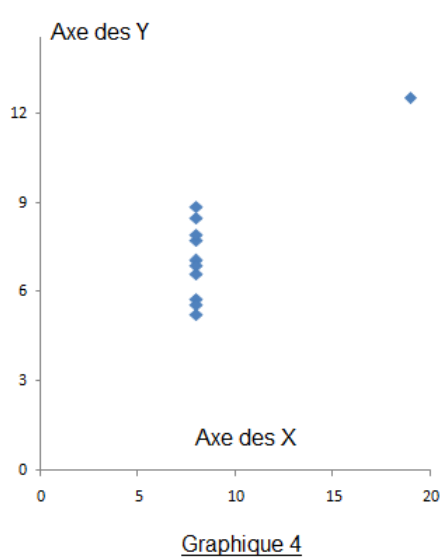
X	Y
10	7,46
8	6,77
13	12,74
9	7,11
11	7,81
14	8,84
6	6,08
4	5,39
12	8,15
7	6,42
5	5,73
Moyenne simple	9	7,5
Ecart-type	3,16	1,94



Graphique 3

Chiffres du graphique IV

X	Y
8	6,58
8	5,76
8	7,71
8	8,84
8	8,47
8	7,04
8	5,25
19	12,5
8	5,56
8	7,91
8	6,89
Moyenne simple	9	7,5
Ecart-type	3,16	1,94



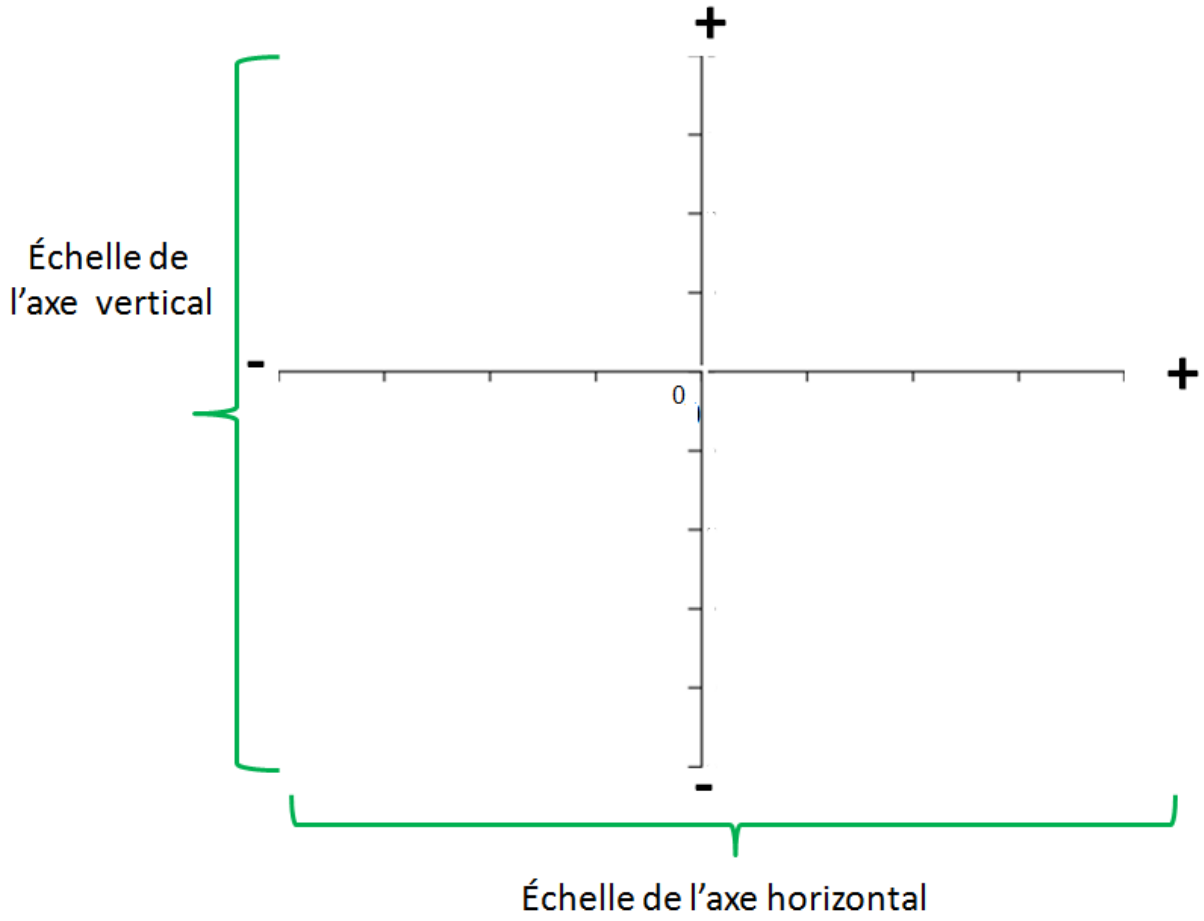
Graphique 4

Source des chiffres : F.J. ANSCOMBE, « Graphs in Statistical Analysis, » American Statistician, 27 février 1973, pages 17-21. Fichier EXCEL.

2 – Les échelles graphiques

f

Beaucoup de graphiques à deux dimensions se présentent sous la forme suivante dite d'un « système de coordonnées cartésiennes » :



Ce qui nous intéresse dans un premier temps ce sont les différentes sortes d'échelles propres à ce type de représentations graphiques.

A – Echelles numériques

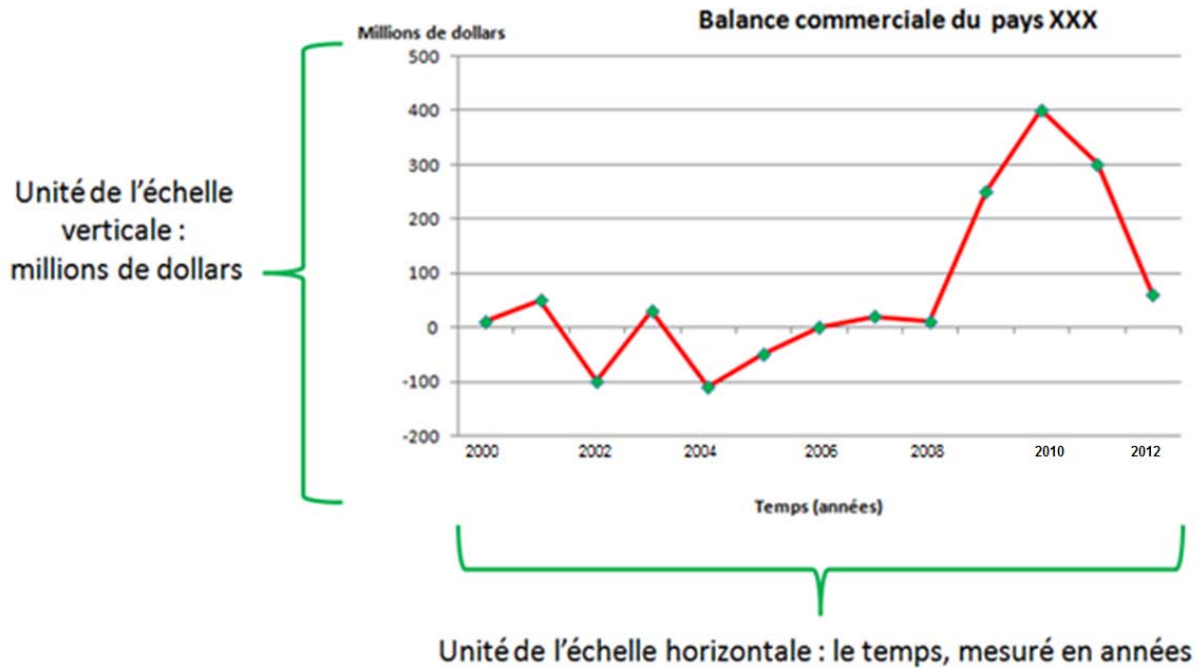
Une **échelle numérique** est une échelle qui mesure des valeurs qui peuvent varier de moins l'infini à plus l'infini. Ci-après, un graphique avec une échelle numérique sur l'axe horizontal et une échelle numérique sur l'axe vertical.

- Sur **l'axe horizontal**, l'unité de mesure numérique est l'année. Les valeurs s'échelonnent entre 1995 et 2008. Si l'on doit dessiner ce graphique à la main sur une feuille de papier, on prendra soin de définir la distance que l'on souhaite consacrer à une année (par exemple : 1 an = 1 cm). Si c'est un logiciel qui réalise le graphique, cette opération devient inutile car les

dimensions du graphique seront choisies par défaut (il est possible cependant de les modifier à son gré en redimensionnant le graphique[6]).

- Sur **l'axe vertical**, l'unité de mesure numérique est le déficit commercial d'un pays, mesuré en millions de dollars. Il varie de -150 millions à + 400 millions. Si l'on doit dessiner ce graphique à la main sur une feuille de papier, on prendra soin de définir la distance que l'on souhaite consacrer à 100 millions de dollars (par exemple : 100 millions de dollars = 1 cm). Si c'est un logiciel qui réalise le graphique, cette opération devient inutile car les dimensions du graphique seront choisies par défaut.

Un exemple d'échelles numériques sur les deux axes



Il est important de toujours bien stipuler sur chaque axe l'unité dans laquelle l'échelle de l'axe considérée est mesurée (ici l'unité est l'année pour l'axe horizontal et les millions de dollars pour l'axe horizontal). En revanche, on peut se dispenser d'indiquer la relation entre la dimension de l'échelle (en centimètres) et la dimension de la variable mesurée sur l'échelle (millions de dollars, temps, etc.).

B – Echelles de catégories

Une **échelle de catégories** est une échelle sur laquelle sont portées des catégories. Il peut s'agir :

- De catégories numériques (échelle quantitative)
- De catégories non numériques dites « nominales » (échelle qualitative)

[6] Dans EXCEL, il faut utiliser la poignée de redimensionnement tout en maintenant enfoncé la touche « Maj », afin de garantir la proportionnalité de la transformation.

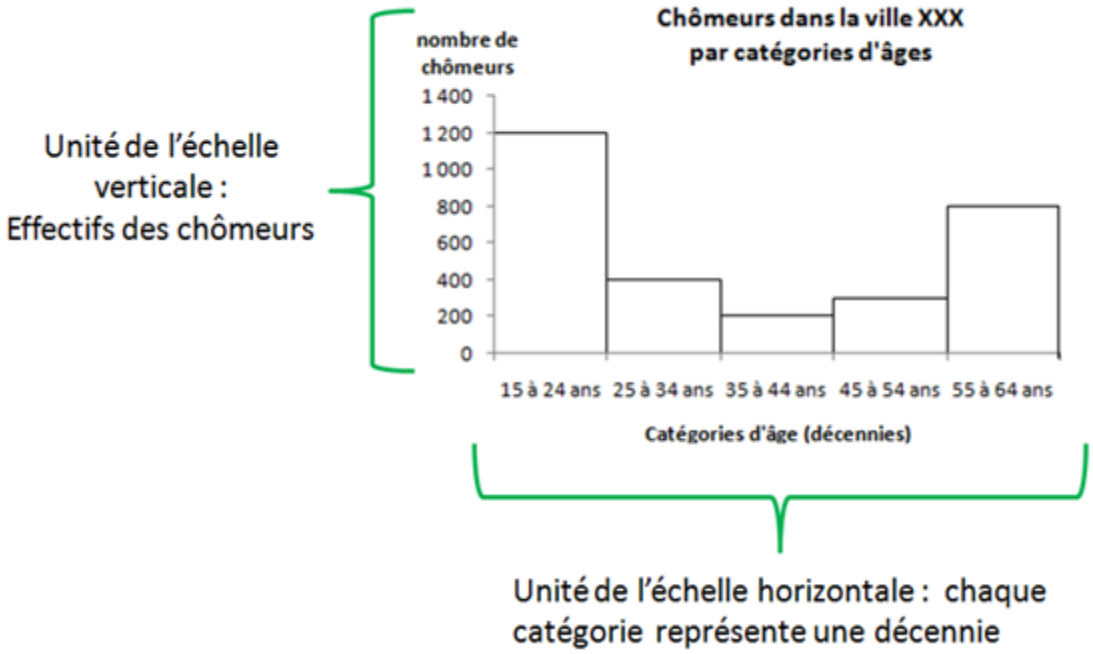
1) Catégories numériques

Ci-après un graphique où l'on a regroupé les chômeurs d'une ville par classes d'âges :

- **Sur l'axe horizontal figurent les catégories d'âges**. C'est une échelle de catégories ou catégorielle. L'unité est l'âge. On remarque aussi que les classes d'âge ont la même amplitude c'est-à-dire que toutes les catégories d'âges ont le même nombre d'années. Il existe aussi des échelles de catégories d'amplitude différentes.

- **Sur l'axe vertical figurent l'effectif des chômeurs qui entrent dans chaque catégorie**. C'est une échelle numérique simple. L'unité est le nombre des chômeurs.

Exemple d'utilisation d'une échelle avec catégories numériques sur l'axe horizontal



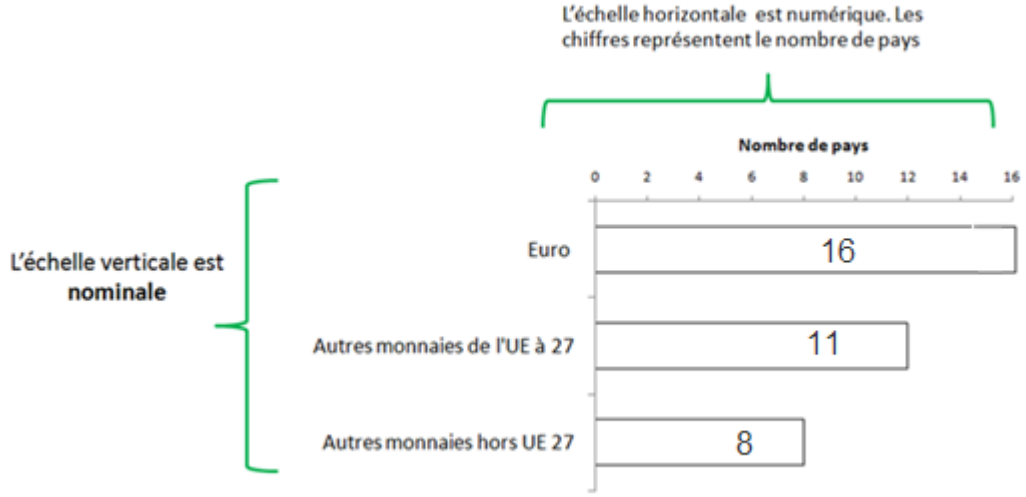
2) Catégories nominales

Ci-après un graphique où l'on a regroupé les 35 pays du tableau 1 du chapitre 1 selon la monnaie utilisée. On a créé trois catégories :

- La catégorie des pays qui font partie de la zone Euro au premier janvier 2009
- La catégorie des pays de l'UE 27 qui n'en font pas partie et utilisent de ce fait d'autres monnaies
- La catégorie des pays hors UE qui sont représentés dans le tableau 1 du chapitre 1.

L'échelle de **l'axe vertical** est donc une échelle nominale. On a placé l'échelle nominale sur l'axe vertical car ainsi il est plus commode d'écrire ce que signifie chaque barre. L'échelle de **l'axe horizontal,** quant-à-elle, est numérique, elle mesure le nombre de pays appartenant à chaque catégorie.

Exemple d'utilisation d'une échelle avec catégories nominales sur l'axe vertical



C – Echelles ordinales

Une **échelle ordinale** est une échelle sur laquelle un ordonnancement des modalités est concevable. Il peut s'agir :

- **D'un classement de préférences**. C'est souvent le cas dans les enquêtes et les sondages d'opinion.

- **D'un classement de rang**. On peut par exemple demander à des investisseurs de classer une liste de pays du plus attractif au moins attractif. Ce classement ne doit pas être confondu avec une échelle numérique simple. En effet, bien qu'il s'agisse de chiffres, l'écart entre les chiffres n'a pas de signification.

Un bon exemple économique d'échelle ordinale est celui des données fournies par le rapport de la Banque mondiale intitulé « Doing Business 2009 Report »[7]). La Banque mondiale classe 181 pays en fonction de la facilité d'y faire des affaires. La première place indique le plus haut niveau de facilité. Un classement élevé (proche de 1) sur l'indice de facilité de faire des affaires signale un environnement réglementaire propice aux opérations commerciales. Le fait que le classement soit représenté sous forme numérique ne doit pas laisser penser cependant qu'il s'agit d'une échelle numérique. En effet, ces chiffres ne représentent que des rangs dans un classement.

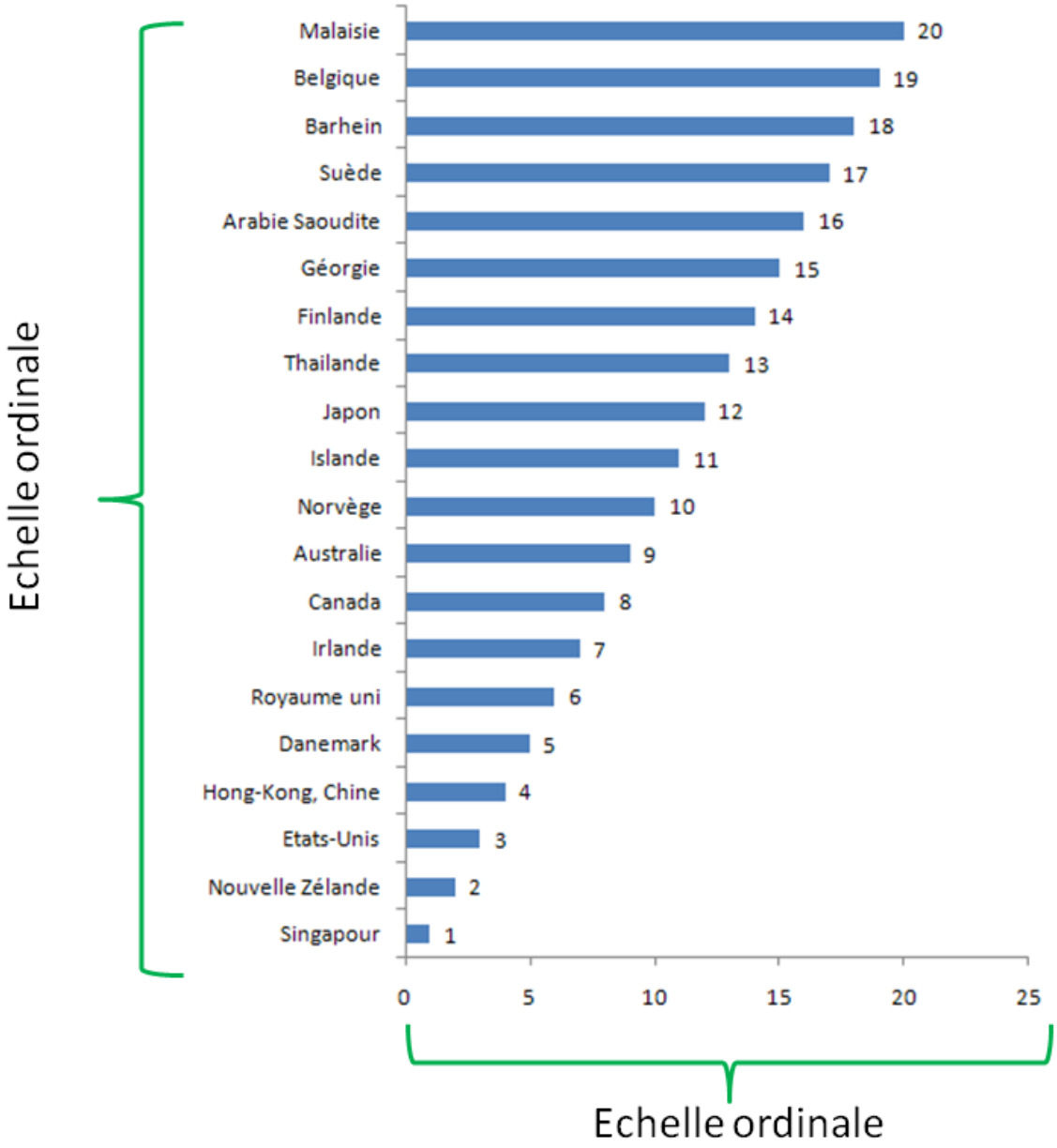
[7] Voir http://www.doingbusiness.org/economyrankings/ (Le classement qui figure dans cette version du cours a été relevé le 10/06/2009 et peut donc être différent de celui qui figure sur le site internet donné en référence).

Classement des pays selon la facilité pour y faire des affaires

Pays	Rang	Pays	Rang	Pays	Rang	Pays	Rang
Singapour	1	Namibie	51	Jordanie	101	Gabon	151
Nouvelle Zélande	2	Koweït	52	Sri Lanka	102	Irak	152
Etats-Unis	3	Colombie	53	Moldavie	103	Djibouti	153
Hong-Kong, Chine	4	Slovénie	54	Seychelles	104	Haïti	154
Danemark	5	Bahamas	55	Guyane	105	Comores	155
Royaume uni	6	Mexique	56	Croatie	106	Sierra Leone	156
Irlande	7	Oman	57	Nicaragua	107	Libéria	157
Canada	8	Mongolie	58	Swaziland	108	Zimbabwe	158
Australie	9	Turquie	59	Uruguay	109	Tadjikistan	159
Norvège	10	Vanuatu	60	Bangladesh	110	Mauritanie	160
Islande	11	Taiwan, Chine	61	Ouganda	111	Côte d'Ivoire	161
Japon	12	Pérou	62	Guatemala	112	Afghanistan	162
Thaïlande	13	Jamaïque	63	Argentine	113	Togo	163
Finlande	14	Samoa	64	Égypte	114	Cameroun	164
Géorgie	15	Italie	65	Paraguay	115	RDP du Laos	165
Arabie Saoudite	16	St. Vincent & Grenadines	66	Éthiopie	116	Mali	166
Suède	17	St. Kitts-et-Nevis	67	Costa Rica	117	Guinée Équatoriale	167
Bahreïn	18	Kirghizistan	68	Nigéria	118	Angola	168
Belgique	19	Maldives	69	Bosnie-Herzégovine	119	Bénin	169
Malaisie	20	Kazakhstan	70	Fédération de Russie	120	Timor Oriental	170
Suisse	21	Macédoine	71	Népal	121	Guinée	171
Estonie	22	Salvador	72	Inde	122	Niger	172
Corée	23	Tunisie	73	Lesotho	123	Érythrée	173
Maurice	24	Dominique	74	Bhoutan	124	Venezuela	174
Allemagne	25	République Tchèque	75	Brésil	125	Tchad	175
Pays-Bas	26	Pologne	76	EF de Micronésie	126	São Tomé & Príncipe	176
Autriche	27	Pakistan	77	Tanzanie	127	Burundi	177
Lituanie	28	Belize	78	Maroc	128	Congo, République du	178
Lettonie	29	Kiribati	79	Indonésie	129	Guinée-Bissau	179
Israel	30	Trinité-et-Tobago	80	Gambie	130	République Centrafricaine	180
France	31	Panama	81	Cisjordanie et Gaza	131	Congo, République Démocratique du	181
Afrique du Sud	32	Kenya	82	Algérie	132		
Azerbaïdjan	33	Chine	83	Honduras	133		
Ste. Lucie	34	Grenade	84	Malawi	134		
Porto Rico	35	Belarus	85	Cambodge	135		
Slovaquie	36	Albanie	86	Équateur	136		
Qatar	37	Ghana	87	République Arabe Syrienne	137		
Botswana	38	Le Brunéi	88	Ouzbékistan	138		
Fiji	39	Îles Salomon	89	Rwanda	139		
Chili	40	Montenegro	90	Philippines	140		
Hongrie	41	Palau	91	Mozambique	141		
Antigua-et-Barbuda	42	Vietnam	92	Iran	142		
Tonga	43	Îles Marshall	93	Cap-Vert	143		
Arménie	44	Serbie	94	Madagascar	144		
Bulgarie	45	Papouasie-Nouvelle Guinée	95	Ukraine	145		
Emirats Arabes Unis	46	Grèce	96	Surinam	146		
Roumanie	47	République Dominicaine	97	Soudan	147		
Portugal	48	Yémen	98	Burkina Faso	148		
Espagne	49	Liban	99	Sénégal	149		
Luxembourg	50	Zambie	100	Bolivie	150		

Source : http://www.doingbusiness.org/economyrankings/

Exemple d'échelles ordinales sur l'axe vertical et sur l'axe horizontal
(Classement des 20 premiers pays du monde où il est facile de « faire des affaires »)



Source : http://www.doingbusiness.org/economyrankings/ (classement relevé le 8 octobre 2008)

L'**échelle horizontale** du graphique ci-dessus donne le classement de 20 pays. Il s'agit donc bien d'une **échelle ordinale** car l'écart qui sépare les pays n'est pas quantifiable. Par exemple, Singapour est classée au premier rang et la Malaisie est classée au 20ème rang. Si l'on fait la différence 20-1 = 19, on ne peut pas en conclure que Singapour est 19 fois plus attractif que la Malaisie du point de vue des investissements internationaux. Il ne s'agit pas d'une échelle réellement numérique, mais d'une échelle spéciale, qualifiée pour cette raison d'« échelle ordinale ».

L'**échelle verticale** du graphique ci-dessus est également ordinale : les pays y sont classés par ordre décroissant, du moins attractif au plus attractif.

D – Echelles verticales doubles

Lorsque que l'on veut représenter les données relatives à deux variables ou à deux caractères sur un même graphique, on a recours à une **échelle verticale double** pour faciliter la lecture.

L'exemple suivant va permettre d'illustrer ce point. Le tableau ci-dessous montre deux séries mensuelles. La première indique le temps passé par une personne sur Internet chaque mois (en heures) et la seconde série indique le total de la somme dépensée sur différents sites marchands.

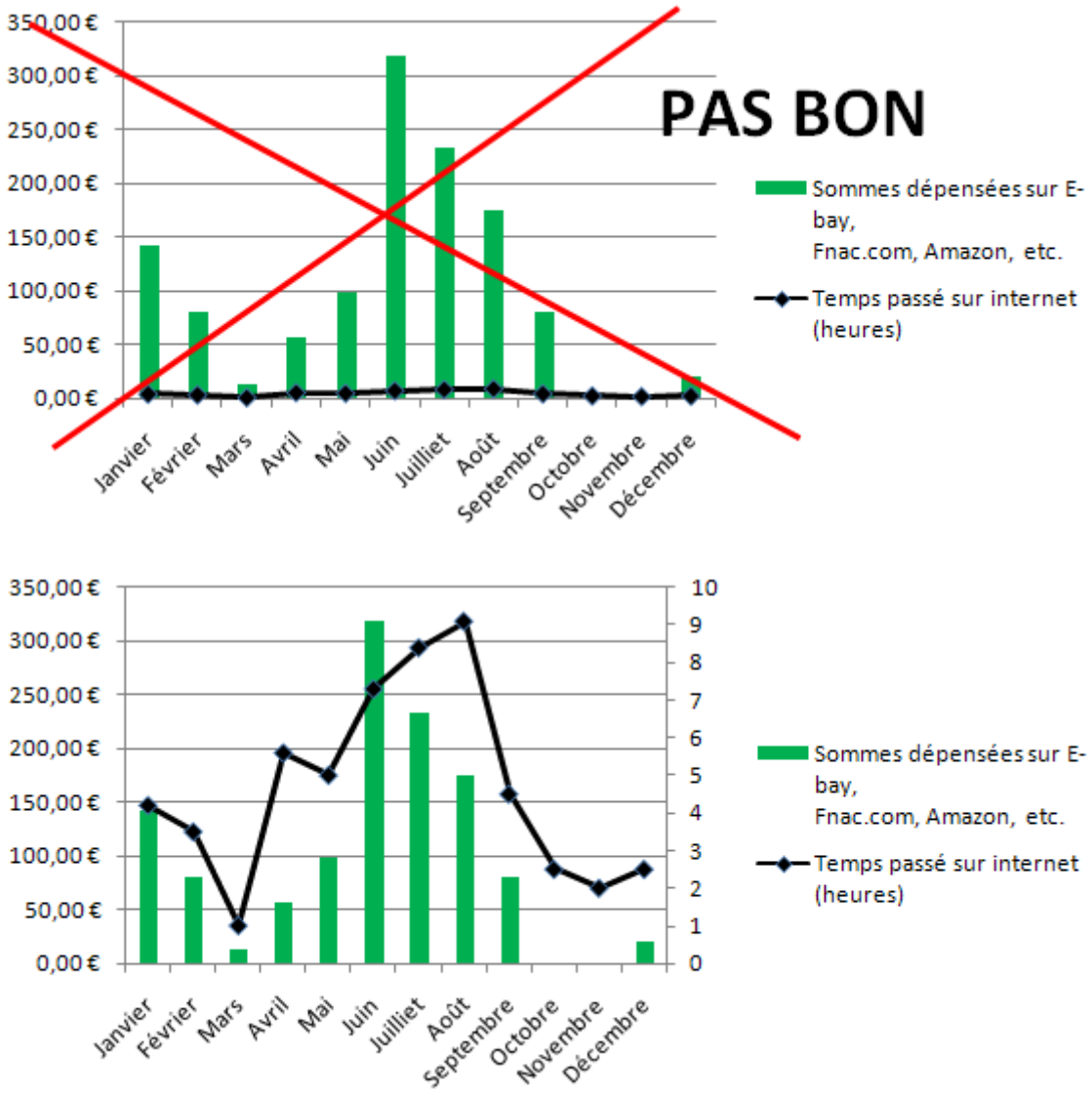
Temps passé sur Internet (heures/mois)
et sommes dépensées sur différents sites marchands (euros)

Mois	Temps passé sur internet (heures)	Sommes dépensées sur E-bay, Fnac.com, Amazon, etc.
Janvier	4,2	142,25 €
Février	3,5	79,59 €
Mars	1	12,50 €
Avril	5,6	56,42 €
Mai	5	98,74 €
Juin	7,3	319,12 €
Juilliet	8,4	232,58 €
Août	9,1	175,91 €
Septembre	4,5	79,50 €
Octobre	2,5	0 €
Novembre	2	0 €
Décembre	2,5	19,93 €

Voir le fichier EXCEL

Source : Matthew McDONALD, « Creating a Combination Chart in EXCEL 2007 », Matthew McDONALD, video Youtube : http://fr.youtube.com/watch?v=WW2IDE4rPCc

Exemple de l'utilité des échelles verticales doubles



Voir le fichier EXCEL

Imaginons que l'on souhaite savoir s'il existe une **covariation** (variation « commune ») entre ces deux séries. Une bonne façon de procéder est de les mettre sous forme d'un graphique. Cependant, comme les deux échelles sont différentes, il faut réserver par exemple l'échelle verticale de droite pour le temps passé sur internet (qui est exprimé en heures) et l'échelle verticale de gauche pour les sommes dépensées (qui sont exprimées en euros). En effet, si l'on utilise seulement l'échelle verticale de gauche pour tracer les deux séries, la plus petite (celle des heures passées sur internet) sera écrasée par la plus grande (celle des euros dépensés) et le graphique ne révèlera rien du tout. Inversement, si l'on réserve une échelle pour chaque série, on obtient alors un graphique beaucoup plus lisible qui semble bel et bien révéler que plus cette personne a passé de temps sur Internet et plus elle a dépensé d'argent (ceci n'est qu'un constat de **covariation** et non une relation de causalité, bien sûr).

E – Echelles logarithmiques

1) Définition

L'échelle logarithmique est une échelle qui mesure le logarithme décimal des valeurs de la variable. C'est un excellent moyen de mettre en évidence une idée ou un résultat grâce aux propriétés des logarithmes décimaux.

Avant de présenter le mode de construction d'un graphique à échelle semi-logarithmique, il convient cependant de faire un rappel sur les logarithmes décimaux.

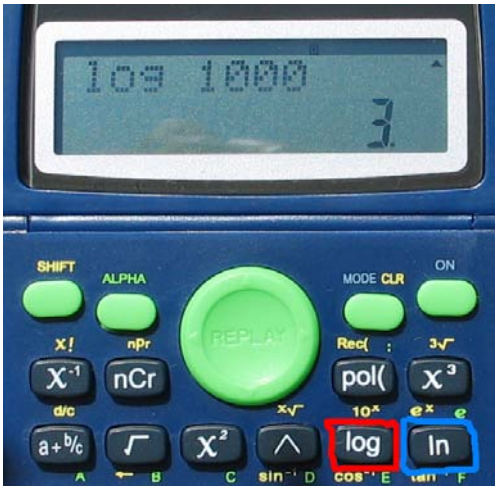
À ce sujet, sur un plan pratique, il est plus important de *savoir obtenir un logarithme décimal avec une machine à calculer*, que de c*omprendre le pourquoi et le comment des logarithmes décimaux*, ce qui est certainement passionnant mais relève d'un cours de mathématiques.

Nous allons donc commencer par voir comment on calcule un logarithme décimal avec une machine à calculer standard (ci-après la "SC-05 Plus") avant de faire un bref rappel sur les logarithmes décimaux.

2) Calcul pratique du log décimal d'un nombre

Il suffit d'appuyer sur la touche "log" d'une machine à calculer pour obtenir le log d'un nombre. Par exemple, l'image ci-dessous illustre le calcul du log décimal de 1000. L'écran indique que le log décimal de 1000 est égal à 3. Pour obtenir ce résultat, on procède ainsi :

1 - Allumer la machine
2 - appuyer sur la touche "log" entourée en rouge (attention de ne pas appuyer sur la touche "ln" qui est entourée en bleu juste à côté et qui sert à calculer les logarithmes naturels)
3 - Entrer le chiffre 1000
4 - Appuyer sur la touche "="
5 - Le résultat (ici le log de 1000 c'est 3) apparaît sur l'écran de la calculatrice.



3) Rappels sur le logarithme décimal

Le logarithme décimal d'un nombre est la puissance à laquelle il faut élever 10 pour obtenir ce nombre. Appliquons cette définition à quelques nombres. Quel est, par exemple, le logarithme décimal de 1 ? Autrement dit, à quelle puissance faut-il élever 10 pour obtenir 1 ?

La puissance à laquelle il faut élever 10 pour obtenir 1 est 0. Par conséquent, le logarithme décimal de 1 est égal à zéro:

$$10^0=1$$

On écrira donc :

$$\log 1 = 0$$

Quel est le logarithme décimal de 100 ? C'est la puissance à laquelle il faut élever 10 pour obtenir 100 :

$10^2=100$, parce qu'il faut élever 10 à la puissance 2 pour obtenir 100. Donc le logarithme décimal de 100 est égal à 2. On écrira par conséquent :

$$\log 100 =2$$

Inversement, si l'on demande "De quel chiffre 3 est-il le logarithme décimal ?", on fera le raisonnement inverse. Sachant que $10^3=1000$, la réponse est donc :

$$\log 3 = 1000$$

Autrement dit, le logarithme décimal de 1000 est égal à 3.

4) Exemples

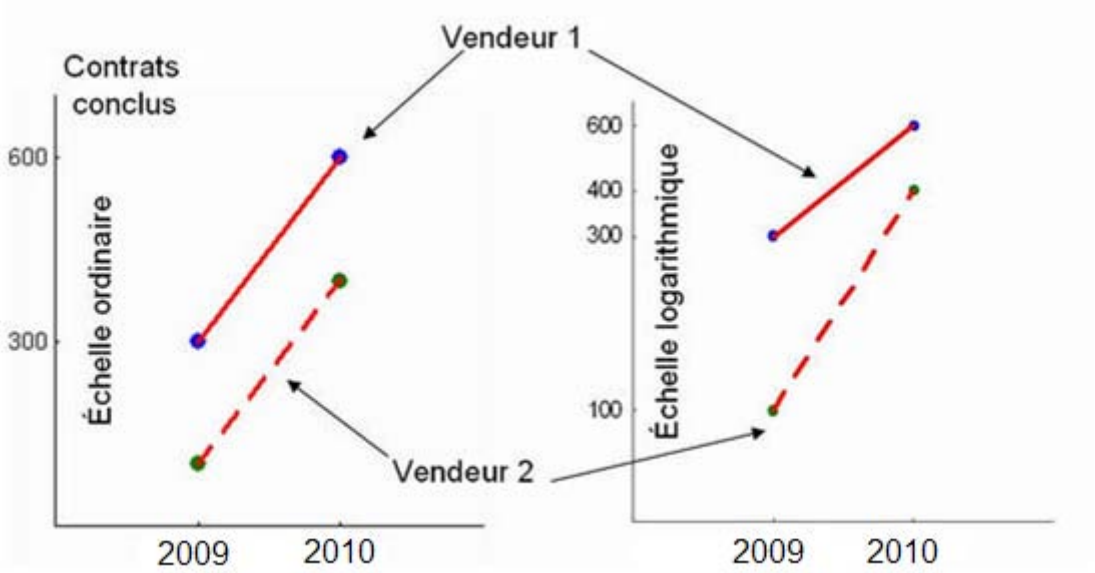
a) L'échelle logarithmique permet de mieux voir les différences de progression

On décide de comparer le nombre de contrats conclus par deux vendeurs en 2009 et 2010 :

	2009	2010	Progression
Vendeur 1	300	600	multiplié par 2
Vendeur 2	100	400	multiplié par 4

Le tableau montre que le nombre de contrats conclus par le vendeur 1 a été multiplié par 2 et que le nombre de contrats conclus par le vendeur 2 a été multiplié par 4.

Sur un graphique ordinaire (à gauche ci-dessous), les deux progressions sont parallèles. En revanche, sur un graphique avec une ordonnée logarithmique, on voit clairement que la progression du vendeur 2 est plus rapide que celle du vendeur 1.



Sur le graphique de droite, l'échelle de l'ordonnée est logarithmique, mais les chiffres indiqués (les nombres de contrats) sont les mêmes que sur le graphique de gauche. Cependant, au lieu d'utiliser les valeurs elles-mêmes, le tracé utilise le logarithme décimal des valeurs, comme indiqué dans le tableau ci-dessous :

	2009	2010
Vendeur 1	log(300)=2,48	log(600)=2,78
Vendeur 2	log(100)=2	log(400)=2,6

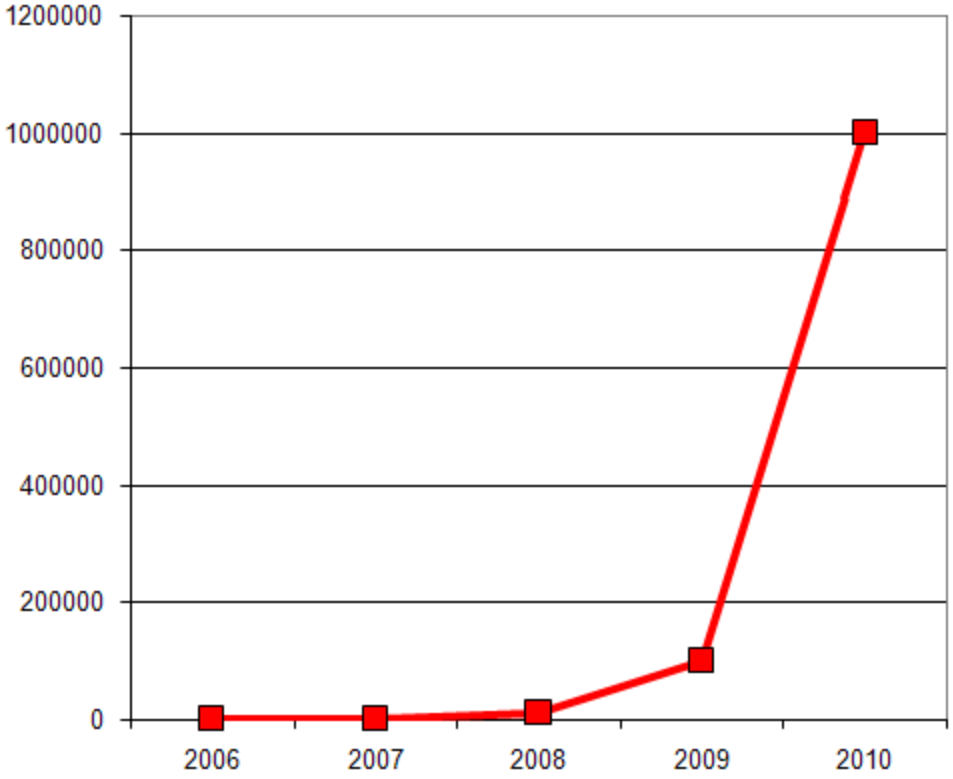
b) L'échelle log linéarise les évolutions à taux constant

On souhaite faire un graphique indiquant l'évolution du chiffre d'affaires d'une entreprise dont la croissance est très rapide :

Années	CA (en euros)
2006	100
2007	1000
2008	10000
2009	100000
2010	1000000

Comme on peut le voir sur le graphique ci-après les valeurs pour 2006, 2007 et 2008 sont écrasées par rapport à celles de 2009 et 2010 :

Graphique avec échelle des ordonnées ordinaires (sans transformation logarithmique)



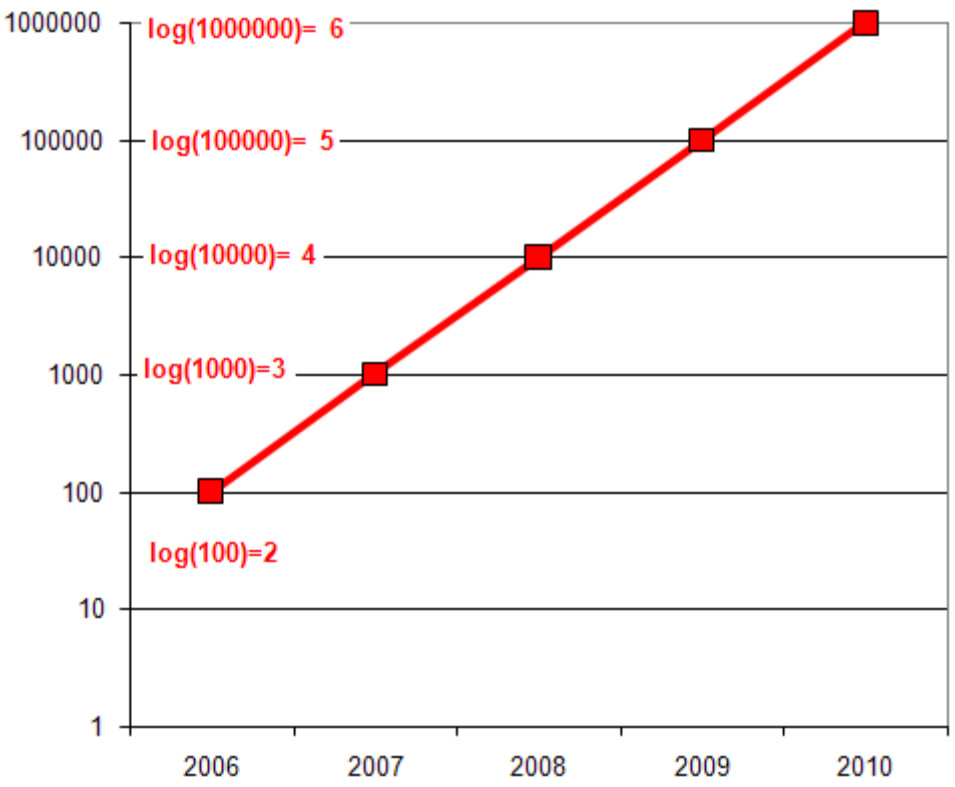
Fichier EXCEL

Appliquons une transformation logarithmique aux valeurs de l'ordonnée :

Années	CA (en euros)	
2006	100	log(100)=2
2007	1000	log(1000)=3
2008	10000	log(10000)=4
2009	100000	log(100000)=5
2010	1000000	log(1000000)=6

On obtient alors le graphique suivant où la progression devient linéaire :

Graphique « semi logarithmique» (l'échelle des ordonnées est logarithmique)

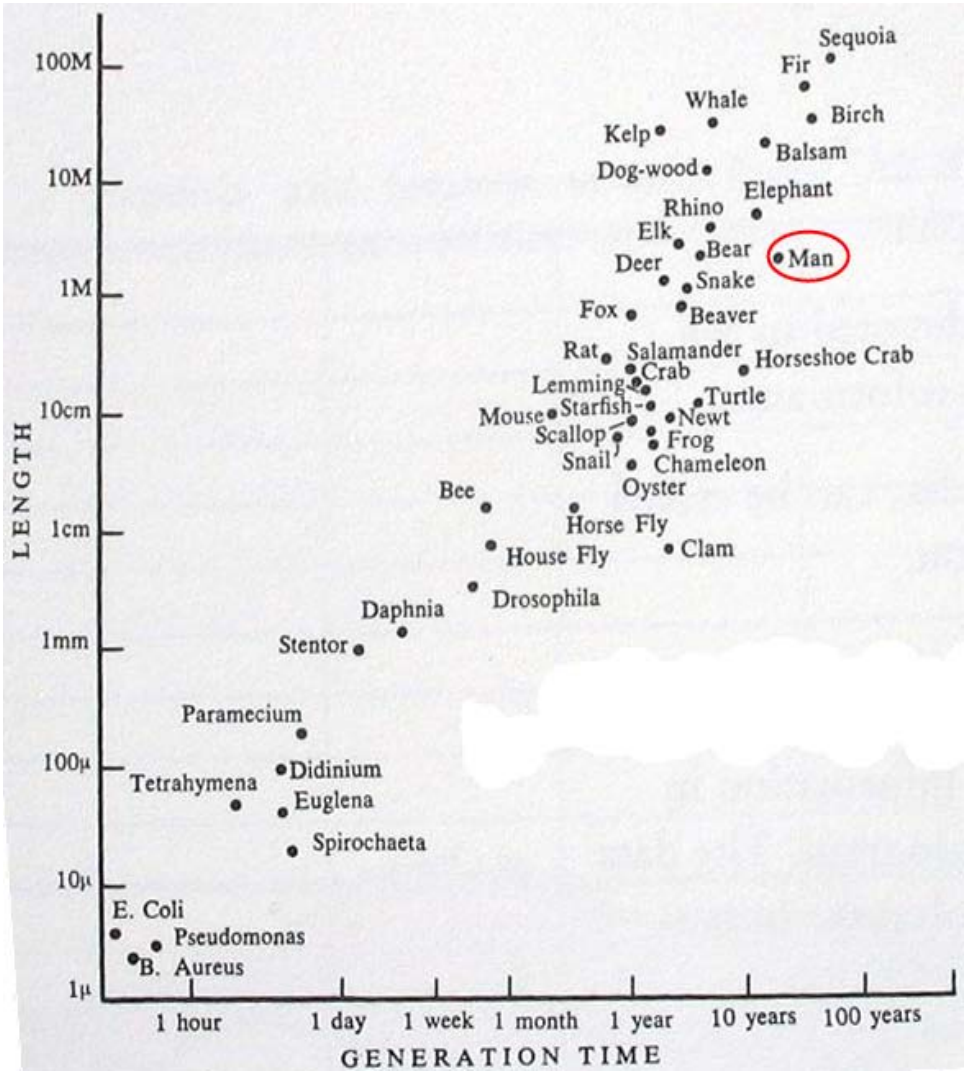


Fichier EXCEL

5 – Echelle doublement logarithmique

Il existe aussi des graphiques avec échelle logarithmique sur les deux axes. Autrement dit, non seulement l'échelle des ordonnées est logarithmique, mais également l'échelle des abscisses. C'est assez peu fréquent en économie. L'exemple donné ci-après est celui de la relation entre le temps de génération (période allant de la naissance à l'âge moyen de reproduction) et la longueur de divers êtres vivants. On voit nettement sur ce graphique que le temps de génération croît avec la longueur. Mais on a ici un cas très intéressant où l'échelle de temps varie entre moins d'une heure et 100 ans et où l'échelle de longueur varie de l'infiniment petit à 100 m. Pour bien contraster ce graphique avec le graphique semi-logarithmique étudié précédemment, on parle de **graphique à échelle doublement logarithmique**.

Exemple d'un graphique à échelle doublement logarithmique

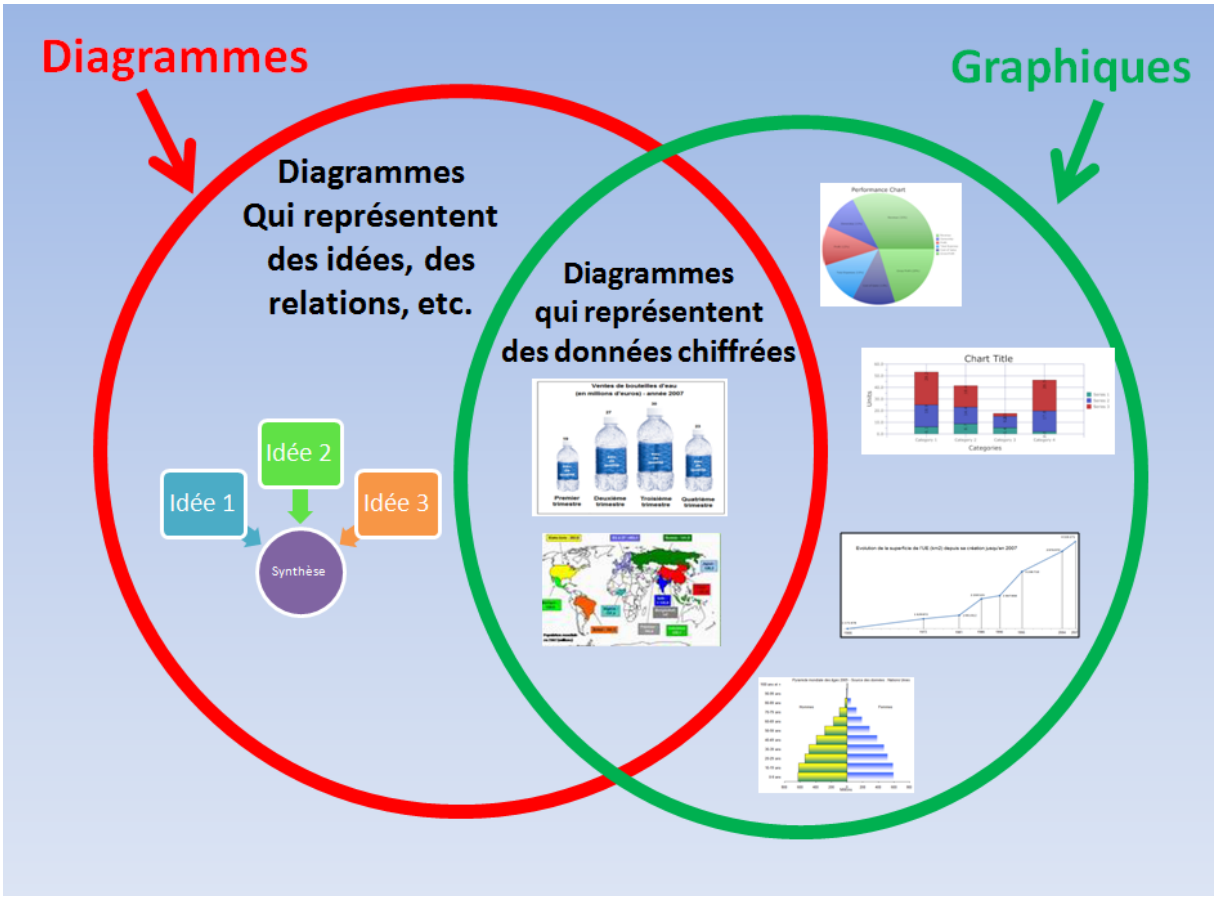


Source: John Tyler BONNER, *Size and Cycle: An Essay on the Structure of Biology* (Princeton, 1965), p.17. Reproduit dans Edward R. TUFTE, *The Visual Display of Quantitative Information* (Graphics Press LLC, 2004), p. 94

3 – Diagrammes statistiques

Un **diagramme** est une représentation visuelle simplifiée et structurée de concepts, d'idées, de constructions, des relations, de l'anatomie **et aussi parfois** (et c'est en cela qu'il nous intéresse ici) **de données statistiques**. Il est employé dans tous les aspects des activités humaines pour visualiser et clarifier. Un diagramme permet aussi de décrire des phénomènes, de mettre en évidence des corrélations en certains facteurs ou de représenter des parties d'un ensemble.

Il existe donc une relation entre diagramme et graphique, que nous pouvons d'ailleurs illustrer par le **diagramme de VENN** ci-après (lequel est un **diagramme non statistique** !).

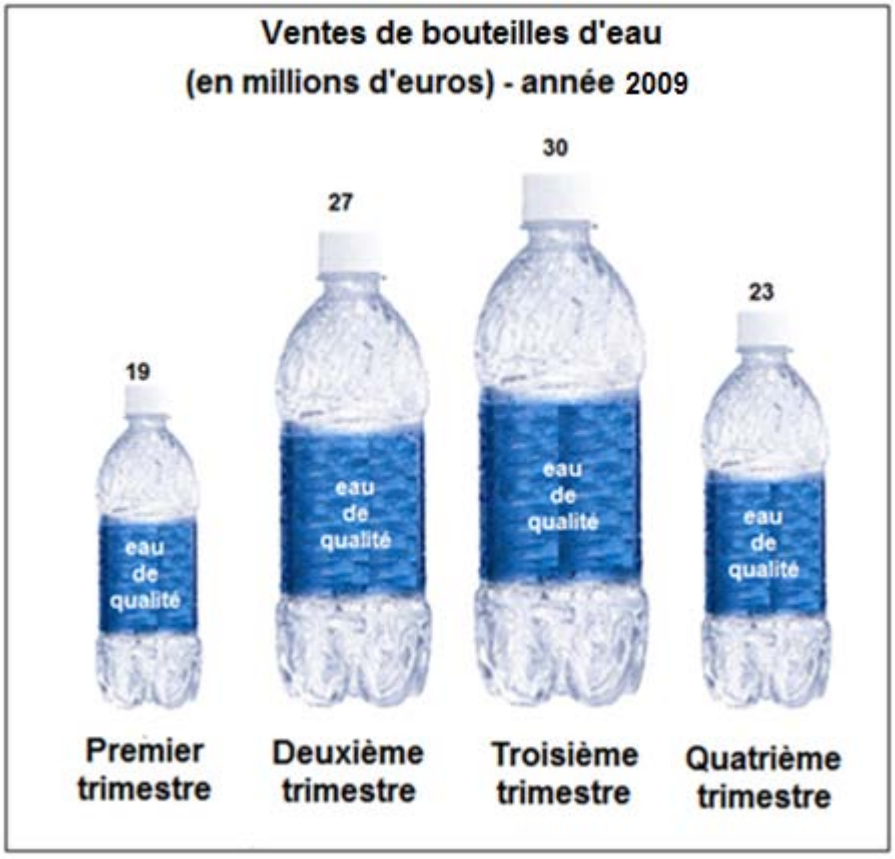


A – Pictogramme

Le **pictogramme** est un dessin, généralement en couleur, conçu de préférence par un infographiste, qui essaie de traduire non seulement des données quantitatives, mais également des informations d'ordre commercial ou esthétique. Dans l'exemple suivant, quatre figurines représentent des bouteilles d'eau : la « corpulence » de chaque bouteille correspond aux ventes de la marque fictive « eau de qualité ». En réalité, il s'agit au départ d'un graphique en colonnes qui a été redessiné pour remplacer les colonnes par des bouteilles. Ce qui compte ici, c'est la hauteur de la bouteille. Plus la bouteille est haute, plus les ventes sont importantes. On voit aussi sur le graphique que la « corpulence » augmente avec la hauteur. Il faudrait donc, pour que le pictogramme reste conforme aux canons habituels (proportionnalité des surfaces aux grandeurs représentées), que la surface des bouteilles soit proportionnelle aux chiffres qu'elles représentent. Ce n'est pas le cas pour des raisons esthétiques et/ou de réalisme visuel. Il y a donc un risque (certains diront « une opportunité ») à utiliser ce genre de graphique, risque qui est d'induire des impressions visuelles qui ne sont pas contenues dans la réalité des chiffres.

Remarquons qu'il s'agit d'un pictogramme qui traduit principalement une information statistique. Mais beaucoup de pictogrammes ne traduisent que des informations diverses, non nécessairement statistiques.

Un pictogramme qui représente des données chiffrées



Un autre pictogramme typique de la puissance évocatrice de ce type de représentation

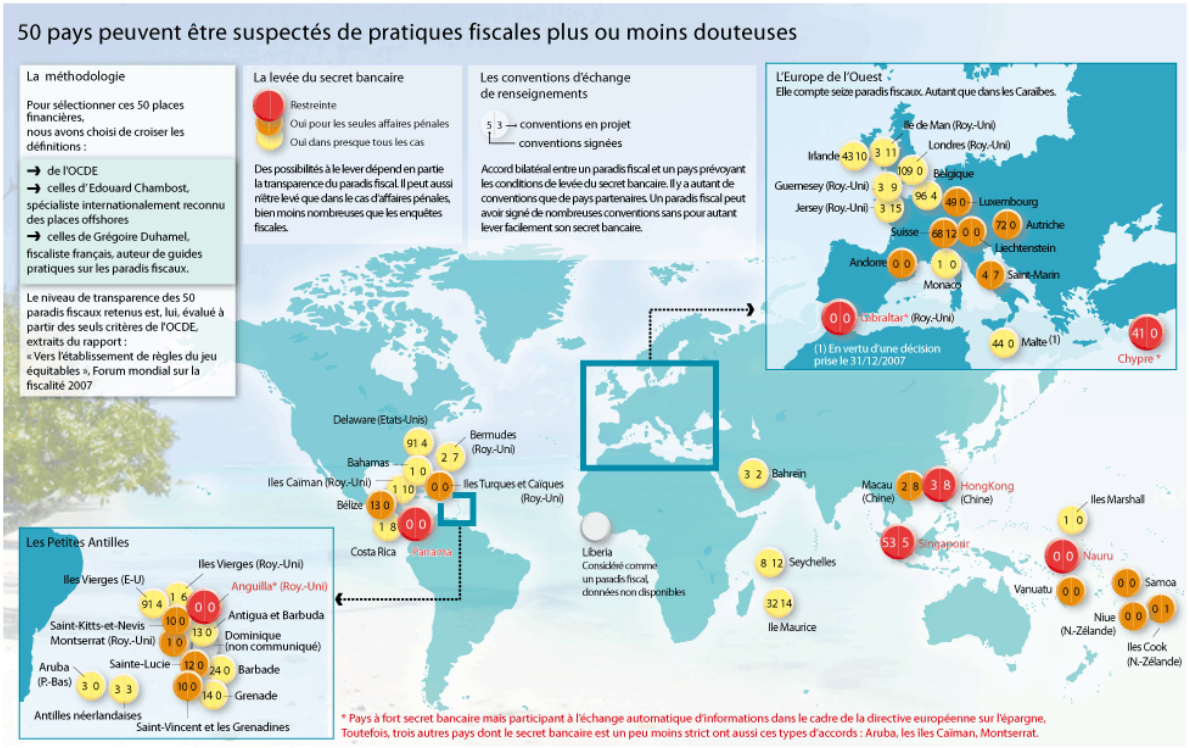


Image agrandie (JPEG)

B - Cartogramme

Le **cartogramme** sert à localiser géographiquement des valeurs ou des modalités relatives à une variable ou à un caractère décrivant les unités statistiques d'une population. Les cartogrammes sont désormais très fréquemment utilisés pour représenter toutes sortes d'informations, de la météorologie, aux résultats électoraux, en passant naturellement par des informations économiques.

Un cartogramme statistique : La carte des paradis fiscaux en 2010



Source : http://medias.lemonde.fr/mmpub/edt/ill/2008/07/08/h_4_PARADIS+X1I1.gif

Image agrandie (JPEG)

Un exemple très intéressant de cartogramme « interactif », est le « planétogramme » mis en place par le magazine « The Economiste », pour suivre l'évolution de la dette publique des pays : http://buttonwood.economist.com/content/gdc

C - Diagramme de GANTT

Le **diagramme de GANTT**, du nom de l'ingénieur américain Henry Laurence GANTT (1861-1919) qui l'a popularisé, est un outil de gestion de projets. Il sert à visualiser dans le temps les différentes étapes qui composent un projet. Il existe des logiciels spécifiques qui permettent de produire ce diagramme, mais il est aussi possible d'utiliser EXCEL 2007 pour le tracer[8].

Nous allons partir d'un exemple simple sous forme d'un tableau décrivant la durée des différentes étapes de la réalisation d'un mémoire et nous transformerons ce tableau en diagramme de Gantt. Ci-après, le tableau qui va servir à faire le diagramme :

Tableau chronologique des tâches à effectuer pour réaliser un mémoire

Liste des tâches	Date de début	Durée (jours)	Date de fin
Choix du sujet	1/1	7	8/1
Documentation	8/1	10	18/1
Calculs, tableaux & graphiques	18/1	12	30/1
Plan	30/1	4	3/2
Rédaction des parties	3/2	25	28/2
Intro et conclusion	28/2	6	6/3
Bibliographie	6/3	7	13/3
TDM, sommaire, index, etc.	13/3	5	18/3
Charte graphique	18/3	5	23/3
Page de garde	23/3	3	26/3
Préparation pour l'impression	26/3	2	28/3
Impression	28/3	4	1/4

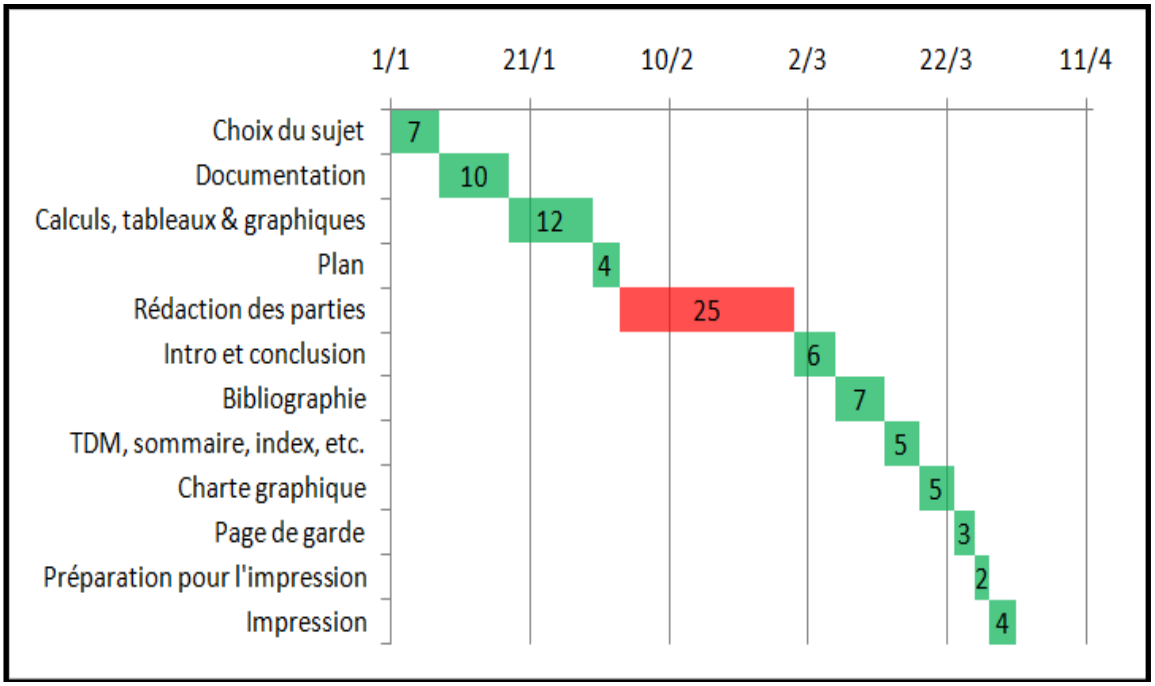
Sur ce tableau figurent successivement :

- Dans la première colonne, l'énumération des tâches
- Dans la seconde colonne la date de début de chaque tâche
- Dans la troisième colonne la durée de chaque étape en jours
- Dans la dernière colonne, la date d'achèvement de chaque tâche

[8] Il existe de nombreuses variantes du Diagramme de GANTT, l'une des plus connue étant la frise chronologique, ou représentation d'une séquence d'événements dans le temps. Pour des exemples voir le cours d'Histoire des faits économiques, à commencer par la frise chronologique de l'introduction : http://www.mazerolle.fr/Histoire-des-faits-economiques/HFEchap00.pdf

Voici maintenant le diagramme tel qu'il apparaît une fois terminé.

Diagramme de GANTT pour la réalisation d'un mémoire



Grâce à ce diagramme, nous pouvons :

- Visualiser la durée totale du projet
- Apprécier la durée de chaque étape et éventuellement réaffecter le temps total entre les différentes tâches
- Vérifier si les chevauchements sont réalistes (ici – pour simplifier- il n'y a pas de chevauchement)
- Ordonner les tâches dans un ordre chronologique

4 – Graphiques usuels

À l'ère du numérique et des télécommunications, les graphiques sont partout. Ces représentations visuelles colorées, aux formes très diverses sont plus agréables à regarder que les tableaux et permettent souvent de mieux faire passer un message au premier coup d'œil.

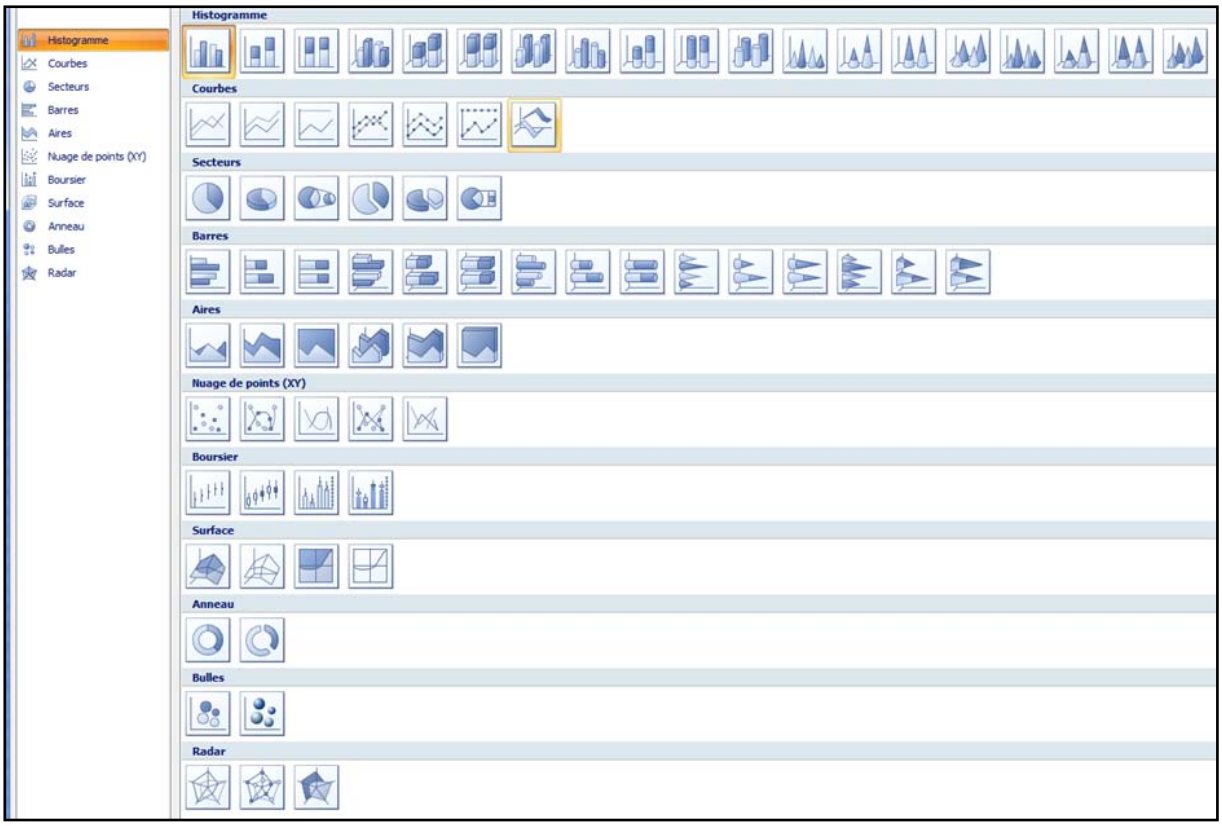
Il est également très facile, grâce aux logiciels tels que EXCEL 2007 de Microsoft de donner une « profondeur » aux différents graphiques, afin qu'ils apparaissent comme ayant trois dimensions (Il est aussi possible dans EXCEL 2007, de représenter « réellement » trois dimensions pour certains graphiques en barres ou pour les graphiques dits « de surface »).

Il ne faut cependant pas abuser de la possibilité qui nous est donnée aujourd'hui de réaliser des graphiques complexes. Car cette complexité peut finir par rendre le graphique difficilement compréhensible. Mieux vaut s'en tenir aux principales représentations graphiques connues de tous : diagrammes en colonne, diagramme en barres, lignes, « camemberts », etc.…

Il existe une grande quantité de graphiques, tous plus imaginatifs les uns que les autres. Ainsi, à titre d'exemple, la figure ci-après montre les 73 possibilités de graphiques simples, regroupées en 11 catégories, qui peuvent être réalisés avec le logiciel EXCEL 2007. Mais en réalité, il est possible d'en faire beaucoup plus, soit en combinant ces formes de base, soit en utilisant certaines astuces.

De plus, grâce à des logiciels tels que Powerpoint de Microsoft, il est possible de réaliser des graphiques animés, ou même des graphiques interactifs (pouvant par exemple être modifiés par l'utilisateur), à l'aide du logiciel Flash d'Adobe ou bien d'Applets Java.

Les 73 représentations graphiques de base dans EXCEL 2007, regroupées dans 11 catégories communes



Les quatre formes graphiques les plus fréquemment utilisées sont :

- Barres verticales ou horizontales
- Courbes ou aires délimitées par des courbes
- Nuages de points
- Secteurs ou « camemberts »

A – Graphiques en barres

On distingue les graphiques en barres verticales et les graphiques en barres horizontales. Mais pour chacune de ces deux catégories, il existe 3 variétés : simple, multiples et tronçonnée. S'agissant des barres « tronçonnées », elles se subdivisent en barres d'effectifs ou en barres de pourcentages empilés à 100%. Le Tableau ci-après donne une représentation schématique de ces 8 variétés.

Pour créer les 8 graphiques les données ci-dessous ont été utilisées :

Ventes 2009 (euros)				
	Leila	Ahmed	Pierre	Elodie
Lyon	18 753 951	8 754 668	9 785 246	16 487 564
Marseille	13 225 478	20 154 287	17 892 555	15 897 233
Paris	37 895 214	35 877 421	32 558 741	22 044 687
	69 874 643	64 786 376	60 236 542	54 429 484

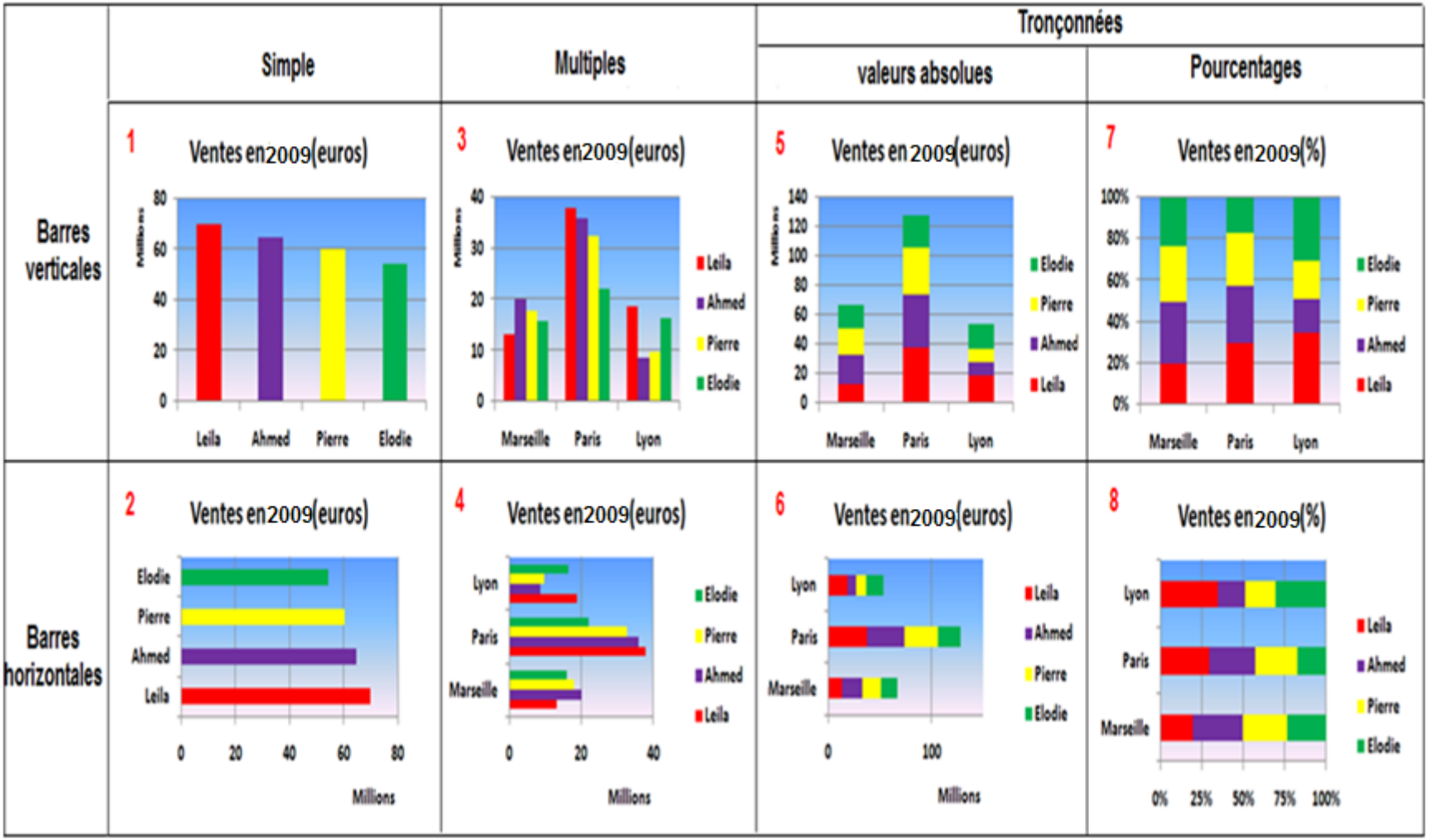
.

Il s'agit du chiffre d'affaires hypothétique qu'une entreprise a réalisé en 2009, soit 249 327 045 euros répartis entre ses 4 vendeurs et dans les trois villes où se trouvent ses clients. Les 8 graphiques ci-après sont les 8 principales façons de représenter ces données (ou une partie d'entre elles) **sous forme de graphiques en barres**. Chaque graphique en barres fait ressortir ces chiffres d'une manière différente et pourra donc être préféré selon les circonstances dans lesquelles le graphique est utilisé.

1) Barres verticales

La première ligne du tableau ci-après intitulé « les 8 principales variétés de graphiques en barres » représente les 4 principales façons de disposer les données du tableau sous forme de barres verticales (ou colonnes). Comme l'axe horizontal représente des catégories nominales, la largeur des colonnes n'a pas d'importance, pourvu qu'elle soit identique pour toutes les colonnes. Cette largeur peut varier d'un minimum qui serait un simple trait vertical jusqu'à un maximum qui serait représenté par le fait que toutes les colonnes seraient « collées ».

Les 8 principales variétés de graphiques en barres



Fichier EXCEL

a) Simple

Le type « barres verticales simples » est représenté par le graphique numéroté 1. Il permet de mettre en évidence le chiffre d'affaire réalisé par chaque vendeur. On voit du premier coup d'œil combien chaque vendeur à réalisé et quel est celui qui a réalisé le plus gros chiffre d'affaires. Les chiffres du graphique correspondent à la dernière ligne du tableau.

b) Multiples

Le type « barres verticales multiples » est représenté par le graphique numéroté 3. Il permet de mettre en évidence le chiffre d'affaire réalisé par chaque vendeur dans chaque ville. Pour chacune des villes (Marseille, Paris, Lyon) on peut voir combien chaque vendeur à réalisé. Ce graphique permet de voir quel est le vendeur le plus performant dans chaque ville.

c) tronçonnées

Le type « barres verticales tronçonnées » est représenté par les graphiques numérotés 5 et 7.

- **Le graphique numéroté 5 montre les valeurs absolues :** il permet de voir d'une part quelle est la ville qui a produit le plus gros chiffre d'affaires. Mais il permet aussi de voir quelle est la contribution de chaque vendeur dans le chiffre d'affaires réalisé dans chaque ville.
- **Le graphique numéroté 7 montre les pourcentages :** il permet seulement de voir la contribution de chaque vendeur dans le CA de chaque ville. Et comme chaque barre verticale correspond à 100%, on peut comparer la performance de chaque vendeur dans chaque ville. On voit par exemple que la contribution de Leila (rouge) est plus forte à Lyon qu'à Marseille. Inversement, la contribution de Ahmed est plus forte à Marseille à qu'à Paris.

2) Barres horizontales

La deuxième ligne du tableau ci-avant intitulé « les 8 principales variétés de graphiques en barres » montre les 4 principales façons de disposer les données du tableau sous forme de barres horizontales. Comme l'axe vertical représente des catégories nominales, la hauteur des barres n'a pas d'importance, pourvu qu'elle soit identique pour toutes les barres. Cette hauteur peut varier d'un minimum (un simple trait horizontal) jusqu'à un maximum (toutes les barres sont « collées »).

a) Simples

Le type « barres horizontales simples » est représenté par le graphique numéroté 2. Il est l'équivalent en barres horizontales du graphique numéroté 1.[9]

b) Multiples

Le type « barres horizontales multiples » est représenté par le graphique numéroté 4. Il est l'équivalent en barres horizontales du graphique numéroté 3.

c) tronçonnées

Le type « barres horizontales tronçonnées» est représenté par les graphiques numérotés 6 et 8. Il est l'équivalent en barres horizontales des graphiques numérotés 5 et 7.

[9] Pour voir un autre exemple intéressant de graphique en barres horizontale, cliquer ici (JPEG)

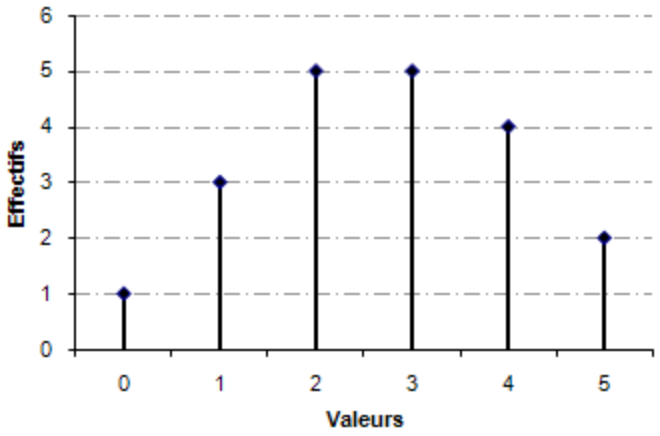
3) Graphique en bâtons

Le graphique en bâtons est un cas extrême du graphique en barres verticales. C'est en quelque sorte sa version « sobre », puisque le graphique en bâtons est en fait un graphique en barres verticales dans lequel la base des barres est réduite à une distance minimale. Au lieu d'avoir des barres verticales on a en réalité des traits verticaux.

Soit par exemple la distribution suivante :

Valeurs	Effectifs
0	1
1	3
2	5
3	5
4	4
5	2
	20

Sa représentation sous forme de diagramme en bâtons est :



Fichier EXCEL

B – Courbes et aires

La courbe est généralement le graphique le plus approprié pour montrer des évolutions dans le temps. Lorsque l'on veut montrer l'évolution de plusieurs séries dans le temps on utilisera également ce type de présentation. Parfois, on veut aussi montrer l'évolution de différentes composantes d'un ensemble dans le temps. Dans ce cas, on peut utiliser des aires délimitées par des courbes. Pour illustrer les différents usages possibles des courbes et des aires, soit le tableau suivant qui représente la répartition du chiffre d'affaires d'une entreprise sur les 3 villes où se situent ses clients.

Répartition du chiffre d'affaires de l'entreprise XXX par villes (Millions d'euros)

Années	Marseille	Paris	Lyon	Total
2003	67,2	128	53,7	248,9
2004	50,8	140,7	40,2	231,7
2005	78,4	200,1	30,6	309,1
2006	80,7	250,2	90,1	421
2007	101,4	180,6	100,1	382,1
2008	202,8	170,1	70,2	443,1
2009	305,1	280,2	60,6	645,9

Fichier EXCEL

Pour compléter ce tableau, ajoutons le tableau en pourcentage suivant, qui donne, pour chaque année, la contribution en pourcentage de chaque ville au chiffre d'affaires total. Ce second tableau s'obtient en divisant les 3 premiers chiffres de chaque ligne du tableau initial par le quatrième chiffre de la ligne correspondante et en multipliant le résultat par 100

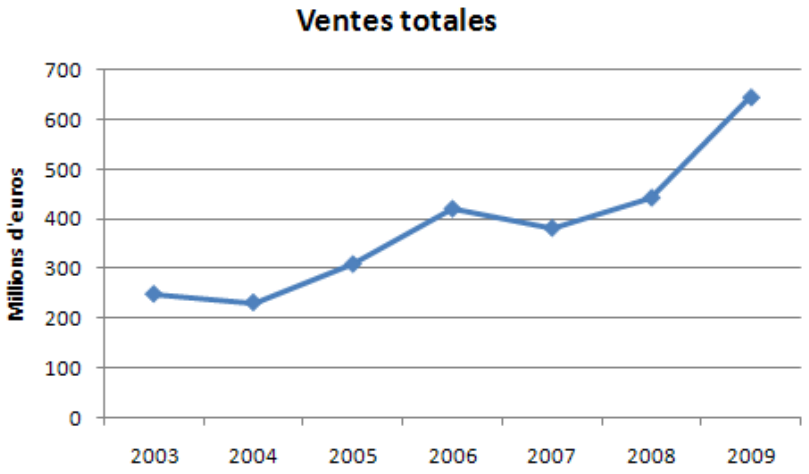
Contribution en % de chaque ville au chiffre d'affaires total de chaque année

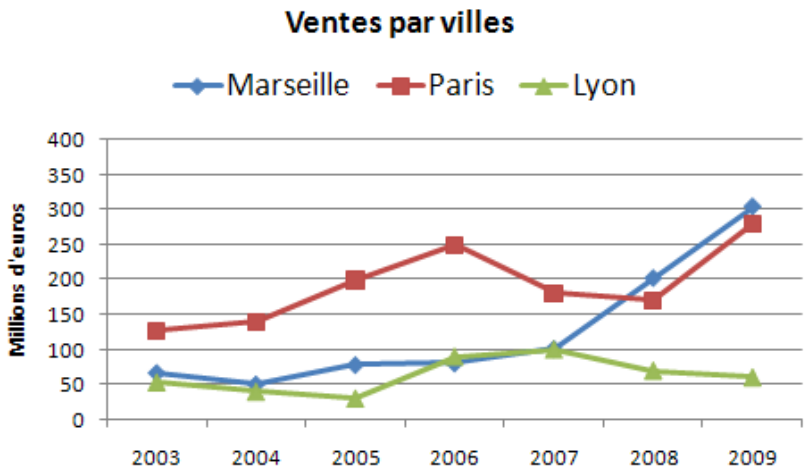
Années	Marseille	Paris	Lyon	Total
2003	27,0	51,4	21,6	100,0
2004	21,9	60,7	17,4	100,0
2005	25,4	64,7	9,9	100,0
2006	19,2	59,4	21,4	100,0
2007	26,5	47,3	26,2	100,0
2008	45,8	38,4	15,8	100,0
2009	47,2	43,4	9,4	100,0

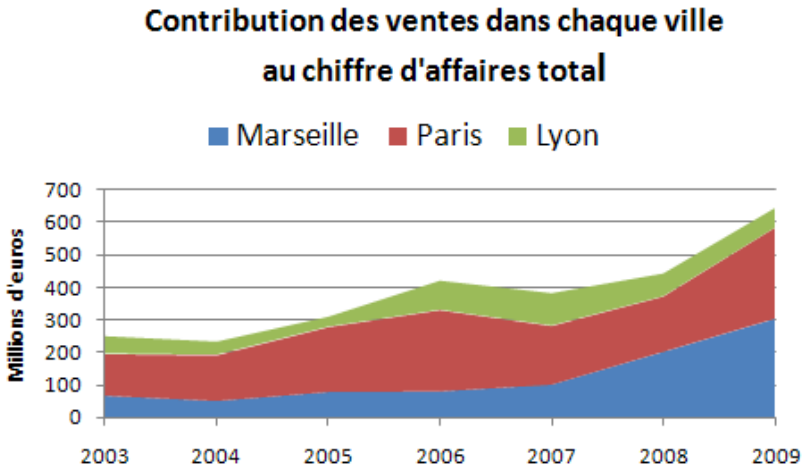
Fichier EXCEL

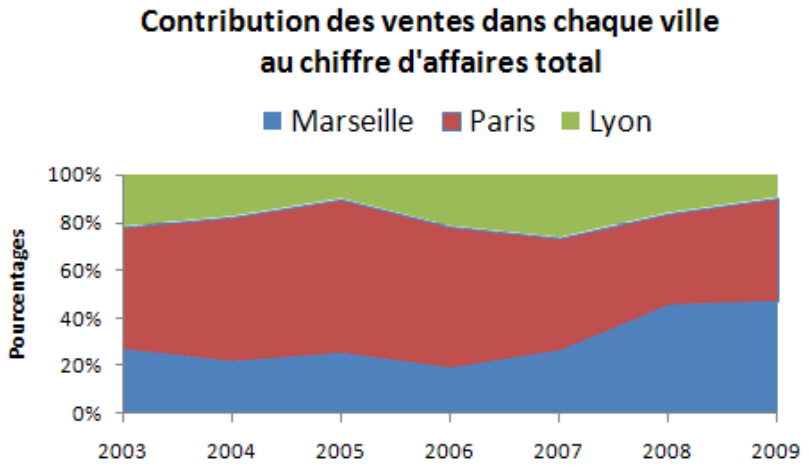
Le tableau de graphiques ci-après illustre les 4 principales possibilités d'exploiter les graphiques en ligne où les aires délimitées par des lignes.

Quatre principales façons d'utiliser les graphiques en courbes et aires









Fichier EXCEL

1) Courbes simples

Le graphique numéroté 1 représente l'évolution des ventes totales (ou chiffre d'affaires total) en millions d'euros de cette entreprise fictive. Il permet de lire à la fois l'augmentation et les périodes où l'évolution a marqué le pas. Il est parfaitement adapté pour illustrer l'évolution globale du chiffre d'affaires de l'entreprise.

2) Coubes multiples

Le graphique numéroté 2 représente l'évolution des ventes totales (ou chiffre d'affaires total) en millions d'euros que cette entreprise réalise pour chacune des 3 villes où se trouvent ses clients. Il est intéressant, mais en fait, on peut lui préférer un graphique qui allierait à la fois les avantages de la courbe simple (graphique numéroté 1) et la décomposition par ville (graphique numéroté 2). C'est ce que propose le graphique numéroté 3.

3) Aires délimitées par des courbes

On peut concevoir deux façons de présenter un graphique d'aires délimitées par des courbes :

- **Soit sous forme de valeurs absolues :** le graphique numéroté 3 représente ainsi à la fois l'évolution des ventes totales et la contribution de chaque ville à cette évolution. La contribution est représentée par le découpage en trois de la surface qui se trouve sous la courbe. La contribution de Marseille est en bleu, celle de Paris en rouge et celle de Lyon en vert. Les 3 contributions additionnées donnent l'évolution totale.

- **Soit sous forme de pourcentages :** le graphique numéroté 4 représente la contribution en pourcentage de chaque ville au chiffre total de l'entreprise. Ce graphique permet de voir par exemple que Paris était prépondérant en 2001 mais que sa part (ainsi que celle de Lyon) a été progressivement grignotée par Marseille.

La méthode de construction est simple : on part du tableau initial et l'on additionne d'abord la colonne de Marseille et de Paris, puis les colonnes Marseille, Paris et Lyon. Ensuite, on trace les 3 courbes sur un même graphique et on effectue un coloriage des 3 zones (voir schéma ci-après). Naturellement, si l'on dispose d'un tableur comme EXCEL 2007, le graphique est tracé automatiquement uniquement en sélectionnant les données du tableau initial en en choisissant le graphique adéquat.

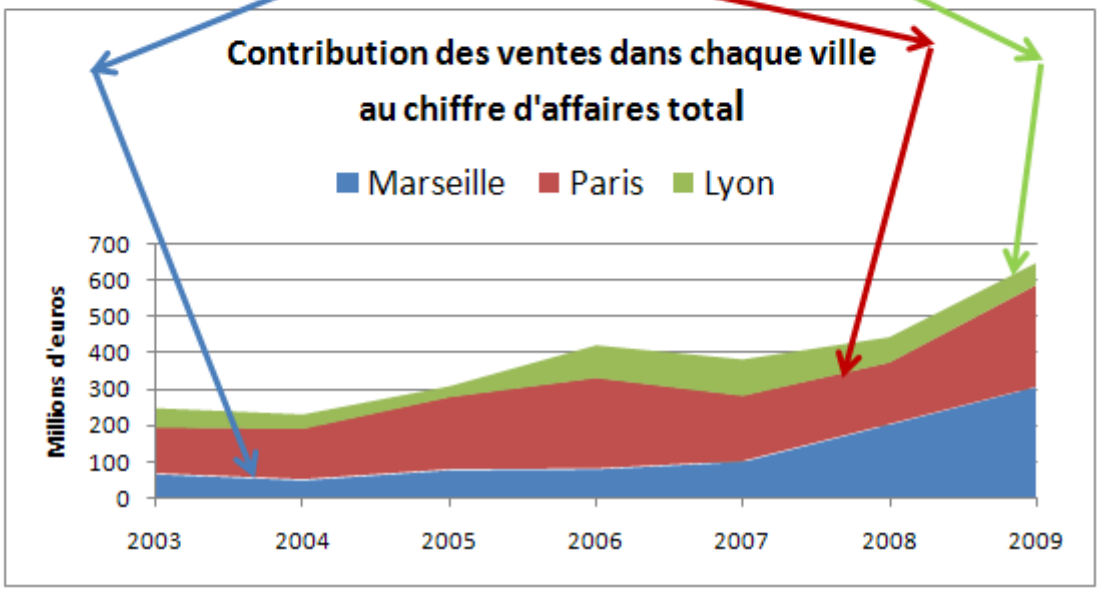
Méthode de construction d'un graphique sous forme d'aires délimitées par des courbes

Tableau initial

Années	Marseille	Paris	Lyon	Total
2003	67,2	128	53,7	248,9
2004	50,8	140,7	40,2	231,7
2005	78,4	200,1	30,6	309,1
2006	80,7	250,2	90,1	421
2007	101,4	180,6	100,1	382,1
2008	202,8	170,1	70,2	443,1
2009	305,1	280,2	60,6	645,9

Tableau final pour tracer le graphique

Années	Marseille	Marseille et Paris	Marseille, Paris et Lyon
2003	67,2	195,2	248,9
2004	50,8	191,5	231,7
2005	78,4	278,5	309,1
2006	80,7	330,9	421
2007	101,4	282	382,1
2008	202,8	372,9	443,1
2009	305,1	585,3	645,9



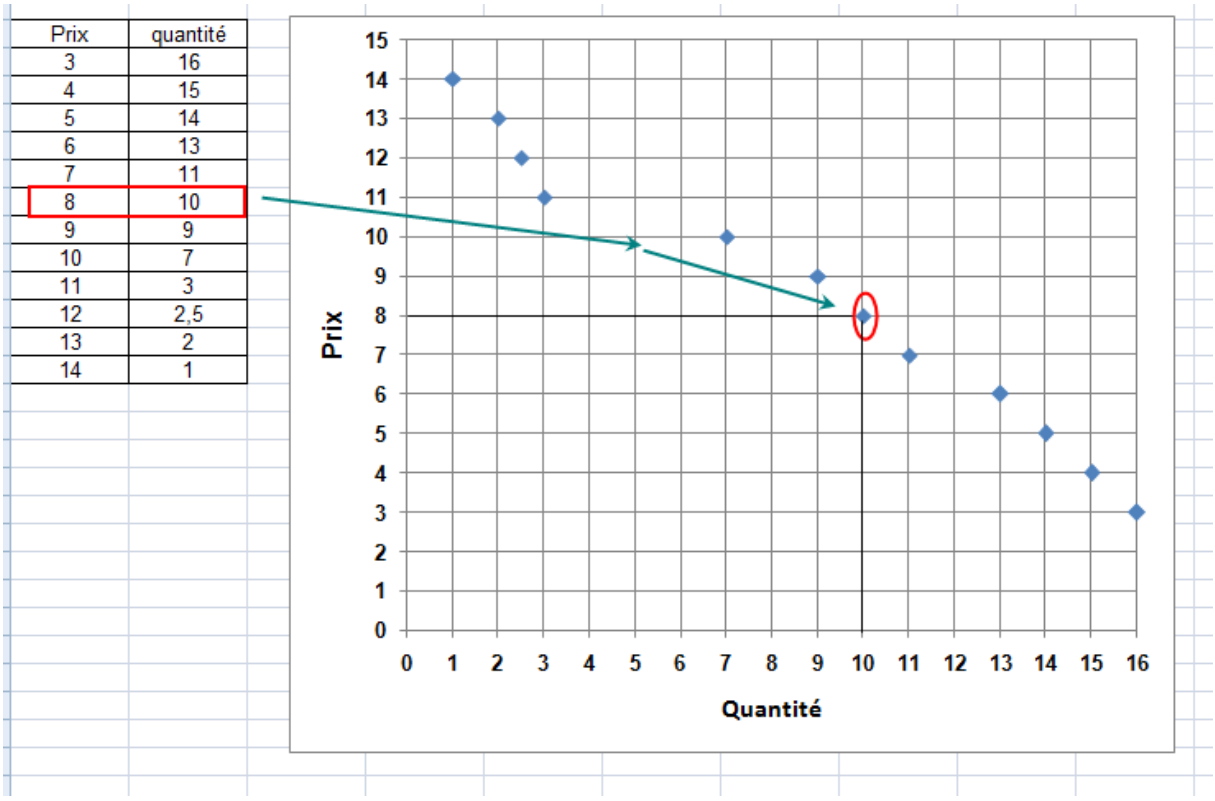
Fichier EXCEL

C - Graphique de dispersion ou nuage de points

Un graphique de dispersion ou nuage de points est un graphique qui met en relation les valeurs de deux variables sur un repère de coordonnées cartésiennes. Nous avons déjà rencontré ce type de graphique dans l'introduction à ce chapitre. En effet, les 4 graphiques du quartet d'ANSCOMBE sont des graphiques de dispersion.

Prenons un autre exemple. Ci-après, un tableau hypothétique qui illustre les différentes quantités d'un certain produit qu'un consommateur XXX est disposé à acheter suivant selon le prix du produit.

Exemple de graphique de dispersion ou « nuage de points »



Prix	quantité
3	16
4	15
5	14
6	13
7	11
8	10
9	9
10	7
11	3
12	2,5
13	2
14	1

Fichier EXCEL

Le graphique de dispersion correspondant permet d'associer à chaque point une coordonnée d'abscisse (la quantité) et une coordonnée d'ordonnée (le prix).

Les graphiques de dispersion ou en nuage de points sont très utilisés pour l'étude des corrélations entre deux variables. (Voir le chapitre 6).

D - Secteurs

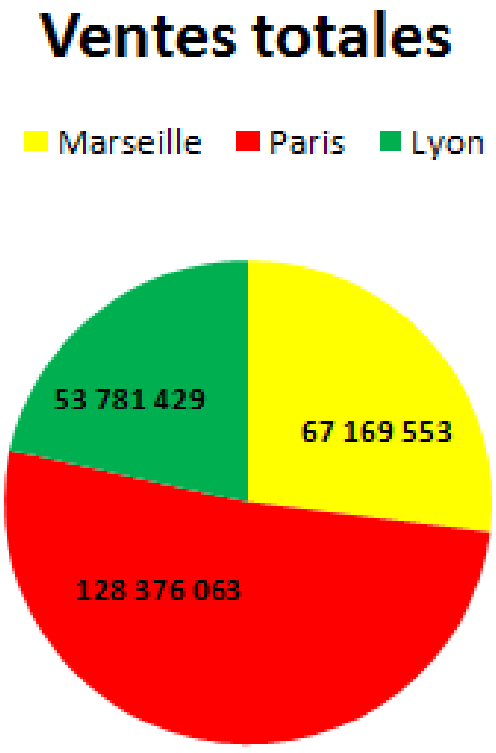
Les graphiques en secteurs sont utiles lorsque l'on veut représenter la relation entre une partie et un tout. On distingue les secteurs à 360° et ceux à 180°. Voyons un exemple de chacun d'eux avant de voir la méthode de construction qui repose sur la conversion des pourcentages en degrés.

Reprenons les données du chiffre d'affaires hypothétique qu'une entreprise a réalisé en 2009 (249 327 045 euros) répartis par ses 4 vendeurs et dans les trois villes où se trouvent ses clients.

Ventes 2009 (euros)				
	Leila	Ahmed	Pierre	Elodie
Lyon	18 753 951	8 754 668	9 785 246	16 487 564
Marseille	13 225 478	20 154 287	17 892 555	15 897 233
Paris	37 895 214	35 877 421	32 558 741	22 044 687
	69 874 643	64 786 376	60 236 542	54 429 484

1) Secteurs à 360 degrés

Le secteur à 360° ci-dessous représente la répartition des ventes totales entre les trois villes (Marseille en jaune, Paris en rouge et Lyon en vert).



Fichier EXCEL

La méthode de construction est la suivante. Pour obtenir la part du chiffre d'affaires réalisé dans chaque ville, on divise le chiffre d'affaires réalisé dans la ville par le chiffre d'affaires total et on multiplie par 360.

Par exemple, pour Marseille on a :

$$\frac{67169553}{249327045} \times 360 = 97$$

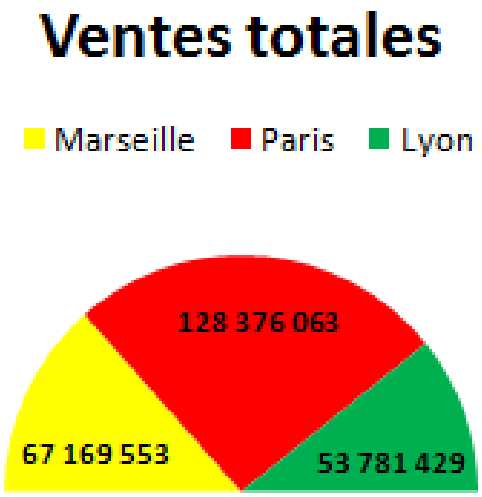
Le tableau ci-dessous donne les résultats en degrés pour les trois villes :

	Total	fréquences	degrés
Marseille	67169553	0,27	97.0
Paris	128376063	0,51	185.3
Lyon	53781429	0,22	77,7
Total	249327045	1,00	360,0

Une fois que l'on a calculé les degrés associés au chiffre d'affaire dans chaque ville, il faut tracer le secteur au moyen d'un compas, puis, avec un rapporteur, le diviser en 3 sous-secteurs ayant pour angle 97° (Marseille), 185,3° (Paris) et 77,7°(Lyon).

2) Secteurs à 180 degrés

Le secteur à 180° ci-dessous représente la répartition des ventes totales entre les trois villes (Marseille en jaune, Paris en rouge et Lyon en vert). Pour le faire avec EXCEL, voir lien internet ici.



Fichier EXCEL

La méthode de construction est la suivante. Pour obtenir la part du chiffre d'affaire réalisé dans chaque ville, on divise le chiffre d'affaires réalisé dans la ville par le chiffre d'affaires total et on multiplie par 180.

Par exemple, pour Marseille on a :

$$\frac{67169553}{249327045} \times 180 = 48,5$$

Le tableau ci-dessous donne les résultats en degrés pour les trois villes :

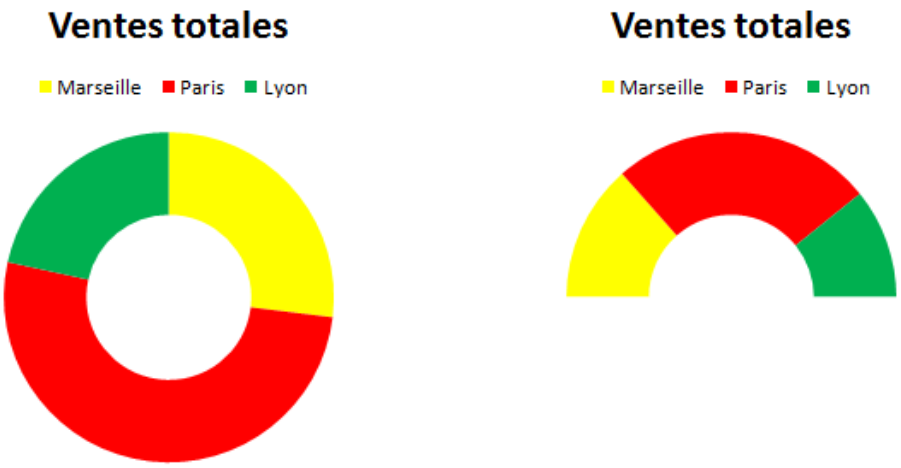
	Total	fréquences	degrés
Marseille	67 169 553	0,27	48,5
Paris	128 376 063	0,51	92,7
Lyon	53 781 429	0,22	38,8
Total	249 327 045	1,00	180,0

Une fois que l'on a calculé les degrés associés au chiffre d'affaires dans chaque ville, il faut tracer le secteur au moyen d'un compas, puis, avec un rapporteur, le diviser en 3 sous-secteurs ayant pour angle 48,5° (Marseille), 92,7° (Paris) et 38,8 degrés (Lyon). Pour le faire avec EXCEL, voir lien internet ici.

3) Anneaux

a) Simples

Les anneaux sont simplement des secteurs « troués » au milieu. Ainsi, par exemple, les deux secteurs précédents peuvent être représentés sous forme d'anneaux comme ci-après. La seule différence étant le cercle vide du milieu dont la taille peu être choisie de façon arbitraire en fonction de considération telles que l'esthétique ou l'économie d'encre (si le graphique est destiné à être imprimé et que l'on doit payer soi-même la cartouche d'encre) par exemple[10].



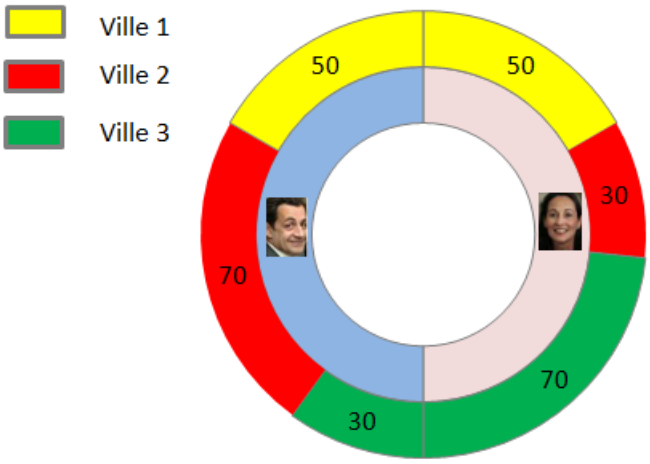
Fichier EXCEL

[10] Certains se dédouanent de leur avarice en étant économe avec l'argent des autres, même l'argent public (c'est vrai qu'il vient des maigres impôts qu'ils paient – quand ils en paient !).

b) Concentriques

Un exemple typique d'**anneaux concentriques** est celui de la représentation des résultats du second tour de l'élection présidentielle de 2007 pour 3 villes hypothétiques. Dans le cercle intérieur, on a la répartition des voix entre les deux candidats (dans cet exemple hypothétique, ils ont obtenu chacun 150 voix) et dans le cercle extérieur, on a la répartition des voix de chaque candidat dans chacune des villes.

	Voix
Ségo	**150**
Ville 1	50
Ville 2	30
Ville 3	70
Sarko	**150**
Ville 3	30
Ville 2	70
Ville 1	50
Total	**300**



5 – Autres graphiques

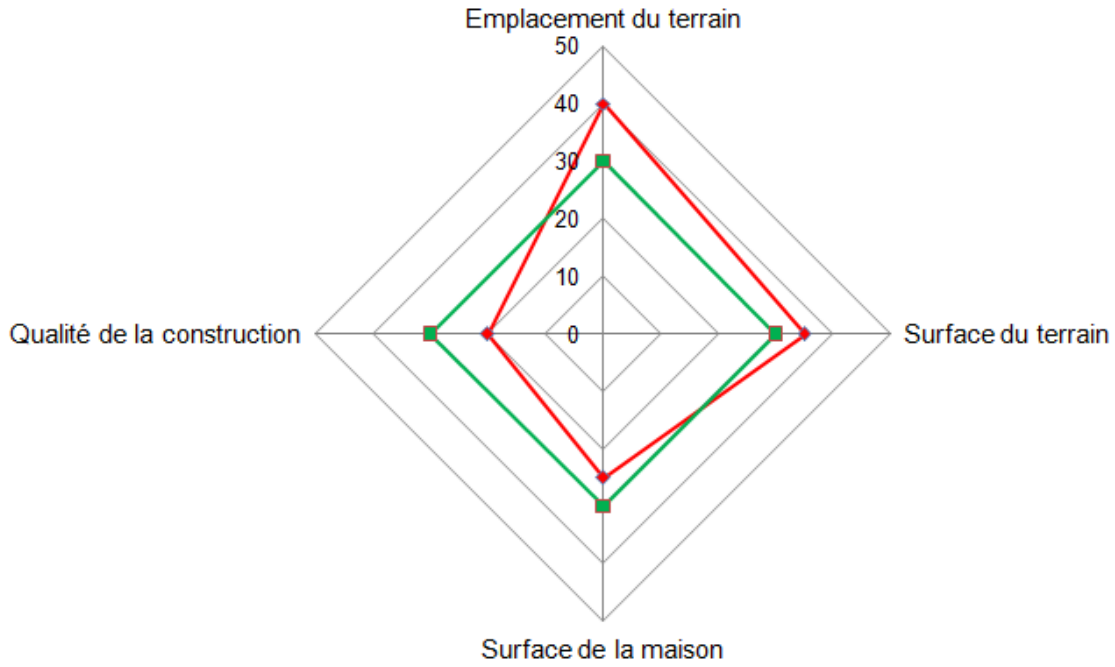
A – Graphiques en radar et toile d'araignée

1 – Radar

Soit par exemple 120 personnes qui sont interrogées dans la ville A et 120 personnes dans la ville B sur la caractéristique qu'elles placent en premier lors de l'achat d'une maison. Il s'agit d'une caractéristique avec 4 modalités. On a le tableau suivant :

	Ville A	Ville B
Modalités	Nombre de personnes	Nombre de personnes
Emplacement du terrain	40	30
Surface du terrain	35	30
Surface de la maison	25	30
Qualité de la construction	20	30
	120	120

On peut alors placer ces données sur un diagramme « en radar » de la façon suivante :



Fichier EXCEL

2 – Toile d'araignée

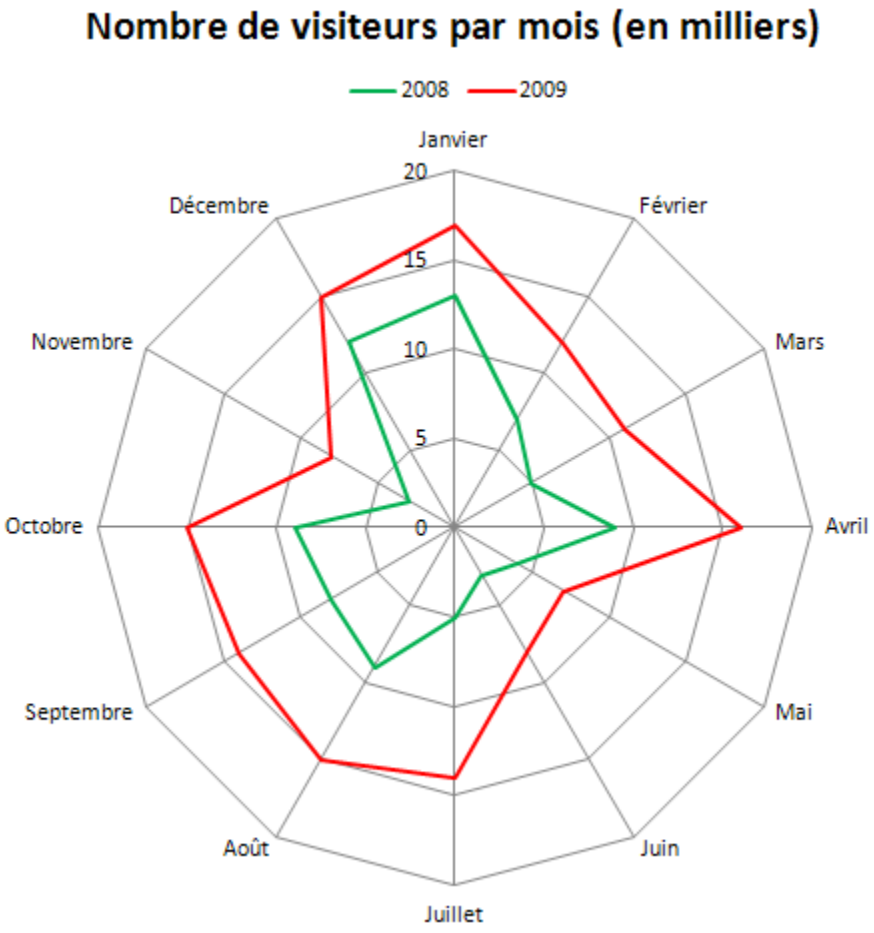
Le graphique en toile d'araignée est une variante du graphique en radar, mais avec un nombre d'axes plus grand. On l'utilise par exemple pour représenter des données chronologiques. Soit par exemple une entreprise qui souhaite comparer le nombre de visiteurs mensuels sur son site internet en 2008 et en 2009.

Nombre de visiteurs mensuels sur le site internet de l'entreprise XXX (milliers)

	2008	2009
Janvier	13	17
Février	7	12
Mars	5	11
Avril	9	16
Mai	4	7
Juin	3	8
Juillet	5	14
Août	9	15
Septembre	8	14
Octobre	9	15
Novembre	3	8
Décembre	12	15

Fichier EXCEL

Une bonne façon de présenter une comparaison visuelle est de faire le graphique en « toile d'araignée» suivant :



.

Fichier EXCEL

Ce graphique donne immédiatement deux informations :

- Les visites en 2009 ont chaque mois été supérieures aux visites en 2008
- Il y a un caractère cyclique dans les visites, car les mois « creux » et les mois « pleins » sont les mêmes en 2008 et en 2009

B – Graphiques à bulles

Semblable au graphique de dispersion ou nuages de points, le **graphique (ou diagramme) en bulles** permet d'ajouter une troisième dimension à l'analyse. Les deux premières dimensions déterminent la position des bulles sur le diagramme tandis que la troisième fixe la surface de chacune des bulles.

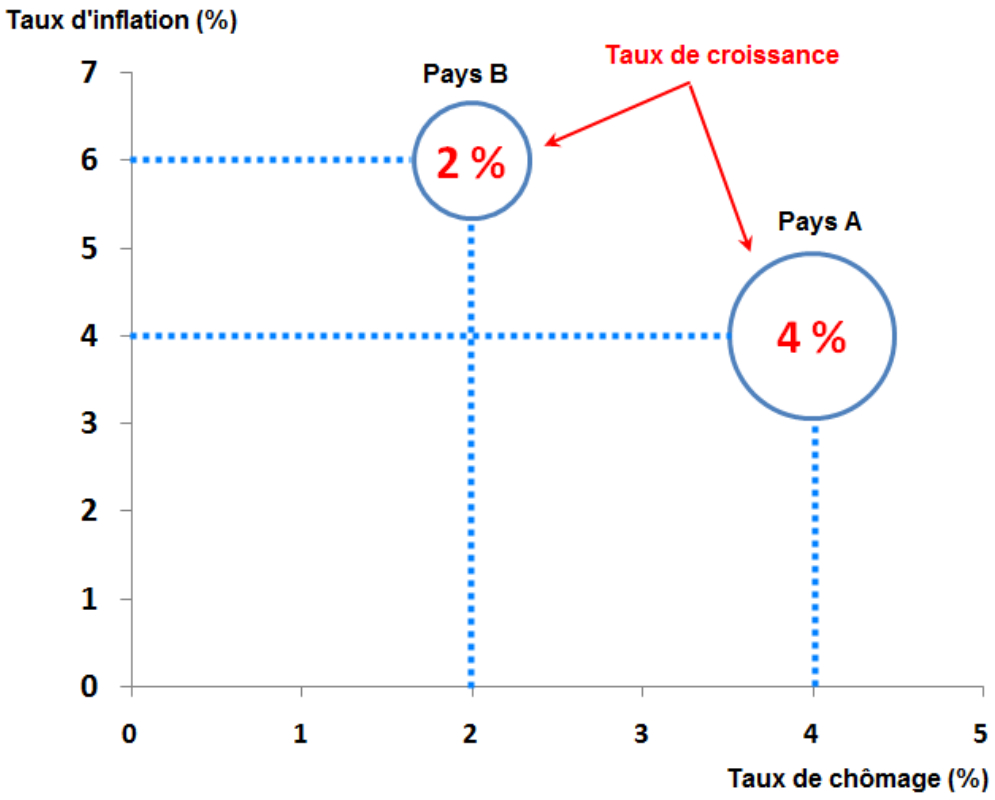
Exemple : Le tableau ci-dessous donne les performances macroéconomiques de 2 pays hypothétiques A et B, en 2009.

	Chômage (%)	Inflation (%)	Croissance (%)
Pays A	4	4	4
Pays B	2	6	2

Représenter ces chiffres sur un graphique « en bulles » avec :

- Le taux de chômage en abscisse
- Le taux d'inflation en ordonnée
- La croissance par un cercle de surface proportionnelle au taux du pays considéré.

On obtient alors le graphique suivant :



Dans cet exemple relativement simple, il suffit de faire en sorte que l'aire du disque qui représente le taux de croissance du pays A soit le double de l'aire qui représente le taux de croissance du pays B.

Pour assurer la proportionnalité des aires, il faut passer par la formule de l'aire du disque :

$$S = \pi R^2$$

On peut choisir pour référence la plus grosse valeur à représenter sous forme de disque.

Ensuite on lui attribue une aire arbitraire, par exemple 2 cm^2 et on en déduit le rayon :

$$S = \pi R^2 \Leftrightarrow 2 = 3,14 \times R^2 \Leftrightarrow R^2 = \frac{2}{3,14} \Leftrightarrow R = \frac{\sqrt{2}}{\sqrt{3,14}} = \frac{1,414}{1,772} = 0,798$$

Pour tracer le cercle, on arrondira le rayon à 8 mm.

On calcule ensuite l'aire de l'autre disque du graphique. Dans notre exemple, si l'aire du disque du pays A représente 4%, le disque du pays B, qui représente 2% doit avoir une aire 2 fois plus petite. Par conséquent, l'aire du disque représentatif de la croissance du pays B sera égale à la moitié de l'aire du disque qui représente la croissance de A, soit 1 cm^2. On en déduit ensuite le rayon du disque de B par la formule :

$$S = \pi R^2 \Leftrightarrow 1 = 3,14 \times R^2 \Leftrightarrow R^2 = \frac{1}{3,14} \Leftrightarrow R = \frac{\sqrt{1}}{\sqrt{3,14}} = \frac{1}{1,772} = 0,56$$

Pour tracer le cercle, on arrondira le rayon à 5,5 mm.

En définitive, le disque de A aura un rayon de 8 mm (environ) et le disque de B aura un rayon de 5,5 mm (environ). Ceci n'est pas facile à tracer de manière précise ! C'est la raison pour laquelle les graphiques à bulles sont généralement réalisés avec un logiciel.

C – Graphiques boursiers

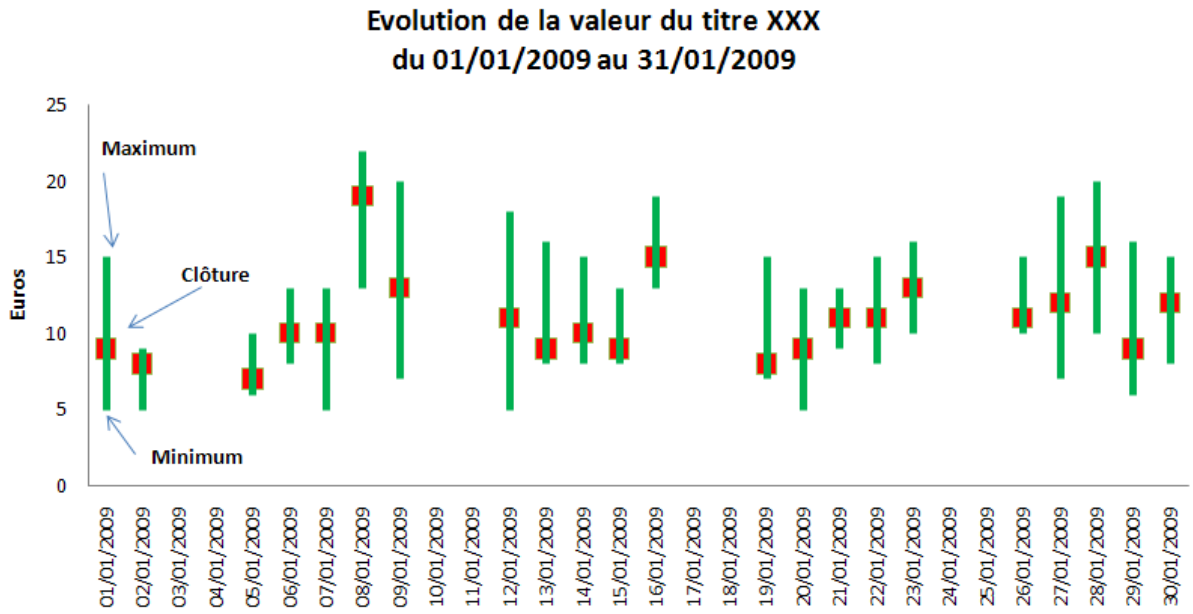
Les graphiques boursiers sont appelés ainsi car ils servent principalement à donner des indications sur l'évolution des cours boursiers. Dans l'exemple ci-après, nous allons voir la version la plus simple du graphique boursier, mais des versions plus complexes sont possibles et facilement réalisables dans EXCEL 2007 un fois que l'on a compris le principe de base.

Soit le tableau ci-après qui donne l'évolution du cours journalier d'un titre boursier (en euros) du 2 janvier au 31 janvier 2009, en ne retenant que les jours ouvrables. On a relevé 3 informations chaque jour : le cours le plus bas, le cours le plus haut et le cours de clôture.

Evolution du cours du titre XXX

Date	Cours le plus bas	Cours le plus haut	Clôture
01/01/2009	5	15	9
02/01/2009	5	9	8
03/01/2009			
04/01/2009			
05/01/2009	6	10	7
06/01/2009	8	13	10
07/01/2009	5	13	10
08/01/2009	13	22	19
09/01/2009	7	20	13
10/01/2009			
11/01/2009			
12/01/2009	5	18	11
13/01/2009	8	16	9
14/01/2009	8	15	10
15/01/2009	8	13	9
16/01/2009	13	19	15
17/01/2009			
18/01/2009			
19/01/2009	7	15	8
20/01/2009	5	13	9
21/01/2009	9	13	11
22/01/2009	8	15	11
23/01/2009	10	16	13
24/01/2009			
25/01/2009			
26/01/2009	10	15	11
27/01/2009	7	19	12
28/01/2009	10	20	15
29/01/2009	6	16	9
30/01/2009	8	15	12
31/01/2009			

Le graphique ci-après, dit « graphique boursier », permet de visualiser les 3 informations. Comme c'est un exemple hypothétique, on a pris soin d'éviter les cas où les maximum et/ou minimum peuvent se confondre avec le cours de clôture. On voit aussi qu'il serait possible d'ajouter d'autres informations, tel que par exemple le cours à l'ouverture. À la limite, chaque barre verticale pourrait être remplacée par un graphique de TUKEY (voir la section qui suit) :



Fichier EXCEL

D - Graphiques de TUKEY

Le graphique « Boite à moustaches » ou diagramme en boîte (box plot) a été inventé en 1977 par le statisticien américain John TUKEY (1915-2000). Inconnu en France, il a été imposé (dans sa version simplifiée) depuis quelques années dans les programmes de mathématiques de l'enseignement secondaire, prenant par surprise les enseignants universitaires qui jusque là s'en souciait comme d'une guigne. Il est pourtant extrêmement utile et d'ailleurs très utilisé en pratique (contrairement à des pans entiers de certains livres de statistique descriptive).

1) Les éléments constitutifs du graphique original

Sur le graphique ci-après, les éléments suivants apparaissent :

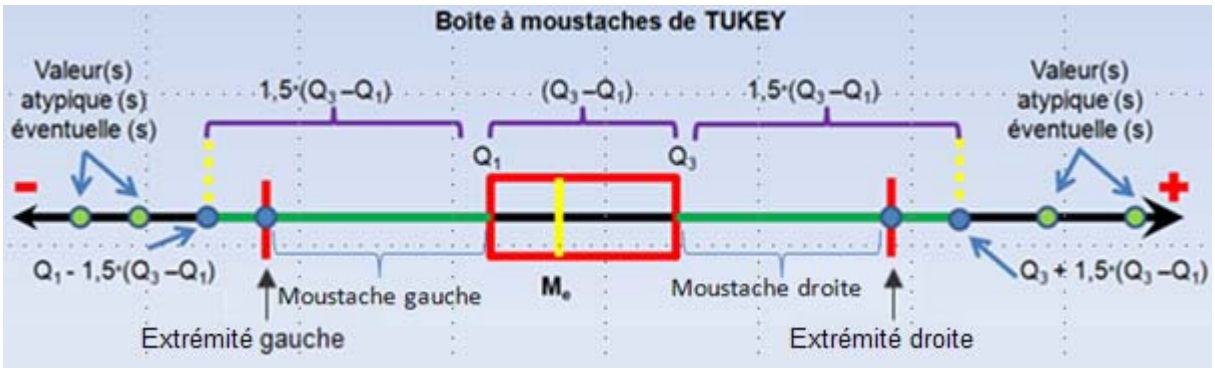
L'intervalle interquartile : il est défini par $Q_3 - Q_1$ et donne les "dimensions" de la boîte (le mode de calcul des quartiles est étudié dans le chapitre 3).

Les extrémités ou « pattes » : Ce sont les extrémités gauche et droite du graphique.

> **L'extrémité gauche** est définie par la **valeur de la série** immédiatement supérieure à $Q_1 - 1{,}5\,(Q_3 - Q_1)$. S'il y a des valeurs inférieures à $Q_1 - 1{,}5\,(Q_3 - Q_1)$ dans la série, elles sont atypiques et représentées par des marqueurs.

L'extrémité droite est définie par la **valeur de la série** immédiatement inférieure à $Q_3 + 1,5 (Q_3 - Q_1)$. S'il y a des valeurs supérieures à $Q_3 + 1,5 (Q_3 - Q_1)$, elles sont représentées par des marqueurs.

La médiane : la valeur de la variable qui partage la population en deux populations égales. On la désigne par l'abréviation Me (le mode de calcul de la médiane est étudié dans le chapitre 3).



2) Exemple[11]

Soit la série des 20 éléments : {4, 0, 1, 1, 2, 2, 2, 3, 3, 4, 2, 3, 4, 5, 2, 1, 3, 3, 4, 5}

La médiane est égale à 3. De plus, $Q_1 = 2$ et $Q_3 = 4$. Enfin, la valeur minimale est 0 et la valeur maximale 5. Au risque de surcharger le graphique, on peut ajouter la moyenne de la série qui est égale à 2,7.

Valeurs atypiques : Pour savoir s'il y a des valeurs atypiques il faut calculer $Q_1 - 1,5 (Q_3 - Q_1) = 2 - 1,5 \times (4 - 2) = 2 - 1,5 \times 2 = 2 - 3 = -1 < 0$ et $Q_3 + 1,5 (Q_3 - Q_1) = 4 + 1,5 \times (4 - 2) = 4 + 3 = 7 > 5$. Conclusion : puisque $-1 < 0$ et que $7>5$, n'y a pas de valeurs atypiques[12]. Par conséquent, puisque la valeur immédiatement supérieure à $Q_1 - 1,5 (Q_3 - Q_1)$ est 0 et que la valeur immédiatement inférieure à $Q_3 + 1,5 (Q_3 - Q_1)$ est 5, les deux moustaches ont respectivement pour extrémité gauche le minimum de la série (soit 0) et pour extrémité droite le maximum de la série (soit 5).

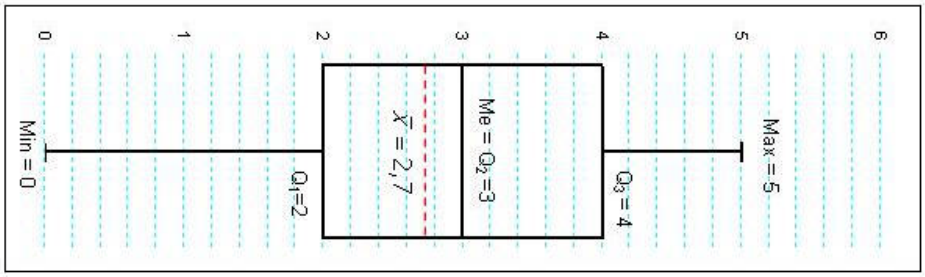
On a donc le graphique boite à moustaches ci-après. La boite est délimitée par Q_1 et Q_3. À l'intérieur de la boîte un premier trait noir indique la médiane (et un second trait rouge (en pointillé et facultatif) indique la moyenne (ce n'est pas le cas dans le graphique de TUKEY original, mais nous l'avons ajouté ici pour souligner qu'il peut

[11] Pour tracer une boite à moustache avec EXCEL 2007, voir : http://blog.immeria.net/2007/01/box-plot-and-whisker-plots-in-excel.html . Voir aussi : http://www.bloggpro.com/box-plot-for-excel-2007/ (mais remplacer les valeurs Min et Max par les valeurs adéquates si la distribution contient des valeurs atypiques). Enfin, voir aussi http://www.coventry.ac.uk/ec/~nhunt/boxplot.htm qui explique la construction d'une boite à moustaches avec EXCEL 97, 2003 et 2007.

[12] Monique LE GUENN indique que « *La valeur 1.5 est selon TUKEY une valeur pragmatique qui a une raison probabiliste. Si une variable suit une distribution normale, alors la zone délimitée par la boîte et les moustaches devrait contenir 99,3 % des observations. On ne devrait donc trouver que 0.7% d'observations atypiques. Si le coefficient vaut 1, la probabilité serait de 0.957, et elle vaudrait 0.999 si le coefficient est égal à 2. Pour TUKEY la valeur 1.5 est donc un compromis pour retenir comme atypiques assez d'observations mais pas trop d'observations* ». http://matisse.univ-paris1.fr/leguen/leguen2001b.pdf

être utile de comparer visuellement la position de la moyenne par rapport à la médiane, voire par rapport au mode, sachant que ni la moyenne, ni le mode, à l'inverse de la médiane, ne sont nécessairement compris entre Q_1 et Q_3). Les valeurs minimale et maximale constituent ici les extrémités des moustaches, car elles sont comprises dans l'intervalle défini par $Q_1 - 1{,}5\ (Q_3 - Q_1)$ à gauche et par $Q_3 + 1{,}5\ (Q_3 - Q_1)$ à droite.

Boite à moustaches de TUKEY pour la série {4, 0, 1, 1, 2, 2, 2, 3, 3, 4, 2, 3, 4, 5, 2, 1, 3, 3, 4, 5}, série utilisée dans le chapitre 3 pour illustrer les calculs des quartiles, de la médiane et de la moyenne



Voir le fichier EXCEL

3) Graphique de TUKEY et forme de la distribution sous-jacente

Le graphique de TUKEY peut être utilisé pour connaître la forme d'une distribution sans avoir à la tracer. On comparera pour cela :

- la position de la médiane par rapport au **milieu de l'intervalle de variation** (minimum de la série augmenté de la moitié de l'**intervalle de variation**).
- La longueur des deux moustaches (mesurée par l'écart en valeur absolue entre chaque extrémité et le quartile qui lui correspond).

a) Aplatissement

La position de la **médiane** par rapport au **milieu de l'intervalle de variation** permet de savoir si la distribution sous-jacente est plus aplatie à droite qu'à gauche ou si elle est symétrique.

On aura ainsi 3 cas de figure :

- Plus la médiane est à droite du milieu de l'intervalle de variation et plus la partie gauche du graphique de la distribution est aplatie (cas A ci-après).
- Plus la médiane est à gauche du milieu de l'intervalle de variation et plus la partie droite du graphique de la distribution est aplatie (cas B ci-après).
- Plus la médiane est proche du milieu de l'intervalle de variation, plus le graphique de la distribution est symétrique, c'est-à-dire pas plus aplatie à gauche qu'à droite (cas C ci-après).

b) Etalement

La comparaison des longueurs de moustaches permet de savoir si la distribution sous-jacente est plus étalée à droite qu'à gauche ou si elle est symétrique.

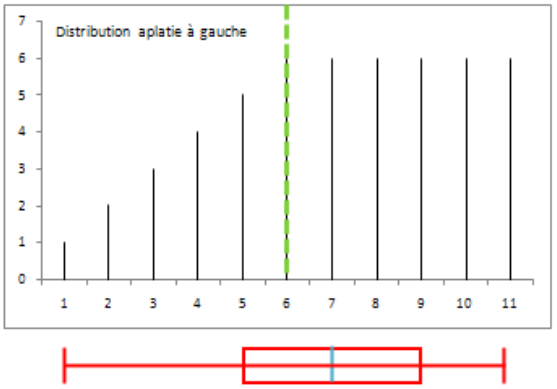
- La longueur en valeur absolue de la moustache de gauche est égale à Q1-extrémité gauche
- La longueur en valeur absolue de la moustache de droite est égale à Q3-extrémité droite

On aura donc 3 cas de figure :

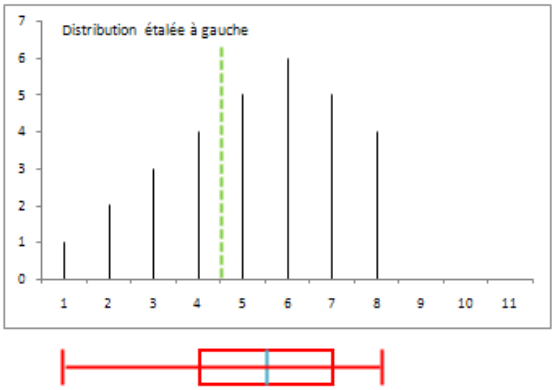
- La partie gauche du graphique de la distribution est d'autant plus étalée que la moustache gauche est plus longue que la moustache droite (cas D ci-après).
- La partie droite du graphique de la distribution est d'autant plus étalée que la moustache droite est plus longue que la moustache gauche (cas E ci-après).
- Quand la distribution est symétrique, les deux moustaches sont de longueurs égales.

Graphique de TUKEY et forme de la distribution sous-jacente (en bleu, la médiane et en vert, le milieu de l'intervalle de variation)

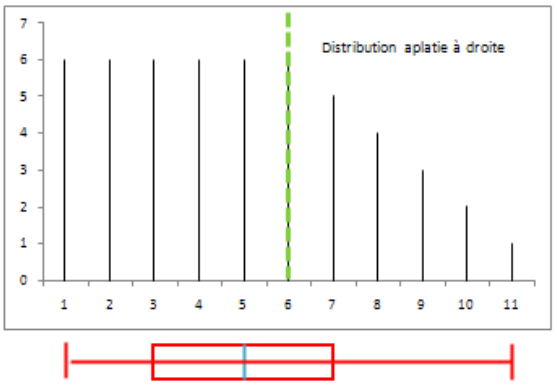
A



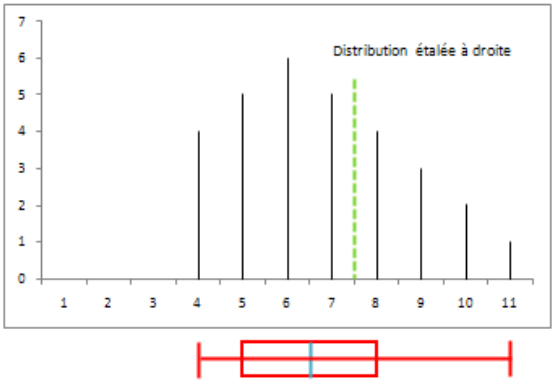
D



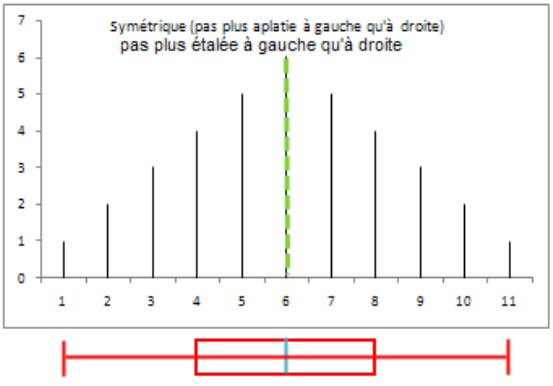
B



E



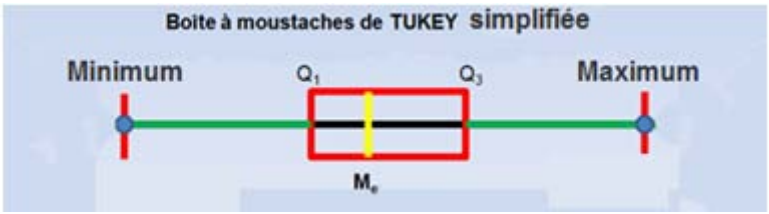
C



Fichier EXCEL

4) Le graphique de TUKEY simplifié

La plupart des logiciels ne tiennent pas compte de la possibilité de valeurs atypiques dans une série et se contentent de prendre pour extrémités les valeurs minimum et maximum de la série. On obtient alors un graphique de TUKEY simplifié, mais vidé de sa substance, comme ci-après.



En outre, même cette version simplifiée ne fait pas l'unanimité, en raison des divers algorithmes d'interpolation employés pour calculer Q_1 et Q_3. On notera cependant, comme indiqué au chapitre 3, que plus la série contient de chiffres, plus les méthodes de calcul des quartiles donnent des résultats identiques. Pour les très grandes séries, les dfférences entre méthodes deviennent négligeables.

E – Graphiques panachés

Il existe une infinité de façons de panacher les différents graphiques. Voyons quelques exemples.

1) Secteur complété par une barre tronçonnée

Soient les données déjà utilisées du chiffre d'affaires par ville et par vendeur d'une entreprise XXX en 2009.

Ventes 2009 (euros)					
	Leila	Ahmed	Pierre	Elodie	Total
Marseille	13 225 478	20 154 287	17 892 555	15 897 233	67 169 553
Paris	37 895 214	35 877 421	32 558 741	22 044 687	128 376 063
Lyon	18 753 951	8 754 668	9 785 246	16 487 564	53 781 429
Total	69 874 643	64 786 376	60 236 542	54 429 484	**249 327 045**

Ventes 2009 (%)					
	Leila	Ahmed	Pierre	Elodie	Total
Marseille					27
Paris					51
Lyon					22
Total					100

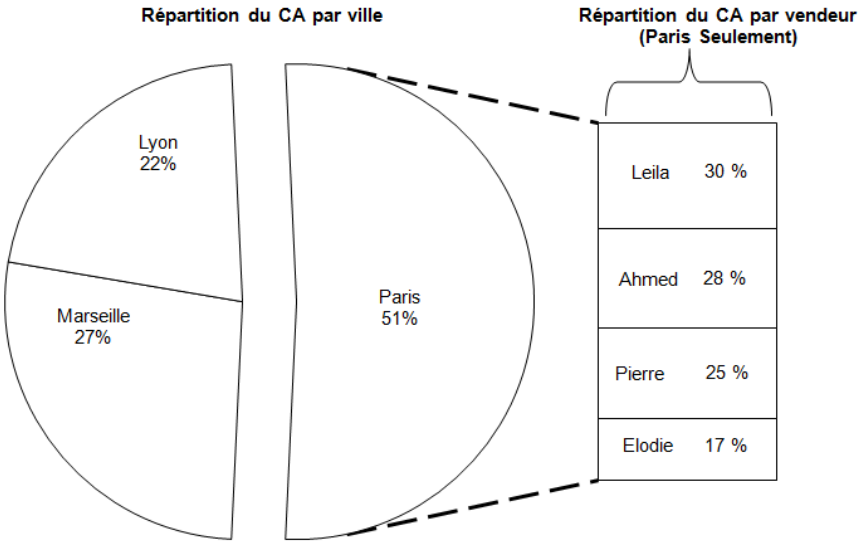
Ventes 2009 (%)					
	Leila	Ahmed	Pierre	Elodie	Total
Marseille					
Paris	30	28	25	17	100
Lyon					
Total					

Fichier EXCEL

On souhaite :

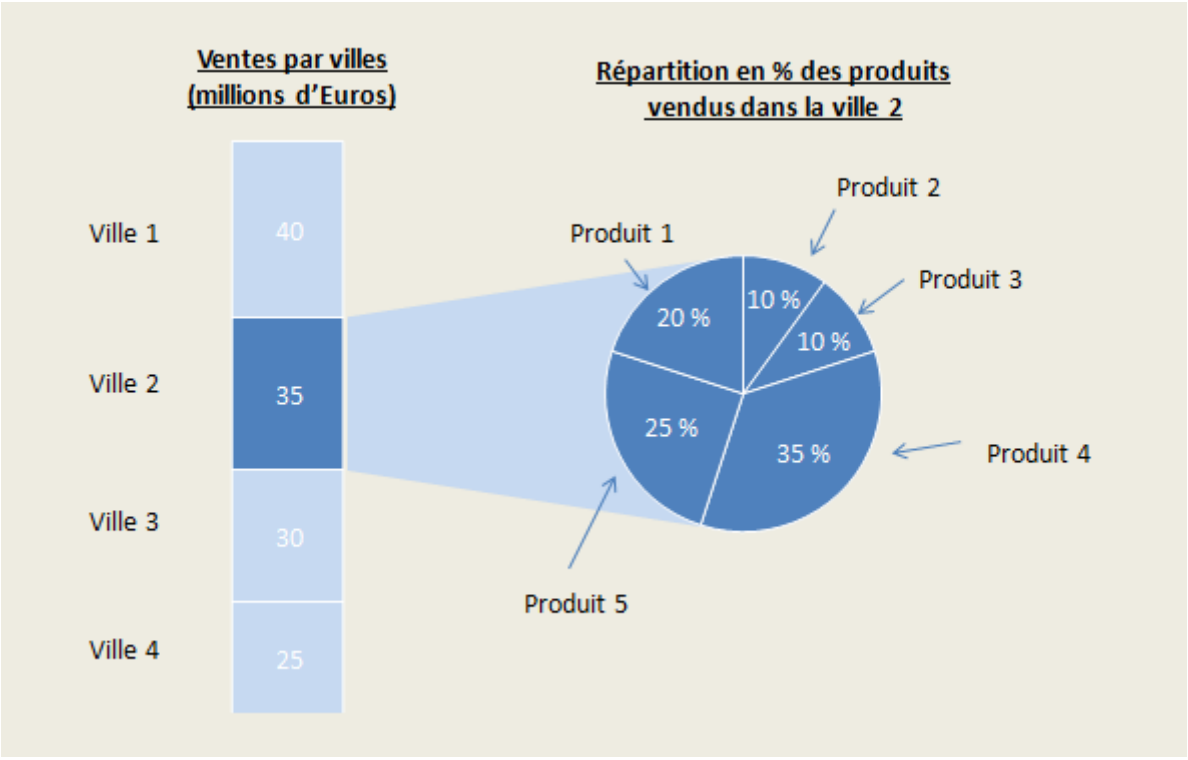
- utiliser un secteur à 360° pour faire apparaitre la répartition du chiffre d'affaires entre les trois villes.
- Décomposer le chiffre d'affaires réalisé sur Paris entre les quatre vendeurs et le faire apparaître au moyen d'une barre verticale tronçonnée à 100%.

On aura alors recours au graphique ci-après :



Un autre exemple de graphique panaché, symétrique du précédent, consiste à faire une barre tronçonnée dont un des « tronçons » est lui-même décomposé en secteur. Voici un exemple réalisé avec PowerPoint 2007.

Un autre exemple de graphique panaché



D'après Kathy VILLELA, "Graphic elements that connect two objects" http://www.powerframeworks.com/makeyourpoint/?p=73 . Télécharger le graphique au format PowerPoint 2007.

2) Graphique de PARETO

Le graphique de PARETO est un cas particulier du graphique à échelle verticale double. Avant de voir comment il se construit, il convient de rappeler la « loi de PARETO » dont il sert à vérifier la validité. Le graphique a en fait été inventé par Joseph Moser DURAN (1904 -), fondateur de la démarche qualité. Le graphique de PARETO est en effet l'un des sept graphiques de base utilisé dans les contrôles de qualité. S'il est appelé graphique de PARETO et non graphique de DURAN (ou graphique de Kaoru ISHIKAWA [1915-1989]), autre fondateur du management de la qualité), c'est en raison de la « loi de PARETO » que nous allons donc exposer pour commencer.

a) De la loi de PARETO au graphique de PARETO

Vilfredo PARETO, économiste et sociologue de la fin du 19ème siècle et du début du 20ème siècle (1848-1923) a notamment étudié la répartition des revenus dans différents pays. Une des principaux constats qui ressort de cette étude est une "loi", dite "**Loi de PARETO**": **dans presque tous les pays, 80% de la richesse sociale était possédée par 20% des individus**[13]. Par la suite, on s'aperçut qu'un grand nombre de phénomènes étaient gouvernés par la loi du 80/20 :

20 % des clients représentent 80 % du chiffre d'affaires
20% des pièces stockées dans une entreprise représentent 80% de la valeur du stock
20% des fournisseurs représentent 80% du volume d'achat total
20% du personnel effectue 80% du travail
20% des salariés d'une entreprise empoche 80% de la masse salariale
20% des automobilistes causent 80% des accidents
20% des vêtements d'une garde-robe sont portés 80% du temps
20% des produits représentent 80% du chiffre d'affaires
20% des ventes représentent environ 80% de la marge bénéficiaire
20% du personnel total est concerné par 80% des accidents du travail
20% des causes peuvent être à l'origine de 80% des défauts
20% des clients sont à l'origine de 80% des réclamations
20% de la population paie 80% des impôts
20% des délinquants génèrent 80% des délits
20% des joueurs de foot marquent 80% des buts
20% des problèmes représentent 80% des préoccupations
20% des pays émettent 80% des gaz à effet de serre.
20% des acteurs jouent dans 80% des films.
20% des personnes imposables génèrent 80% des recettes fiscales
20% de l'effort de l'entraînement sportif permet d'atteindre 80% de la performance
20% des indicateurs de gestion fournissent 80% des informations pertinentes, etc.

Les exemples précédents illustrent la loi de PARETO. Quelques causes majeures, une fois isolées, permettent de résoudre la plus grande partie d'un problème de

[13] PARETO présenta cette loi en 1906 et l'exprima d'abord ainsi : 20% de la population italienne détient 80% des richesses.

qualité[14]. Une fois ces causes majeures identifiées, on peut concentrer les efforts et les ressources à les éliminer. Le **graphique de PARETO** est une façon de visualiser la loi des 80/20. Mais, comme on va le voir dans l'exemple étudié plus loin, cette loi n'est pas systématique, il arrive très souvent aussi que 20% des causes expliquent beaucoup moins que 80% des résultats.

b) Définition, construction, exemple et interprétation

Un **diagramme de PARETO** est un graphique qui combine un **graphique en barre** et une **courbe cumulative**. Il sert principalement à l'étude des données qualitatives, mais rien n'empêche de l'utiliser pour des données quantitatives.

Procédure pour construire le graphique :

i) Si les données sont sous forme d'une série, les regrouper par **modalités** (données) ou par **valeurs** (données quantitatives), de façon à obtenir une **distribution** par modalités ou par valeurs (éventuellement, une distribution par **classes de modalités** ou par **classes de valeurs**).

ii) Classer les **valeurs** ou les **modalités** par ordre décroissant des effectifs

iii) Ajouter une colonne pour la distribution en pourcentages

iv) Ajouter ensuite une colonne de pourcentages cumulés

v) Faire un graphique pour représenter simultanément :

- La distribution des pourcentages par un **graphique en barre**, en mettant l'axe des y à gauche

- Les pourcentages cumulés par une **courbe des pourcentages cumulés**, en mettant l'axe des y à droite

Exemple : Supposons que l'on veuille étudier les raisons de la résiliation d'un abonnement en ligne. Le problème ici est de comprendre pourquoi les abonnés résilient leur abonnement (afin de réduire le nombre de résiliations). On recherche donc les causes. Pour ce faire, lorsque les clients résilient leur abonnement, on leur propose un questionnaire (volontairement simplifié dans cet exemple) où ils sont invités à cocher la case qui correspond à la raison de la résiliation de leur abonnement. Supposons que les 5 choix suivants leurs soient proposés (Remarque : nous sommes en présence de **modalités nominales**, sans possibilité d'effectuer sur ces modalités des opérations numériques ayant un sens, ni d'ailleurs la possibilité de classer intrinsèquement des modalités [elles vont être classées par la suite en ordre décroissant du nombre de réponses qu'elle récoltent mais c'est un critère extrinsèque]) : le contenu du site ne correspondait pas à mes attentes (réponse codifiée par "A"), le contenu n'est pas bon (réponse codifiée par "B"), le

[14] Une vidéo imagée sur le regroupement des problèmes de qualité, puis le classement des groupes par ordre d'importance décroissant est disponible sur Youtube (en anglais) : http://www.youtube.com/watch?v=GVGdtlnZ7xM

contenu n'est pas renouvelé assez souvent (réponse codifiée par "C"), difficultés techniques pour accéder au contenu (réponse codifiée par "D"), Autres (réponse codifiée par "E").

On a obtenu les résultats suivants en étudiant 15 cas de résiliations : {**E**, **B**,**D**, **E**, **D**, **E**, **A**, **B**, **B**,**C**, **D**, **A**, **B**,**B**, **E**}. Evidemment, en réalité, on étudierait un nombre de cas beaucoup plus grand, mais le principe de construction resterait identique.

Construction du graphique :

i) Si les données sont sous forme d'une série, les regrouper par **modalités** (données qualitatives) ou par **valeurs** (données quantitatives), de façon à obtenir une **distribution** :

Modalités	Nombre de réponses
Le contenu ne correspond pas à mes attentes (A)	2
Le contenu n'est pas bon (B)	5
Le contenu n'est pas renouvelé assez souvent (C)	1
Difficultés techniques pour accéder au contenu (D)	3
Autres (E)	4
Total	15

ii) Classer les **valeurs** ou les **modalités** par ordre décroissant des effectifs :

Modalités	Nombre de réponses
Le contenu n'est pas bon	5
Autres	4
Difficultés techniques pour accéder au contenu	3
Le contenu ne correspond pas à mes attentes	2
Le contenu n'est pas renouvelé assez souvent	1
Total	15

iii) Ajouter une colonne pour la distribution en pourcentages

Modalités	Nombre de réponses	Pourcentages
Le contenu n'est pas bon	5	33,3
Autres	4	26,7
Difficultés techniques pour accéder au contenu	3	20,0
Le contenu ne correspond pas à mes attentes	2	13,3
Le contenu n'est pas renouvelé assez souvent	1	6,7
Total	15	100,0

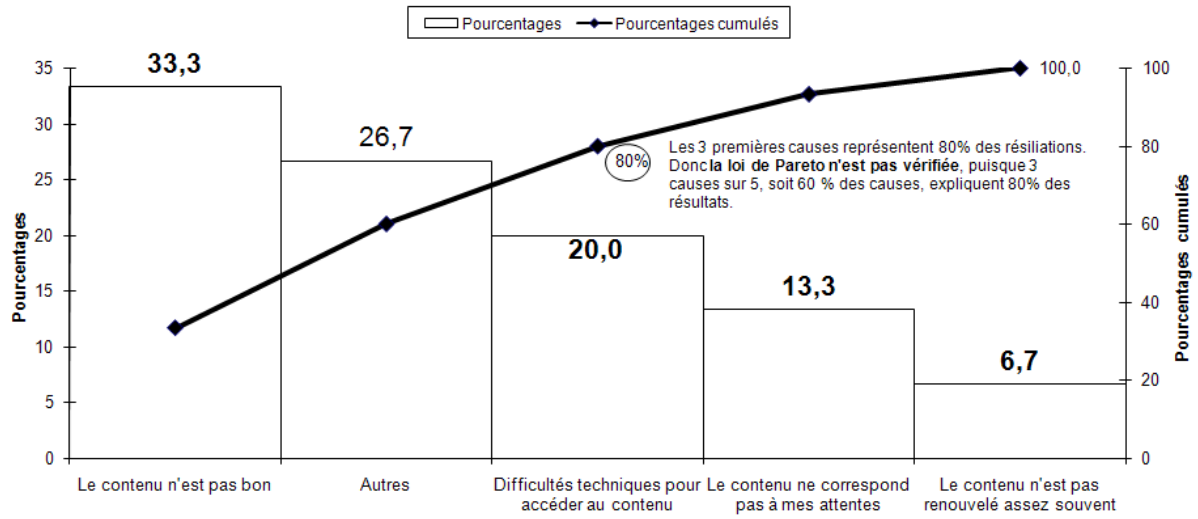
iv) Ajouter ensuite une colonne de pourcentages cumulés

Modalités	Nombre de réponses	Pourcentages	Pourcentages cumulés
Le contenu n'est pas bon	5	33,3	33,3
Autres	4	26,7	60,0
Difficultés techniques pour accéder au contenu	3	20,0	80,0
Le contenu ne correspond pas à mes attentes	2	13,3	93,3
Le contenu n'est pas renouvelé assez souvent	1	6,7	100,0
Total	15	100,0	

v) Faire un graphique pour représenter simultanément :

- La distribution des pourcentages par un **graphique en barre**, en mettant l'axe des y à gauche

- Les pourcentages cumulés par une **courbe des pourcentages cumulés**, en mettant l'axe des y à droite



Fichier EXCEL

c) Interprétation

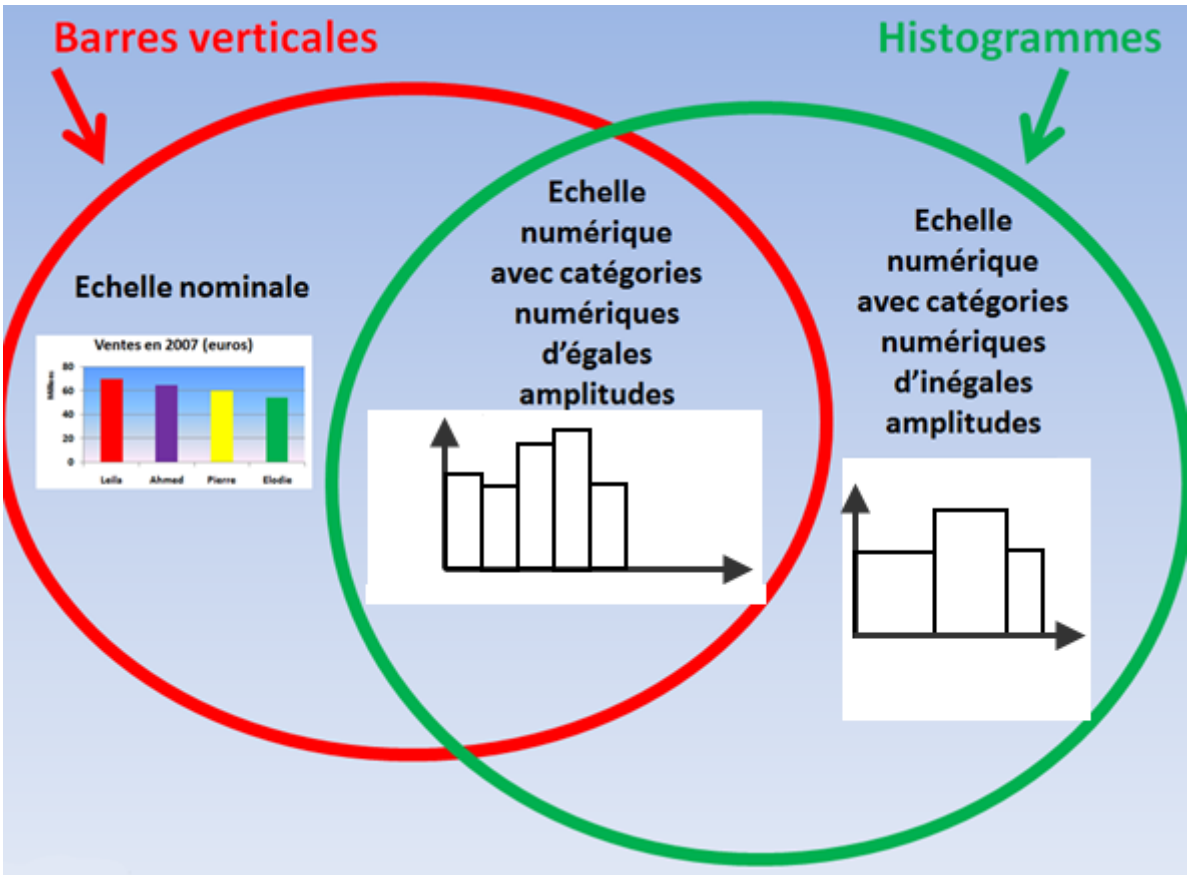
Dans notre exemple, on voit que la loi de PARETO n'est pas vérifiée. En effet, la loi de PARETO veut que 20% des causes expliquent 80 % des résultats. Or ici, il y a 5 causes. Donc une cause représente à elle seule 20% des causes. Pour que la loi de PARETO soit vérifiée, il faudrait qu'une seule cause (20% des causes) explique 80% des résultats (80% des résiliations). Or, ici, la première cause n'explique que 35% des résiliations et il faut 3 causes, soit 60% des causes, pour parvenir à expliquer 80% des résultats.

F – Histogramme

L'histogramme peut parfois être assimilé (au niveau de sa construction par un tableur) à un graphique en barres verticales. C'est le cas lorsque le graphique en barres verticales a pour échelle horizontale des catégories numériques d'amplitudes égales[15]. Mais l'histogramme se distingue du graphique en barres verticales lorsque les catégories numériques de l'axe horizontal sont d'amplitudes inégales. Le diagramme de VENN ci-après permet d'illustrer ce point et les exemples qui suivent permettent de le préciser.

À noter que le tracé informatique d'un histogramme avec classes d'amplitudes inégales, quoique possible, reste un tour de force[16].

Barres verticales et histogrammes : pas toujours la même chose !

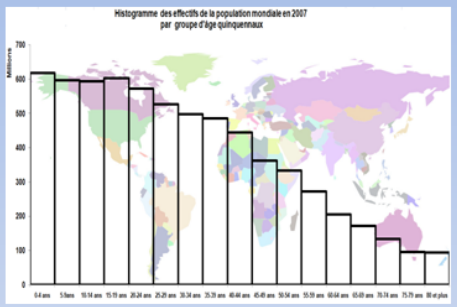
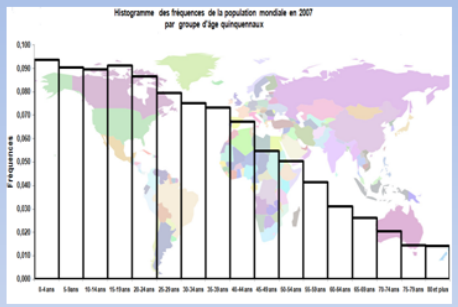
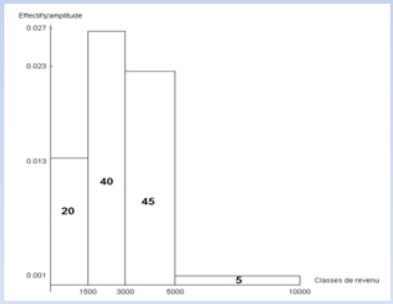
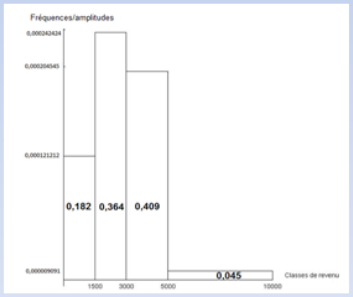


[15] Même dans ce cas, les barres de l'histogramme sont collées les unes aux autres, alors que les barres d'un diagramme en barres classique ne le sont généralement pas (mais on peut faire en sorte qu'elles le soient, d'où la convergence).

[16] C'était possible et assez simple à faire avec Mathematica 6. C'est devenu impossible dans Mathematica 7, la fonction correspondante n'assurant plus la proportionnalité des aires aux effectifs (ou aux fréquences).

S'agissant des histogrammes, il convient en outre de distinguer les histogrammes d'effectifs et les histogrammes de fréquences. Nous allons donc être amenés à étudier 4 types d'histogrammes comme indiqué dans le tableau ci-dessous.

Les 4 types d'histogrammes

	Histogramme d'effectifs	Histogramme de fréquences
Amplitudes de classes égales	1 	2 
Amplitudes de classes inégales	3 	4 

Nous allons construire chacun de ces 4 types d'histogrammes ci-après en les identifiant par leur numéro dans le tableau (de 1 à 4).

1) Amplitude de classes identiques

Soir le tableau ci-dessous qui donne la population mondiale en 2008 (estimation dite « en milieu d'année ») par groupes d'âges quinquennaux (hommes et femmes confondus). La dernière colonne, intitulées « fréquences » est simplement calculée en divisant l'effectif de chaque classe d'âge par la population mondiale totale. Par exemple, pour obtenir le premier chiffre de la colonne des fréquences (classe d'âge des 0-4 ans), on a effectué le calcul suivant :

$$\frac{631830670}{6706992932} = 0,094204761$$

Soit, en arrondissant : 0,094.

On remarque que toutes les classes d'âges sont identiques (5 ans)[17]. Les classes ont la même amplitude. A chaque classe d'âge est associé un effectif (colonne des effectifs) ou une fréquence (colonne des fréquences). La somme des effectifs donne la population mondiale en 2008, tandis que la somme des fréquences est égale à 1.

Nous allons d'abord voir comment se présente l'histogramme des effectifs, puis ensuite l'histogramme des fréquences.

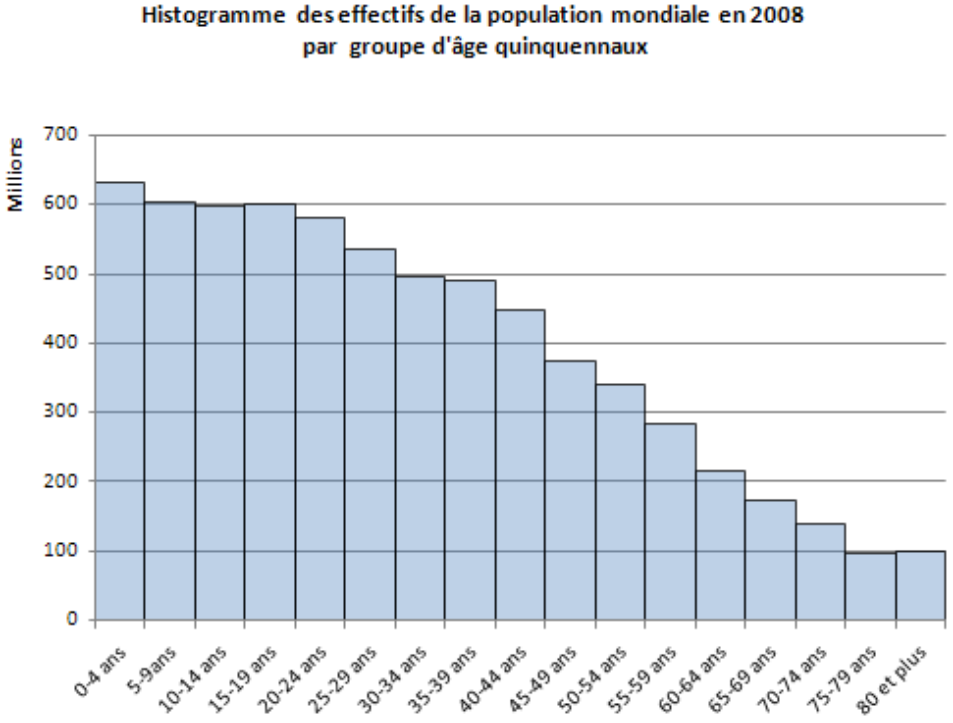
Population mondiale en 2008
par groupe d'âge quinquennaux
(Source : http://www.census.gov/ipc/www/idb/tables.html

Groupes d'âge	Effectifs	Fréquences
0-4 ans	631830670	0,094
5-9ans	603252033	0,090
10-14 ans	597053767	0,089
15-19 ans	601271944	0,090
20-24 ans	582319765	0,087
25-29 ans	536412822	0,080
30-34 ans	497152332	0,074
35-39 ans	490257604	0,073
40-44 ans	446771081	0,067
45-49 ans	375195788	0,056
50-54 ans	341483006	0,051
55-59 ans	282446909	0,042
60-64 ans	214089542	0,032
65-69 ans	174129418	0,026
70-74 ans	137647461	0,021
75-79 ans	97620891	0,015
80 et plus	98057899	0,015
Total	6 706 992 932	1

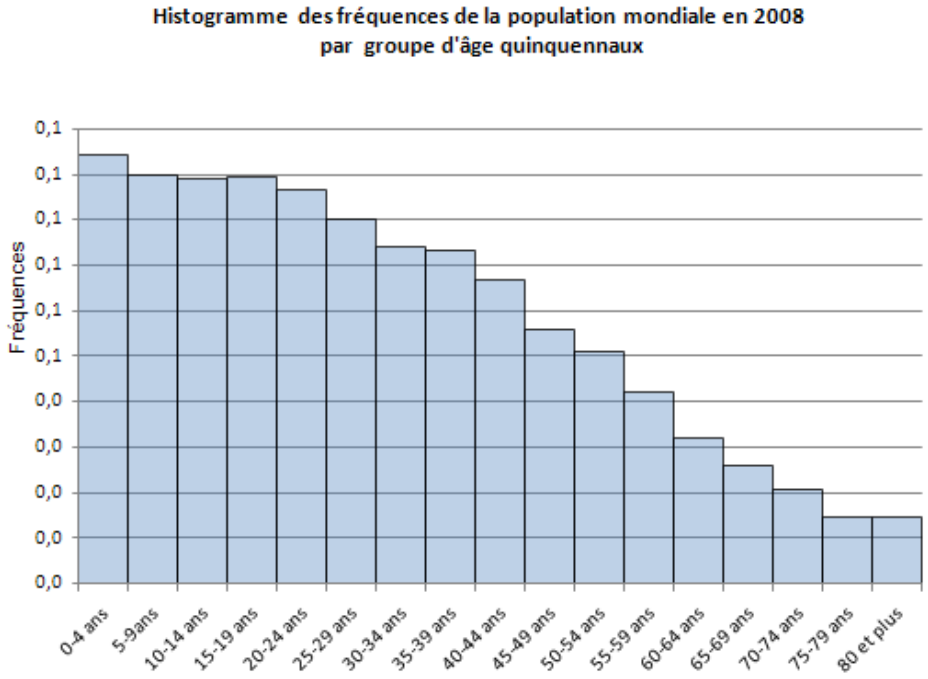
Voir le fichier EXCEL

[17] La dernière classe va en fait de 80 à plus de 110 ans, mais pour simplifier nous la supposons égale à 5 ans, en nous basant sur le fait que le nombre des 85 ans et plus reste encore minime comparé à l'ensemble de la population mondiale, même s'il est appelé à augmenter.

Exemple d'histogramme d'effectifs quand les catégories numériques sont d'amplitudes égales. Correspond à l'histogramme numéroté 1 dans le tableau « Les 4 types d'histogrammes »



Exemple d'histogramme de fréquences quand les catégories numériques sont d'amplitudes égales Correspond à l'histogramme numéroté 2 dans le tableau « Les 4 types d'histogrammes »



a) Histogramme des effectifs

Sur l'histogramme des effectifs de la figure précédente, on peut voir que :

- L'axe horizontal a pour échelle des catégories numériques d'amplitude égales (5 ans)
- L'axe vertical correspond à l'effectif de population associé à la classe d'âge correspondante

b) Histogramme des fréquences

Sur l'histogramme des fréquences ci-avant, on peut voir que :

- L'axe horizontal a pour échelle des catégories numériques d'amplitude égales (5 ans)
- L'axe vertical correspond à la fréquence associée à la classe d'âge correspondante.

Les deux histogrammes (celui des effectifs et celui des fréquences) ont la même forme, mais diffèrent par l'échelle de l'axe vertical.

2) Amplitude de classes différentes

Pour bien comprendre ce qu'est un histogramme (d'effectifs ou de fréquences) lorsque que les catégories numériques de l'axe horizontal sont d'amplitudes différentes, prenons l'exemple d'un échantillon de 110 ménages dont le revenu mensuel en euros est donné par la série classée ci-après :

Revenu = {1100, 1130, 1150, 1200, 1220, 1300, 1300, 1310, 1350, 1400, 1400, 1400, 1400, 1450, 1460, 1480, 1490, 1490, 1495,1495, 1500, 1500, 1550, 1600, 1600, 1630, 1640, 1700, 1900, 2000, 2020, 2050, 2070, 2090, 2100, 2200, 2220, 2400, 2500, 2540, 2560, 2600, 2710, 2730, 2750, 2800, 2810, 2810, 2820, 2840, 2850, 2850, 2850, 2870, 2890, 2900, 2920, 2960, 2980, 2990, 3000, 3000, 3000, 3000, 3000, 3030, 3050, 3070, 3080, 3090, 3090, 3090, 3095, 3100, 3200, 3210, 3250, 3280, 3300, 3350, 3400, 3400, 3400, 3400, 3420, 3450, 3500, 3550, 3560, 3570,3575, 3600, 3610, 3800, 4000, 4100, 4250, 4300, 4310, 4380, 4500, 4560, 4580, 4590, 4590, 5000, 6000, 7500, 9000, 9800}.

Supposons que l'on souhaite répartir ces ménages dans les catégories de revenu suivantes : [0 – 1500[; [1500 – 3000[;[3000 – 5000[; [5000 – 10000[. On va alors obtenir le tableau d'effectifs suivant :

Répartition des 110 ménages par classe de revenu

Classes de revenu	Effectifs
[0 -1500[	20
[1500-3000[	40
[3000-5000[	45
[5000-10000[	5
	110

L'amplitude de chaque classe n'est pas la même. Les 2 premières classes ont une amplitude de 1500 euros, la troisième a une amplitude de 2000 euros et la dernière classe a une amplitude de 5000 euros. Par conséquent, si l'on veut représenter ces données sous forme d'un histogramme d'effectifs, nous ne pouvons pas procéder de la même façon que lorsque les amplitudes de classes sont identiques. En effet, sur l'axe horizontal, les bases des « barres » n'auront plus nécessairement la même **longueur**. L'échelle de l'axe horizontal est la suivante :



On voit que dans ces conditions, la **hauteur** des barres verticales ne peut plus être proportionnelle aux effectifs, car cela aboutirait à donner une image fausse de l'importance des effectifs inclus dans chaque classe.

a) Histogramme d'effectifs

Pour tracer l'histogramme des effectifs, il faut donc modifier l'échelle de l'axe vertical en divisant les effectifs de chaque classe par l'amplitude de classe correspondante. On ajoute pour cela deux colonnes au tableau précédent :

Calcul des effectifs corrigés (effectifs sur amplitudes)

Classes de revenu	Amplitudes de classe	Effectifs	Effectifs /amplitudes
[0 -1500[	1 500	20	0,013
[1500-3000[	1 500	40	0,027
[3000-5000[	2 000	45	0,023
[5000-10000[	5 000	5	0,001
		110	

La colonne « amplitude de classe » donne l'écart en euros entre les deux extrémités de chaque classe. La colonne « effectifs corrigés » se calcule en divisant chaque effectif par l'amplitude de classe qui lui correspond ; Ainsi, l'effectif corrigé de la classe de revenu [0 – 1500[s'obtient par l'opération suivante :

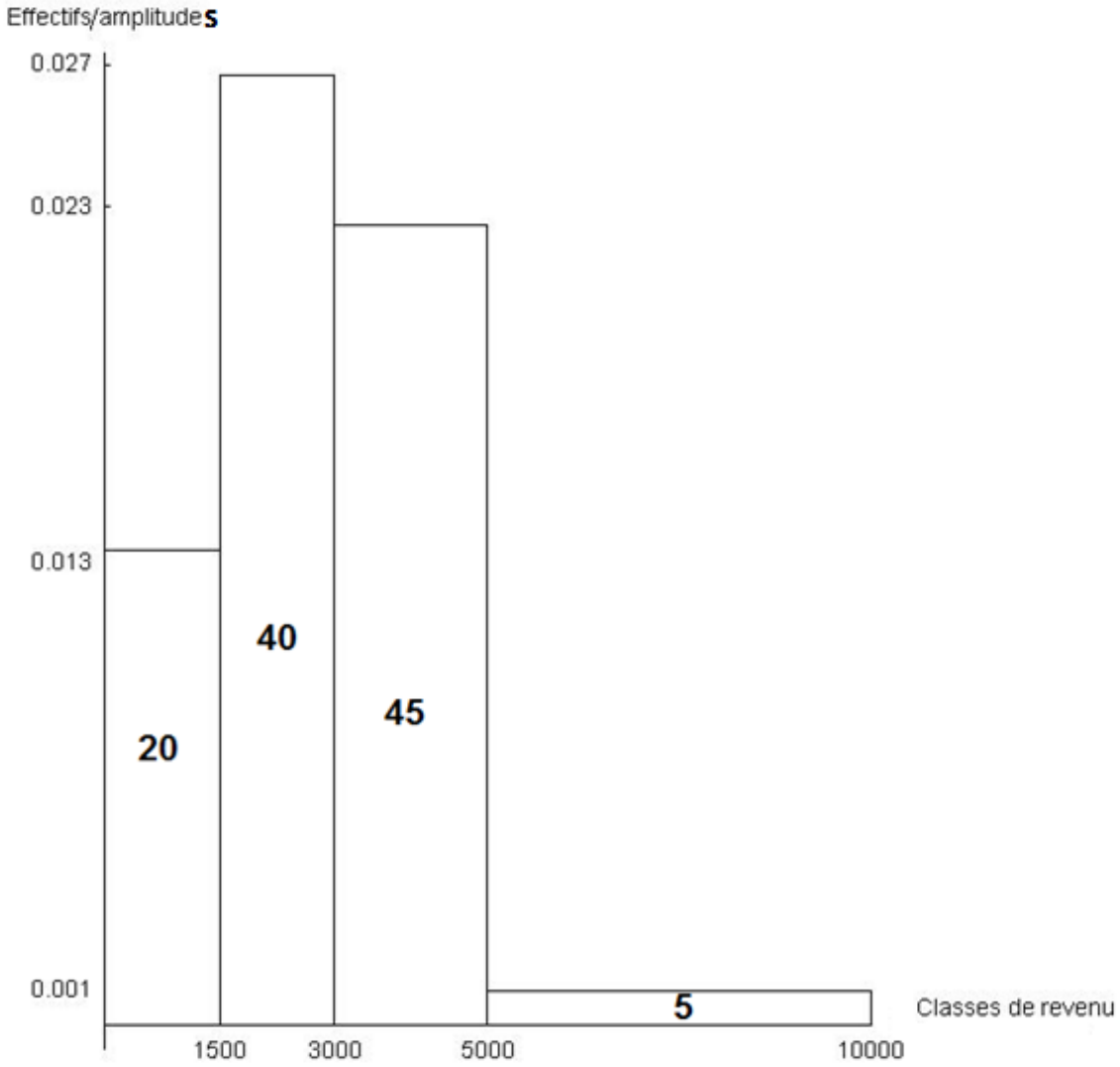
$$\frac{20}{1500} = 0,01333333$$

C'est-à-dire 0,013 en ne conservant que les 3 premières décimales. Les autres chiffres de la colonne s'obtiennent de la même façon.

Nous pouvons maintenant tracer l'histogramme des effectifs (voir graphique ci-après). Dans cet histogramme, ce n'est plus la **hauteur** de chaque barre qui indique l'effectif, mais sa **surface**. C'est la raison pour laquelle *l'effectif est reporté directement sur chaque barre*, tandis que l'axe vertical mesure **l'effectif corrigé**, ou plus précisément **l'effectif divisé par l'amplitude de classe**. Sur cet histogramme,

ce n'est plus la hauteur qui correspond à l'effectif, mais la surface. On peut voir facilement que la barre qui correspond à 40 a une surface double de celle qui correspond à 20. Et, bien que cela ne soit pas évident visuellement, la barre qui correspond à 45 a une surface qui est 45/40=1,125 plus grande que celle qui correspond à 40 et une surface 40/5 = 8 fois plus grande que celle qui correspond à 5.

Histogramme d'effectifs avec catégories numériques d'amplitudes différentes Correspond à l'histogramme numéroté 3 dans le tableau « Les 4 types d'histogrammes »



b) Histogramme de fréquences

Pour tracer l'histogramme des fréquences, il faut donc modifier l'échelle de l'axe vertical en divisant les fréquences de chaque classe par l'amplitude de classe correspondante.

On construit pour cela le tableau suivant :

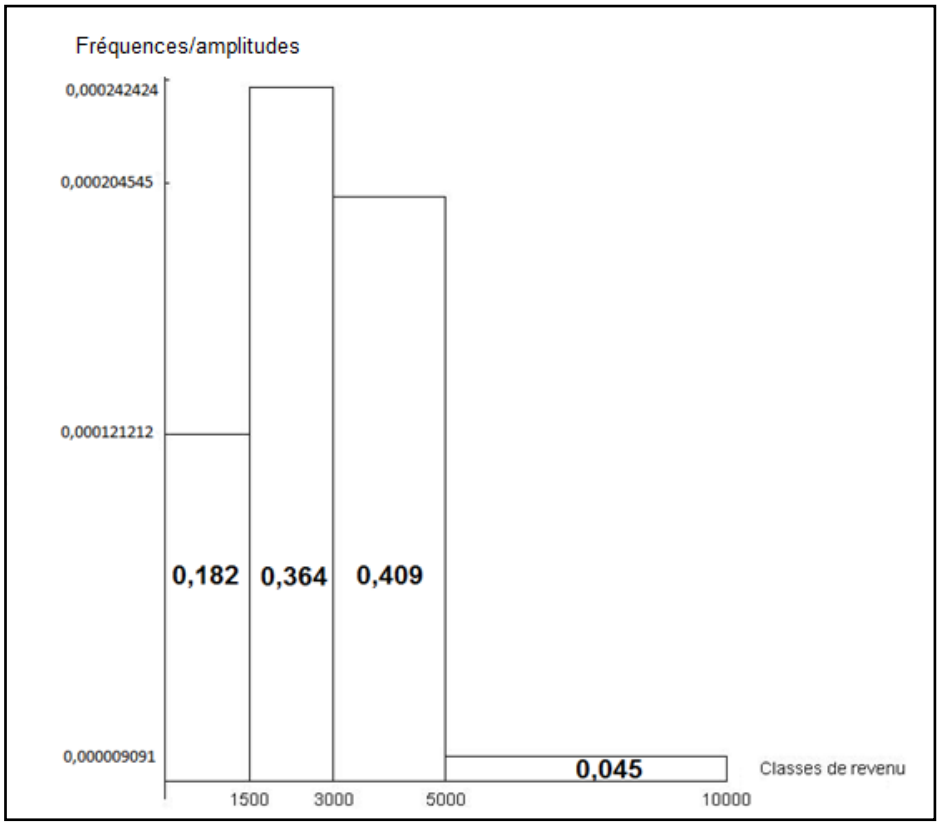
Calcul des fréquences corrigées (fréquences sur amplitudes)

Classes de revenu	Amplitudes de classe	Effectifs	Fréquences	Fréquences/amplitudes
[0 -1500[	1 500	20	0,182	0,000121212
[1500-3000[	1 500	40	0,364	0,000242424
[3000-5000[	2 000	45	0,409	0,000204545
[5000-10000[	5 000	5	0,045	0,000009091
		110	1,000	

Dans ce tableau, la colonne des fréquences correspond aux effectifs de chaque classe divisés par l'effectif total et la colonne des « fréquences corrigées » est égale au rapport de chaque fréquence à l'amplitude de classe correspondante.

On obtient un histogramme de fréquence qui a exactement la même forme que l'histogramme des effectifs, mais l'échelle verticale est graduée différemment (c'est l'échelle des amplitudes corrigées). Et dans chaque barre figure maintenant la fréquence qu'elle représente.

Histogramme de fréquences avec catégories numériques d'amplitudes différentes Correspond à l'histogramme numéroté 4 dans le tableau « Les 4 types d'histogrammes »



G – Pyramide des âges

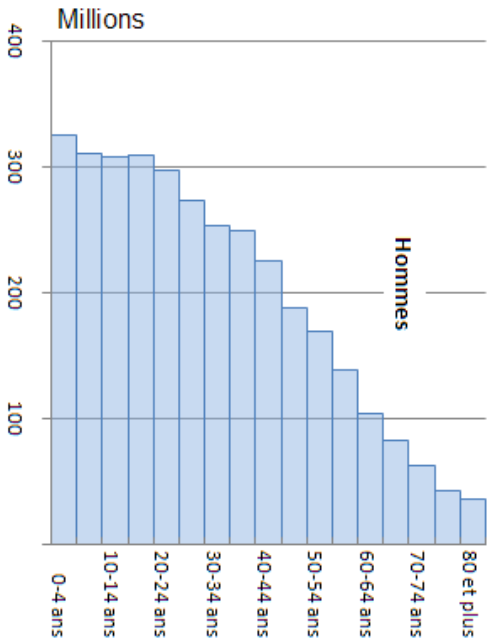
La pyramide des âges est un outil de l'analyse démographique inventé en 1870 par le Général WALKER, alors directeur du *Bureau of Census*, organisme américain chargé du recensement de la population et des études démographiques.

C'est en fait un **double histogramme** qui indique la composition d'une population par classes d'âges et par sexe, à un moment donné.

Faire une pyramide, c'est faire le graphique du nombre d'hommes et de femmes en tenant compte de leur date de naissance et de leur âge au moment où l'on arrête le calcul. Pour cela on place dos à dos, deux histogrammes décrivant la structure par classe d'âge d'une population masculine pour le premier et d'une population féminine pour le second. Le tableau ci-dessous répertorie la population mondiale par groupe d'âges quinquennaux et par âge en 2008. Nous allons l'utiliser à titre d'exemple pour construire la pyramide des âges.

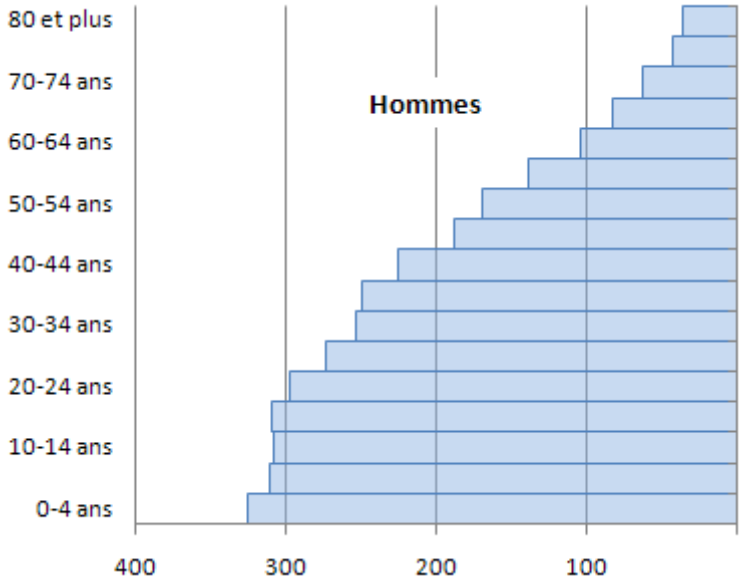
L'histogramme A ci-dessous représente la répartition de la population masculine mondiale en 2008 suivant les données du tableau précédent. Il s'agit d'un histogramme dont les amplitudes de classe sont identiques (5 ans) et qui peut donc être assimilé à un graphique en barres (voir le diagramme de VENN qui explique les différences et les similitudes entre barres verticales et histogramme).

Histogramme A



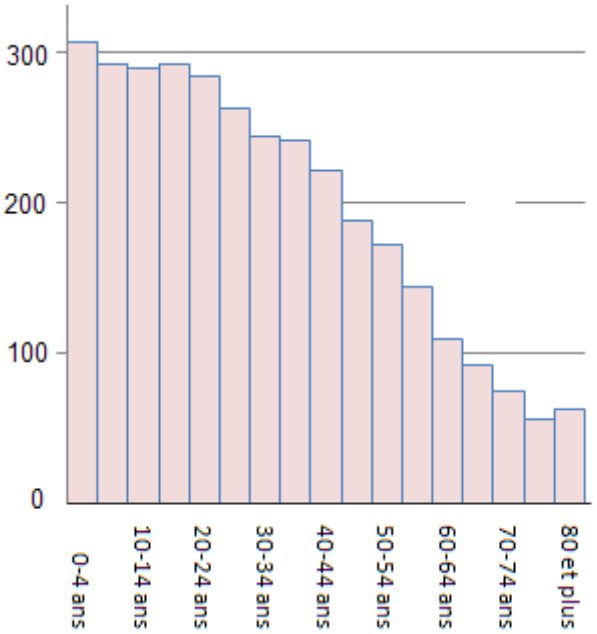
Nous pouvons ensuite faire pivoter ce graphique de 90° vers la gauche de façon à obtenir l'histogramme B.

Histogramme B



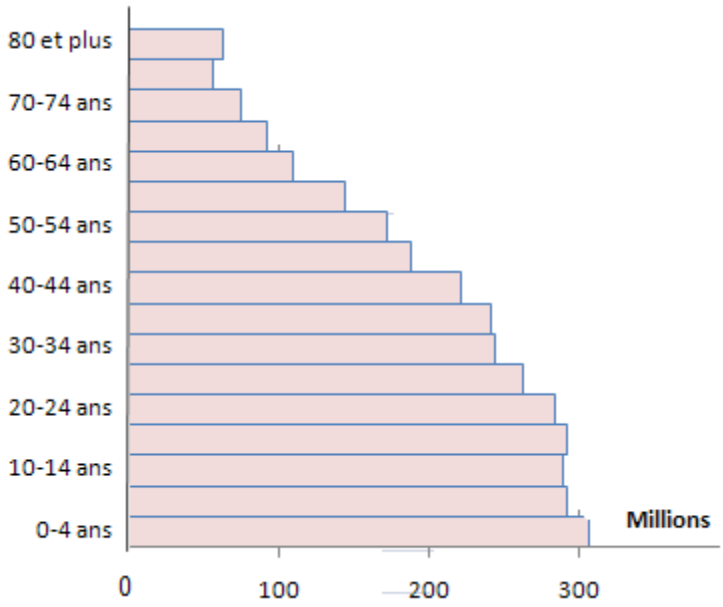
L'histogramme C ci-dessus représente la répartition de la population féminine mondiale en 2008

Histogramme C



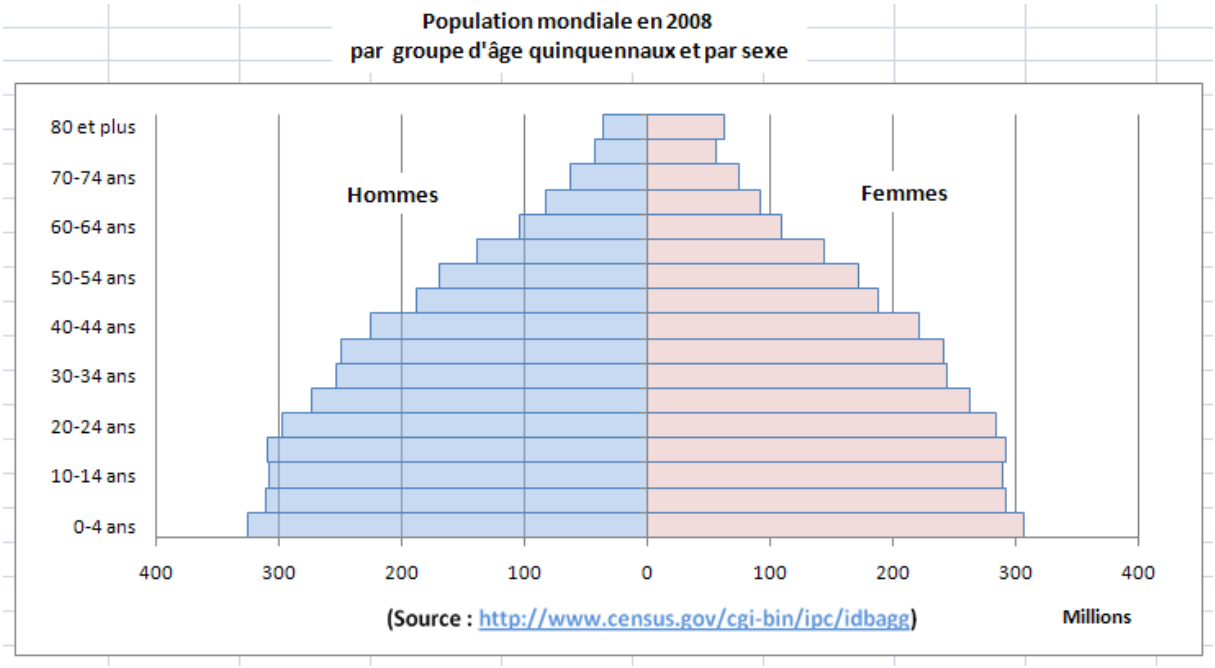
Nous pouvons ensuite faire effectuer à ce graphique une rotation de 90 degrés dans le sens inverse des aiguilles de la montre puis un pivotement de gauche à droite autour de l'axe vertical de façon à obtenir l'histogramme D.

Histogramme D



Fichier EXCEL

Et finalement, en mettant côte à côte les histogrammes B et D, nous obtenons la pyramide classique des âges de la population mondiale en 2008 :



Il s'agit en fait d'un graphique qui représente TROIS dimensions : l'âge, le sexe et les effectifs associés à ces deux catégories.

H – Graphique en cascade

Le graphique en cascade (« waterfall graphic ») est une variante du graphique en barres. Il sert à faire apparaître :

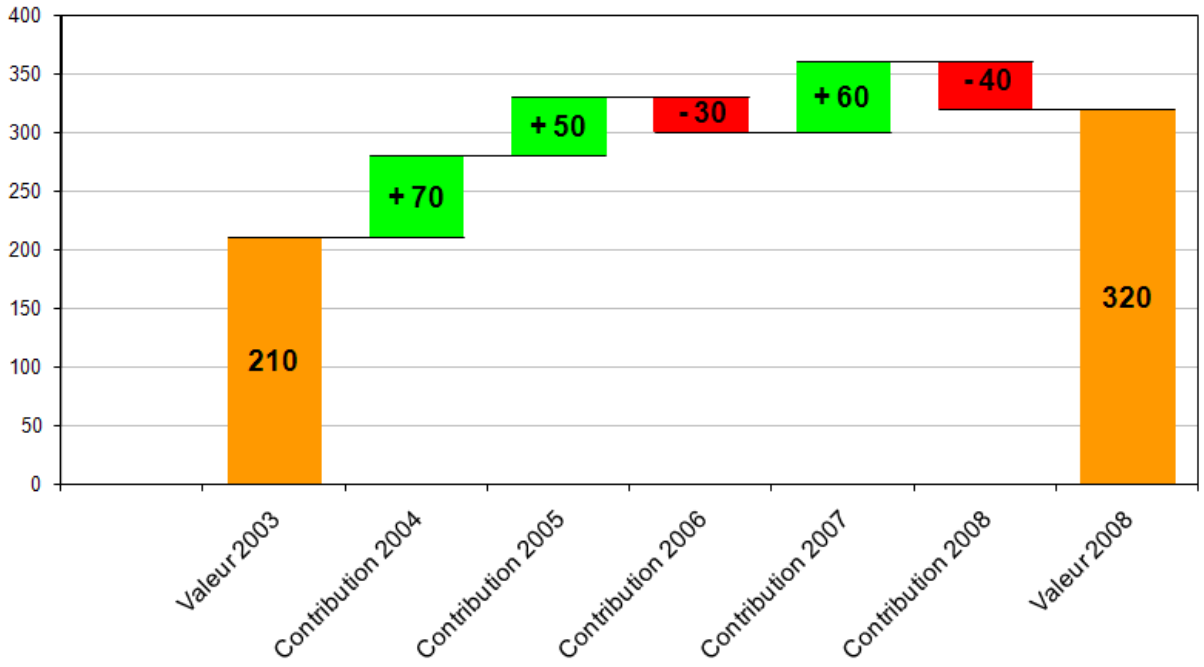
- les contributions positives et négatives à une grandeur

- les variations successives d'une grandeur.

Exemple : Le tableau ci-dessous indique l'évolution du chiffre d'affaire d'une entreprise de 2003 à 2008. La troisième colonne du tableau met en évidence les augmentations et la quatrième les diminutions. On voit ainsi que le CA a augmenté en 2004, 2005 et 2007, et qu'il a baissé en 2006 et 2008.

Années	CA (euros)	Plus	Moins
2003	210000		
2004	280000	70000	
2005	330000	50000	
2006	300000		-30000
2007	360000	60000	
2008	320000		-40000

Le graphique en cascade va permettre de faire apparaître ces variations (en milliers d'euros sur le graphique):



Voir le fichier Excel (il faut d'abord installer la macro : téléchargeable ici)

On voit ainsi immédiatement la contribution de chaque année et on peut visualiser :

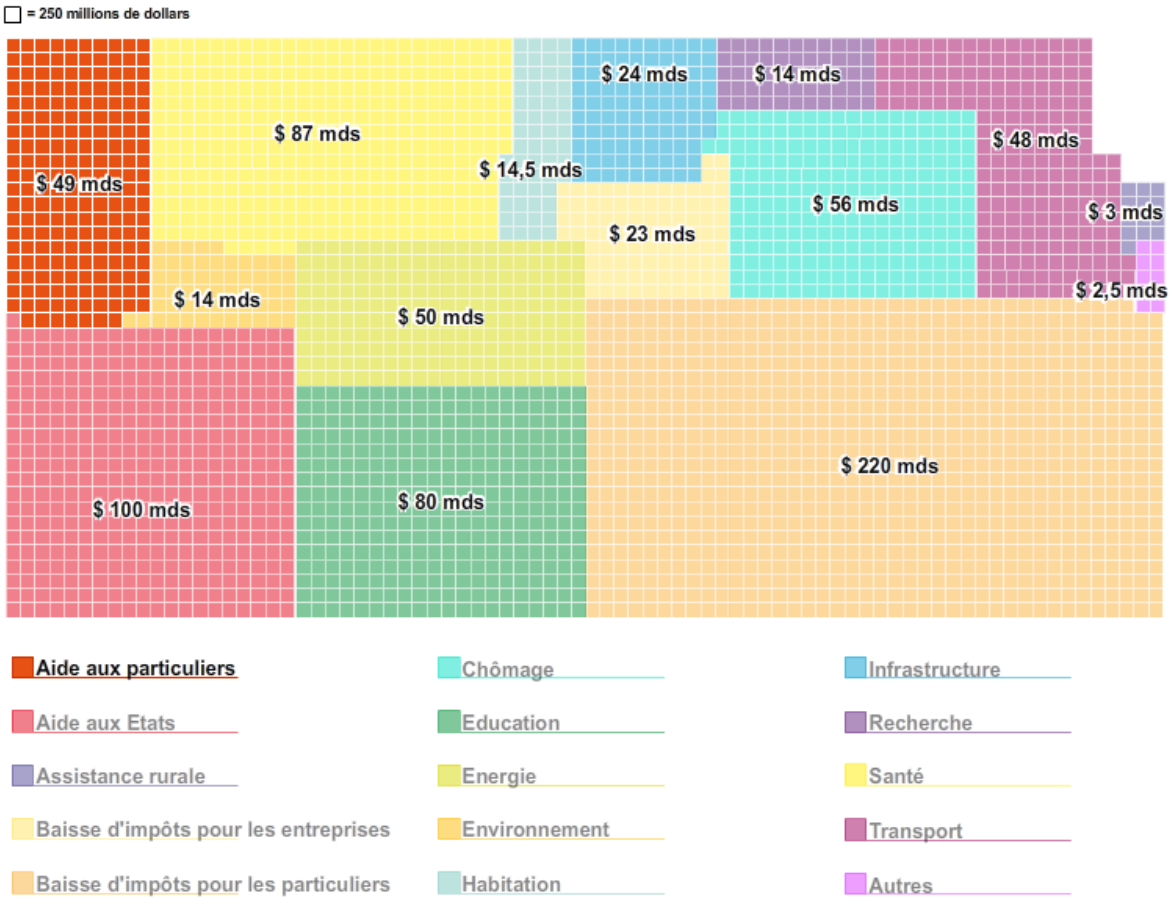
- l'importance de la contribution
- Si la contribution est positive ou négative.
- comment on est passé de 210 à 320 par variations successives.

I – Autres graphiques en forme d'aires

Nous avons déjà étudié plusieurs graphiques en forme d'aires :

1. le graphique en forme d'aires délimitées par des courbes.
2. Les histogrammes
3. Les secteurs (180° et 360°)

Mais les aires peuvent aussi être utilisées de manières plus générales pour visualiser des données. Un exemple est donné ci-après : celui de la répartition des fonds du plan GEITHNER de relance de l'économie américaine à la suite de la crise financière des années 2007-10. Une surface de type « pixel carré » est définie (ici un « pixel carré » vaut 250 millions de dollars). Ensuite, pour représente 49,250 milliards de dollars, on va utiliser 197 carrés (197 x 0,250 = 49,250 milliards) et ainsi de suite.



Source : Le Monde, http://medias.lemonde.fr/mmpub/edt/flash/20090216/1156269_5d37_relance.swf

Cette présentation est particulièrement adaptée aux graphiques interactifs, mais aussi à la vision de la répartition d'une masse en plusieurs composantes susceptibles de varier. **Son impact visuel est néanmoins discutable** (comparer par exemple les 14 milliards destinés à l'environnement et les 14,5 milliards destinés à l'habitat : sans les chiffres, comment savoir que ces deux surfaces sont approximativement identiques ?).

J – Graphique à trois dimensions

Grâce aux ordinateurs et aux logiciels il est devenu très facile aujourd'hui de réaliser de beaux graphiques en 3D. La troisième dimension reste cependant une construction visuelle dans la mesure où elle doit être affichée sur des écrans à 2 dimensions[18]. EXCEL 2007 offre diverses possibilités, tout comme d'autres logiciels, tels que Mathematica. En revanche, à moins d'être très bon dessinateur, il est impossible de réaliser ce type de graphique avec la règle, le rapporteur, le compas et les crayons de couleur (à l'inverse de tous les autres graphiques vus jusqu'à présent).

On peut distinguer 3 catégories de graphique en 3D :

- Les « faux » graphiques en 3D qui ne sont que des graphiques en 2D auxquels on a ajouté une profondeur à des fins visuelles.
- Les graphiques en barres à 3 dimensions
- Les graphiques dits « de surface » (non traité)

1) Graphiques en 2D avec ajout de profondeur

Pour illustrer ce type de graphique, reprenons les données du chiffre d'affaires hypothétique qu'une entreprise a réalisé en 2009 (249 327 045 euros) répartis par ses 4 vendeurs et dans les trois villes où se trouvent ses clients.

	Ventes 2009 (euros)			
	Leila	Ahmed	Pierre	Elodie
Lyon	18 753 951	8 754 668	9 785 246	16 487 564
Marseille	13 225 478	20 154 287	17 892 555	15 897 233
Paris	37 895 214	35 877 421	32 558 741	22 044 687
	69 874 643	64 786 376	60 236 542	54 429 484

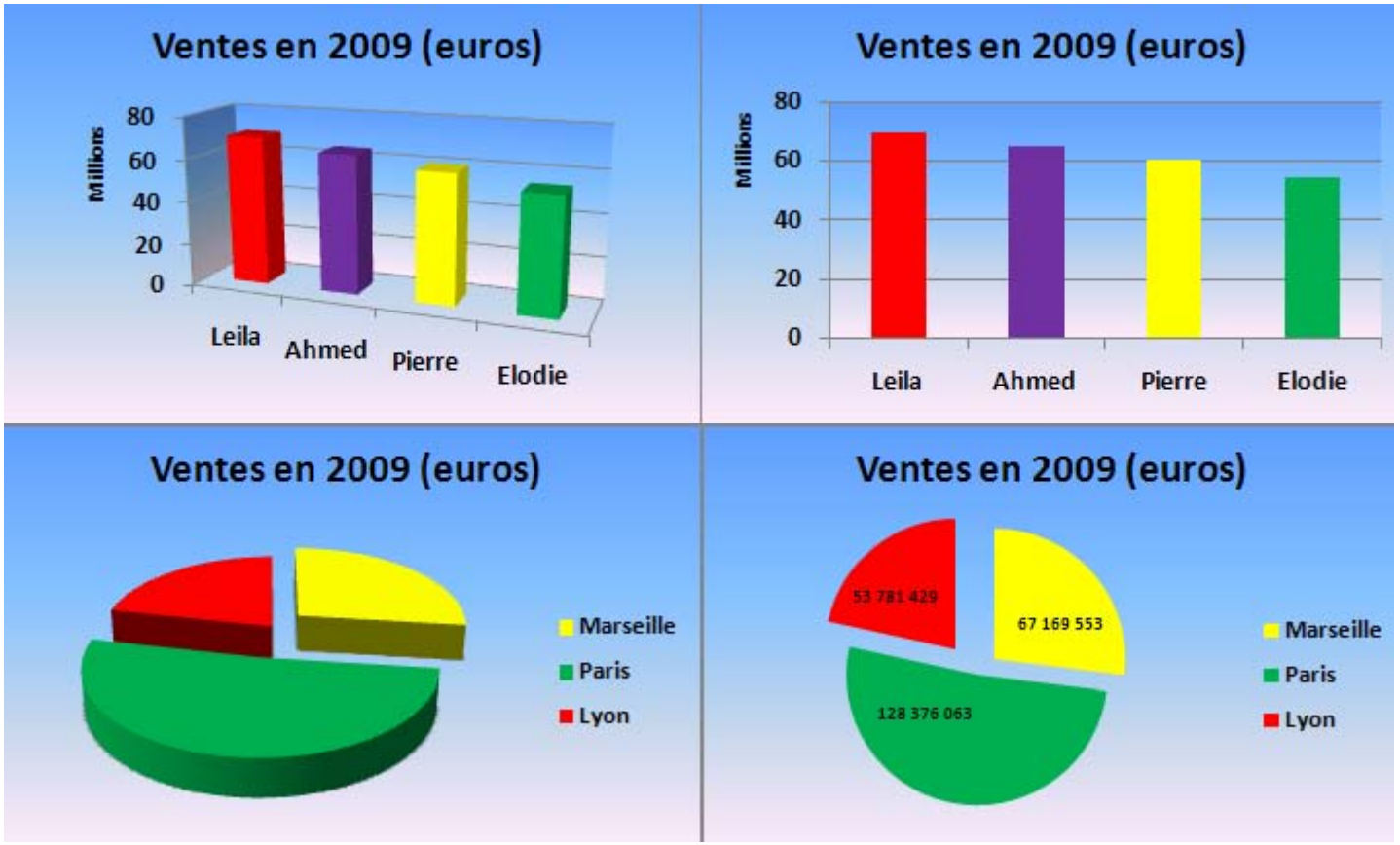
Lorsque l'on ajoute de la « profondeur » ou de la « perspective » au graphique en barres verticales ou aux secteurs classiques, on obtient ce genre de résultats (voir graphiques ci-après).

Bien entendu, le nombre de dimensions n'a pas changé par rapport aux équivalents 2D de ces deux graphiques qui n'ont que l'inconvénient de paraître « plats » par comparaison.

Les graphiques 2D avec ajout de profondeur ou de perspective attirent davantage les regards. Il ne faut donc pas se gêner pour les utiliser surtout étant donné la facilité avec laquelle on peut les réaliser grâce aux logiciels.

[18] Ceci est appelé à changer dans les décennies à venir, avec l'essor du virtuel.

À gauche, graphiques en 2D avec ajout de « profondeur » ; à droite : leurs équivalents 2D

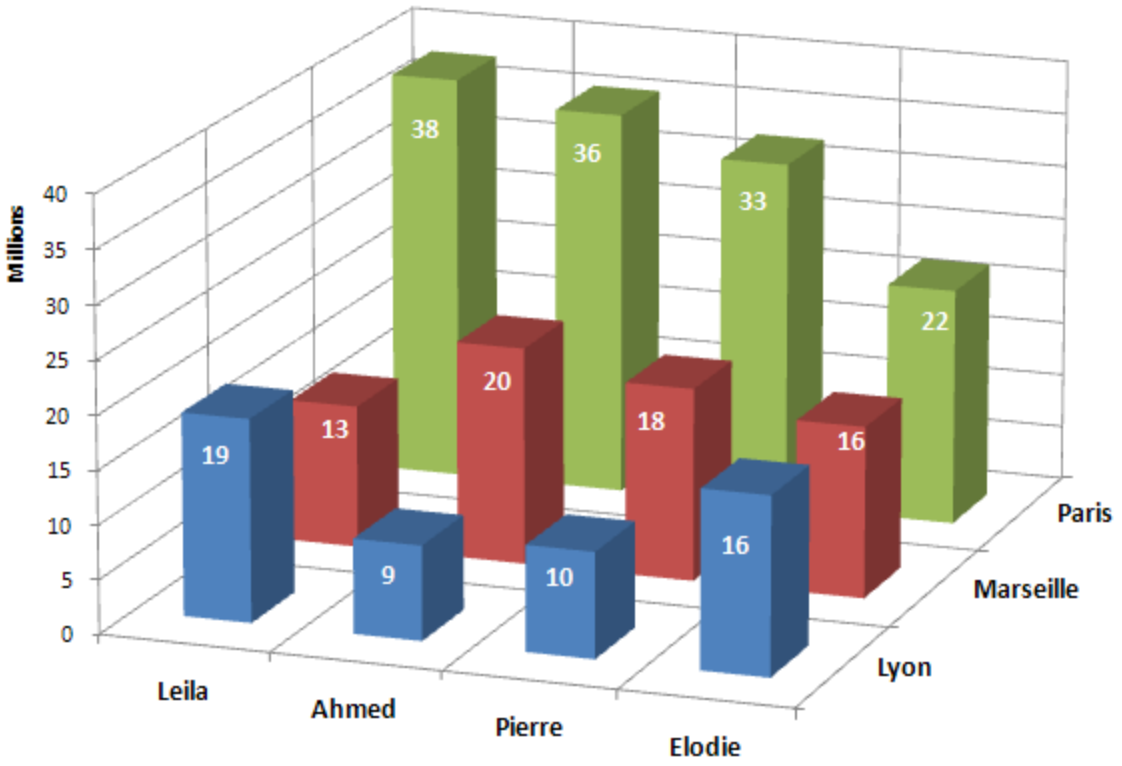


Fichier EXCEL

2) Graphique en barres avec 3 dimensions "réelles"

Par comparaison avec les graphiques en 2D avec ajout de profondeur, les graphiques en barres avec 3 dimensions ajoutent réellement une dimension supplémentaire.

Cette fois, nous allons utiliser un seul graphique en barres verticales pour montrer à la fois la répartition du CA par villes et par vendeur en 2009.



	Ventes 2009 (euros)			
	Leila	Ahmed	Pierre	Elodie
Lyon	18 753 951	8 754 668	9 785 246	16 487 564
Marseille	13 225 478	20 154 287	17 892 555	15 897 233
Paris	37 895 214	35 877 421	32 558 741	22 044 687
	69 874 643	64 786 376	60 236 542	54 429 484

Fichier EXCEL

6 – Résumé

Les **diagrammes** et les **graphiques** sont des représentations visuelles simplifiées, mais néanmoins rigoureuses, de chiffres, de concepts, de catégories, d'images, etc. Bien que certains diagrammes représentent aussi des chiffres, c'est principalement au graphique que cette fonction incombe. Plus précisément, en statistique descriptive, le graphique sert à représenter visuellement des informations sur les dimensions (caractères ou variables) d'une population d'unités statistiques.

Les graphiques révèlent des informations sur la forme des séries que les tableaux et les statistiques résumées ne peuvent pas toujours montrer.

Beaucoup de graphiques à deux dimensions se présentent sous la forme suivante dite d'un « système de coordonnées cartésiennes ». Il existe **différentes sortes d'échelles** propres à ce type de représentations graphiques.

Une **échelle numérique** est une échelle qui mesure des valeurs qui peuvent varier de moins l'infini à plus l'infini.

Une **échelle de catégories** est une échelle sur laquelle sont portées des catégories. Il peut s'agir :

- De catégories numériques (échelle quantitative)
- De catégories non numériques dites « nominales » (échelle qualitative)

Un **graphique à échelle verticale double** permet de représenter les valeurs ou les modalités d'une variable ou d'un caractère sur l'échelle verticale de droite et les modalités d'une variable ou d'un caractère sur l'échelle verticale de gauche.

L'échelle logarithmique est une échelle qui mesure le logarithme décimal des valeurs de la variable. C'est un moyen de mettre en évidence les propriétés d'une progression dans le temps et de comparer cette progression avec une autre progression. L'échelle logarithmique permet aussi de mieux visualiser les écarts entre valeurs, lorsque ces écarts varient dans de très fortes proportions.

Parmi les **diagrammes statistiques**, on distingue notamment le **pictogramme**, dessin généralement en couleur, conçu de préférence par un infographiste, qui essaie de traduire non seulement des données quantitatives, mais également des informations d'ordre commercial ou esthétique, le **cartogramme**, pour localiser géographiquement (« géo localiser » fait plus branché) des valeurs ou des modalités relatives à une variable ou à un caractère décrivant les unités statistiques d'une population et le **diagramme de GANTT**, qui est un outil de gestion de projets. Il sert à visualiser dans le temps les différentes étapes qui composent un projet.

Parmi les 73 représentations graphiques de base proposées par le logiciel EXCEL on distinguera plus particulièrement les graphiques usuels suivants : graphique en barres, horizontales ou verticales, simples, multiples ou tronçonnées, les courbes, les aires délimitées par des courbes, les nuages de points et les secteurs à 360 ou 180 degrés, concentriques ou non.

Mais la liste des graphiques ne peut pas par définition, être exhaustive, car c'est seulement l'imagination qui en limite le nombre. Ainsi a-t-on aussi étudié dans ce chapitre, outre les graphiques usuels, d'autres graphiques tels que les anneaux et les « radars » et autres graphiques en toile d'araignée, les graphiques à bulles, le graphique de TUKEY, les graphiques « panachés », le graphique de PARETO, l'histogramme (qu'il faut distinguer du graphique en barres verticales quand les amplitudes de classes sont inégales), la pyramide des âges, le graphique en cascade et les représentations tridimensionnelles.