Chapitre 3

Graphiques, diagrammes et autres messages visuels

Ce sont les **figures très colorées** sous forme de rectangles, de cercles ou de lignes brisées montantes ou descendantes, que l'on voit dans la presse, spécialisée ou non, dans les émissions de télévision, dans les rapports professionnels et même dans certaines publicités ou encore sur des boîtes de produits de consommation courante.

On ne peut pas y échapper : il y en a un peu partout, et il est vrai qu'on les « comprend » assez bien généralement **au premier coup d'œil.** Du moins on comprend **assez bien le** « **message** » qu'elles veulent faire passer : telle proportion de réussite au baccalauréat, telle part de cadres dans les entreprises du bâtiment, telle progression d'intentions de vote pour chaque candidat à l'élection présidentielle, telle croissance espérée de nos actions en bourse, etc.

En ce sens, **l'effort de communication** est indéniable. À la différence d'un tableau statistique (chapitre précédent), terne et monochrome en général, rempli de nombres plus ou moins longs, placés systématiquement les uns au-dessous des autres sans grande possibilité d'originalité, un peu « lourd », *un peu rébarbatif* pour tout dire, le diagramme ou le graphique[1] apparaît lumineux, éclatant, *attractif*.

Cependant, rien n'étant jamais parfait (même en statistique), il faut bien reconnaître que le graphique donne, de façon visuelle, plutôt des **ordres de grandeur** que des chiffres précis, contrairement au tableau. Le graphique intégrera, bien sûr, quelques chiffres exacts à la virgule près, placés dans une partie de cercle, ou sur le haut d'un bâtonnet, ou sur une échelle verticale, mais jamais on ne pourra donner tous les chiffres, au risque de le charger à tel point qu'il en devienne incompréhensible et dès lors « non communicant ».

Les deux instruments (*tableaux et graphiques*) sont donc complémentaires.

1. Les termes de *graphique* et de *diagramme* pourront, ici, **être substitués** sans trop de problèmes de sémantique. Cependant, on qualifiera plutôt de diagrammes les représentations visuelles générales et de graphiques celles qui sont plus particulièrement construites sur un repère normé en au moins deux axes d'abscisses et d'ordonnées.

Chacun conserve sa fonction, *faire passer un message* ; l'un est complet et un peu triste, l'autre plus direct et plus gai, mais tous les deux sont *parfaitement rigoureux* notamment dans leur construction.

Tout comme pour un tableau, le diagramme est évidemment dépendant de la *consigne de présentation* décrite en fin de chapitre précédent, respectant les principes : ***source-titre-intitulés-unités.***

Mais ici, il faut rajouter un principe : c'est la « **fameuse règle d'or** » des diagrammes, qui s'applique à toutes les figures statistiques que l'on peut voir, construire ou imaginer.

Elle est basée sur la **vitesse de compréhension** du lecteur (celui à qui s'adresse le message). Une fois que le lecteur a lu le titre du diagramme, repéré la source et déchiffré la légende, il ne faut pas qu'il dépasse **quelques secondes supplémentaires** pour comprendre le sens du message. Il peut rester béat d'admiration pendant l'heure qui suit s'il estime que le diagramme est magnifique, mais il aura **compris le message au tout début**, sinon ce n'est pas de la communication.

Règle d'or en matière de messages statistiques visuels

UN DIAGRAMME DOIT ÊTRE COMPRIS EN DIX SECONDES, UNE FOIS LE TITRE LU.

Précisons bien : si le lecteur dépasse la dizaine de secondes pour cette compréhension, cela ne veut surtout pas dire qu'il est idiot… bien au contraire, c'est que le graphe est mal fait (volontairement ou non). Cela arrive quelquefois quand le concepteur ne se met pas à la place de son lecteur. C'est donc dans ce cas **l'auteur du diagramme (et non l'utilisateur) qui est « limité » au point de ne pas avoir bien réalisé son devoir de « communicateur »** en la matière.

On ne construit pas un diagramme pour se faire plaisir ou pour rester la seule personne au monde à pouvoir le déchiffrer.

§ 1 – LES REPRÉSENTATIONS GRAPHIQUES À TRADUCTION DIRECTE

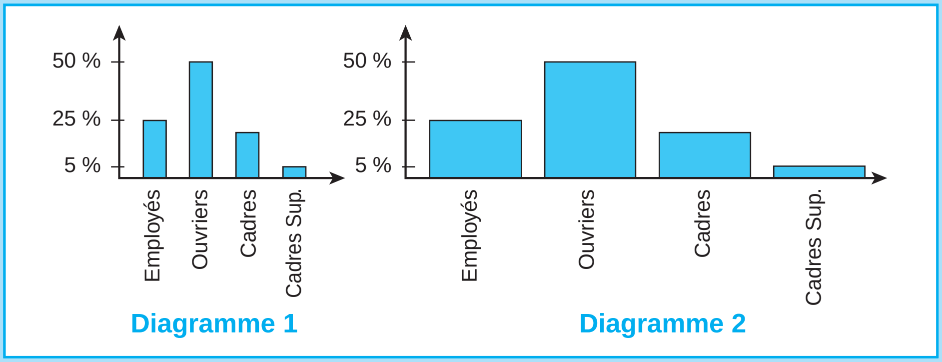
Ce sont celles, les plus classiques, qui n'utilisent **pas d'autres instruments de transformation que la géométrie traditionnelle pour leur construction** : on utilise ici des rectangles, des cercles, ou des courbes, mais pas des illustrations (pictogrammes ou figurines), ou de la cartographie, ou des échelles spéciales comme les échelles logarithmiques.

A – Les diagrammes de hauteur

L'effet visuel est ici rendu par l'appréciation de différentes *hauteurs* (et non pas, par exemple, de surfaces) des éléments qui composent le diagramme. Si un étroit rectangle vertical représente la proportion d'employés dans une entreprise, ce sera *sa hauteur* et non *sa base* qui exprimera la mesure, comme le montre l'exemple ci-après :

Diagramme en tuyaux d'orgues

Figure 3.1 Répartition du personnel salarié de l'entreprise X par type d'activité au 1er janvier 2007

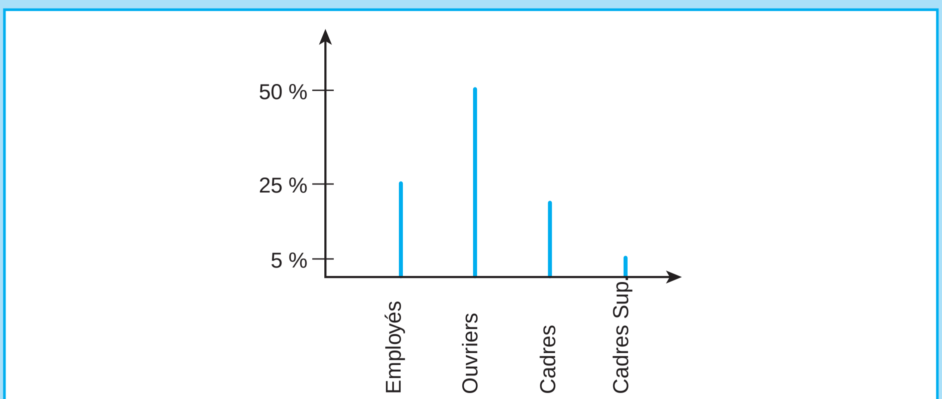


Plusieurs remarques peuvent être faites grâce à cet exemple :

- Ce que jusqu'à présent nous avons appelé « rectangles » va maintenant prendre sa terminologie réelle de « tuyaux d'orgues ». Il s'agit donc d'un **diagramme en tuyaux d'orgues**.
- On voit bien que la *hauteur* des tuyaux d'orgues est la même dans les deux cas de figure. Seule change la **dimension de la base**, et son choix ne dépend que du rendu esthétique que l'on cherche : quel diagramme préférons-nous ? Celui de droite ou celui de gauche ? Ou un qui soit entre les deux ? Tant que les hauteurs sont conservées **et les bases constantes**, nous pourrons faire comme il nous plaira.
- Dans cet exemple, on a affaire à un **diagramme à une dimension**, comme pour les tableaux. Il montre un seul caractère (la PCS) face à sa fréquence. On aurait pu le dessiner de la manière suivante :

Diagramme en bâtons

Figure 3.2 Répartition du personnel salarié de l'entreprise X par type d'activité au 1er janvier 2007

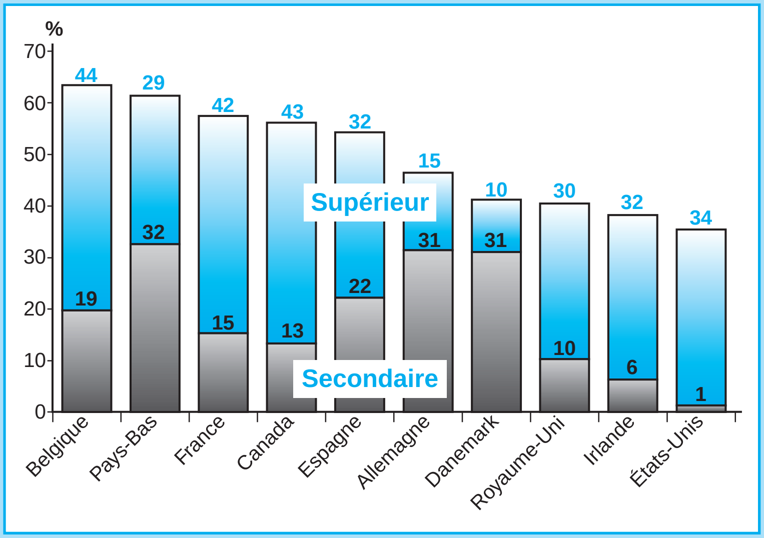


Ce diagramme-là, qui a exactement la même fonction que le précédent, est appelé « **diagramme en bâtons** ». On a simplement réduit les bases à un point[1] et les hauteurs sont les mêmes.

Le diagramme en bâtons peut être jugé moins esthétique, mais c'est surtout un problème de fonctionnalité qui nous fait souvent préférer l'autre : en effet, avec les tuyaux d'orgues, on peut insérer **deux dimensions** et, ce faisant, *on double la richesse de l'information* :

Diagramme en tuyaux d'orgues à 2 dimensions

Figure 3.3 Taux de scolarisation à 20 ans, selon les pays et les niveaux d'enseignement, en 1996



Source : « Les grands chiffres de l'Éducation nationale », Ministère de l'Éducation nationale, de la recherche et de la technologie, septembre 1999.

Ce diagramme en tuyaux d'orgues est bien **à deux dimensions** : la dimension quantitative « taux de scolarisation selon les pays » et la dimension qualitative « taux de scolarisation selon l'enseignement », *supérieur* ou *secondaire*.

Un autre moyen d'insérer deux dimensions est **de juxtaposer deux modalités**, soit de façon verticale, soit de façon horizontale, comme dans le diagramme ci-après.

Notons que ce diagramme s'appelle *diagramme en tuyaux d'orgues couchés*, ou plus généralement « **diagramme en barres** ». La logique de construction est la même que précédemment ; on a simplement « renversé » les axes du graphique.

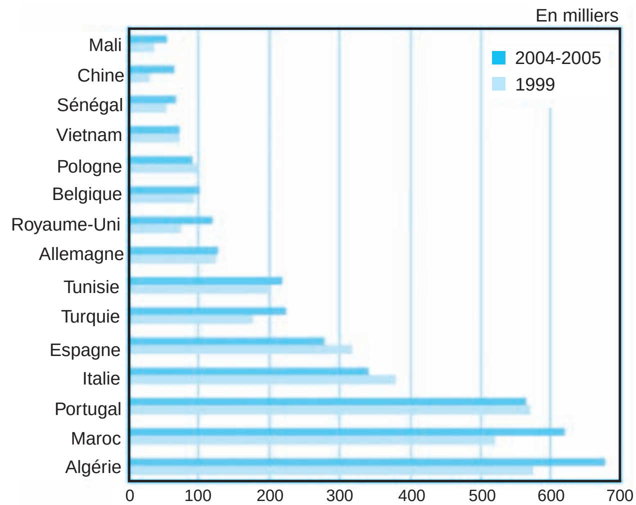
Pour bien comprendre la diversité des possibilités de choix, et pour se mettre en situation pratique, le lecteur peut se reporter aux deux exercices intitulés « **Choisir le message visuel adapté** », page 98 et page 103.

Il est fréquent de trouver de tels diagrammes (*barres* ou *tuyaux d'orgues* à deux dimensions). À plus de deux ou trois dimensions, le diagramme est difficilement lisible.

1. En anglais, on fait la différence entre *Rod Graph*, les bâtons, et *Column Chart*, les tuyaux d'orgues.

Diagramme en barres

Figure 3.4 Les immigrés selon le pays de naissance, en 1999 et en 2004-2005



Source : INSEE, enquêtes de recensement de 2004 et 2005.
Champ : France métropolitaine
Référence : *INSEE Première*, n° 1098, août 2006

Au-delà de l'effet visuel, il faut noter que les caractères de ces tableaux sont soit **qualitatifs** (comme ici, la nationalité ou la date), soit **quantitatifs discrets** (par exemple le nombre de pièces d'un appartement) ou **quantitatifs regroupés par blocs** (par exemple des normes de surface d'appartement).

En aucun cas, ces diagrammes ne s'adaptent au cas continu. Nous verrons que le cas des variables *quantitatives continues* est traité par un autre type de graphe : *l'histogramme*, dont l'effet visuel n'est pas la hauteur mais **la surface**. C'est d'ailleurs une erreur que commettent bien des gens et même parfois certains logiciels : confondre les deux types de diagrammes et tomber dans le piège de la mauvaise interprétation en cas de classes inégales (voir plus loin, page 77 et page 78).

B – Les diagrammes à axes multiples : toile d'araignée et radar

Ce sont encore des diagrammes de hauteur, mais leur ambition est de schématiser **des comparaisons** visuelles sur deux – mais souvent beaucoup plus – modalités du caractère étudié. Il y a **autant d'axes que de modalités**.

Les axes, **gradués depuis un même centre**, donnent une image d'*étoile* ou de *toile d'araignée*, selon leur nombre[1]. Certains logiciels utilisent le terme de « **radar** » ; en anglais, on parle de *spider chart* ou de *radar chart*.

1. Ces graphiques ressemblent aux ***graphiques en coordonnées polaires***, dans lesquels la lecture est identique, mais pas la construction, celle-ci étant angulaire (ou trigonométrique) dans ces derniers cas.

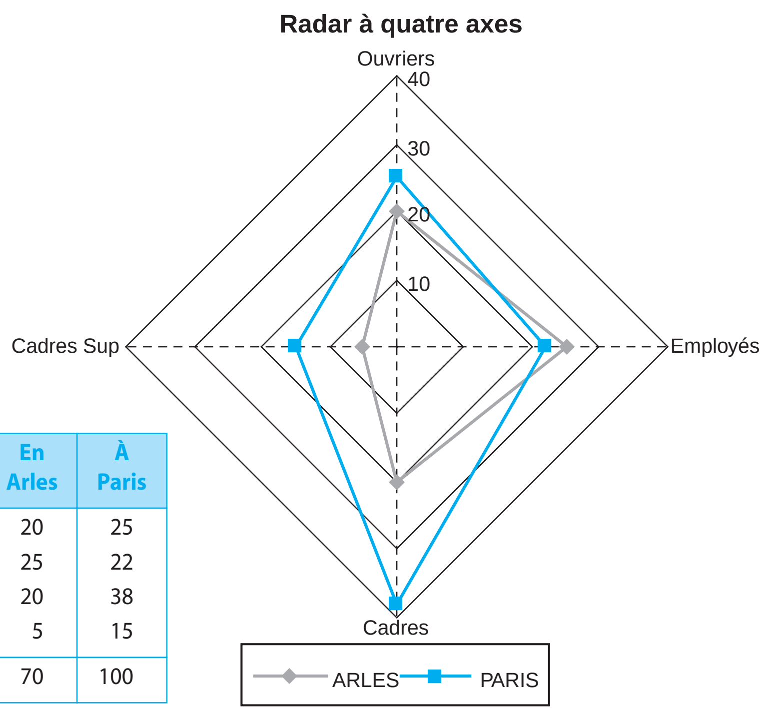
Les comparaisons peuvent être *temporelles* (comparer des chiffres d'affaires à des dates différentes), *spatiales* (comparer des équipements sur des régions, ou des pays différents), *physiques* (comparer des performances pour des imprimantes d'ordinateurs) ou autres. On peut aussi, dans le même diagramme, comparer les modalités *à la moyenne* de ces modalités sur une série d'informations plus large.

Les segments de droite qui relient les valeurs des modalités forment les *pourtours* d'**un polygone**, bien apparent à l'œil, comme on le voit dans la figure ci-après. C'est la **superposition des figures ainsi formées** qui permet les comparaisons. L'effet visuel est renforcé par le **choix de couleurs** dans ces segments de droite, ou même dans les surfaces des polygones[1].

Dans l'exemple ci-après, nous prenons *quatre modalités d'activités professionnelles*. Nous aboutissons donc à quatre axes et nous comparons les valeurs des modalités pour deux agglomérations différentes (comparaison spatiale).

Diagramme en « radar »

Figure 3.5 Répartition du personnel salarié de l'entreprise X par type d'activité au 1er janvier 2007 et par lieu d'établissement



	En Arles	À Paris
Ouvriers	20	25
Employés	25	22
Cadres	20	38
Cadres sup	5	15
TOTAL	70	100

Dans le même ordre d'idée, le graphique suivant (figure 3.6) montre la **décomposition mensuelle sur trois ans**, du chiffre d'affaires d'un commerce, en milliers d'euros. *Douze axes*

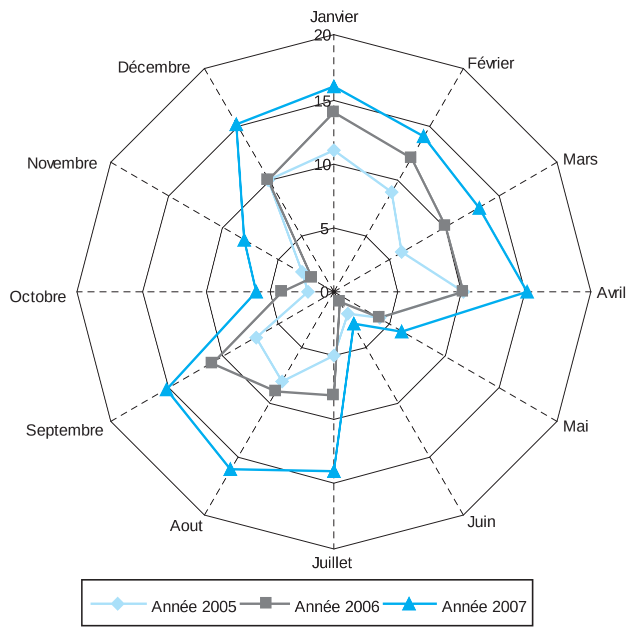
1. Voir les exercices 5 et 6, page 115 et suivantes, pour des résolutions manuelles et avec Excel.

et *trois polygones* rendent compte de l'évolution : il est possible de comparer les différentes performances, mois par mois, sur les trois années[1].

Diagramme en toile d'araignée

Figure 3.6 Chiffre d'affaires du commerce Y (en milliers d'euros)

Mois \ Année	2005	2006	2007
Janvier	11	14	16
Février	9	12	14
Mars	6	10	13
Avril	10	10	15
Mai	4	4	6
Juin	2	1	3
Juillet	5	8	14
Août	8	9	16
Septembre	7	11	15
Octobre	2	4	6
Novembre	3	2	8
Décembre	10	10	15



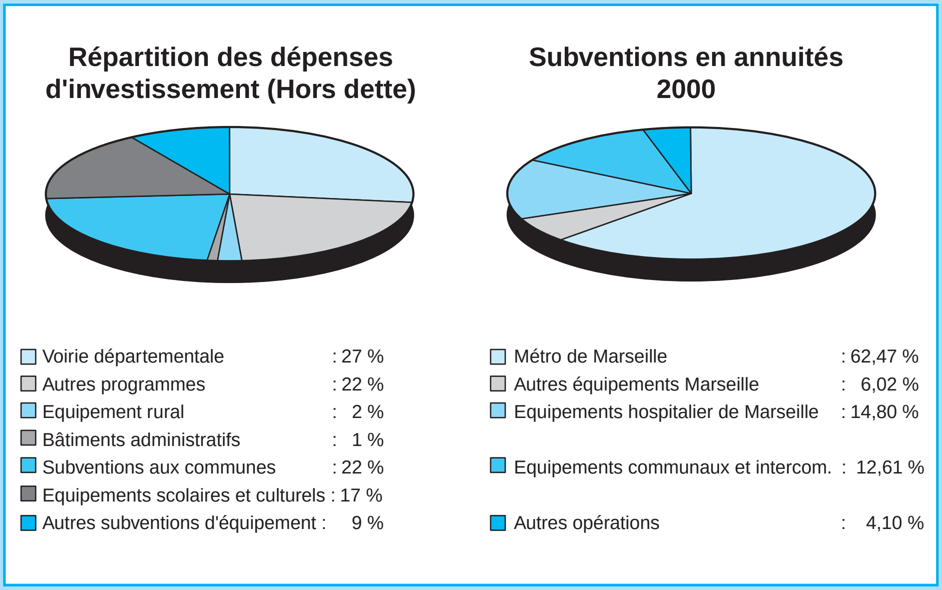
1. C'est-à-dire de comparer 3 « mois de janvier » entre eux, 3 « mois de février » entre eux, etc. On retrouvera cette notion de *correction de variations saisonnières* au chapitre 6, avec la notion de CVS, page 241.

C – Les diagrammes angulaires

Voilà deux exemples de diagrammes angulaires **à une dimension** : le premier nous servira d'*observation*, le second de *construction*.

Diagrammes à secteurs circulaires en 3D

Figure 3.7



Source : *Rapport financier 2000*, Conseil général des Bouches-du-Rhône.

Si vous demandez à un Français non averti : « Comment s'appelle ce genre de figure ? », il y a une probabilité non nulle qu'il vous réponde : « Camembert ! ». Évidemment, vu la quantité de variétés de fromages au lait de vache que nous connaissons et l'image traditionnelle de cette appellation, cette réponse spontanée n'a rien d'étonnant ; mais posez la même question à un Mexicain ou un Danois, vous n'obtiendrez pas forcément la même réponse. Beaucoup d'Anglo-Saxons, adoptant également *une référence gustative*, l'appellent *Pie Chart*, ou encore *Cake Chart*, c'est-à-dire *graphique en tarte* ou *en gâteau*.

Oublions les goûts, mais gardons les couleurs : quel est, finalement, le nom de ce diagramme ? C'est un « **diagramme à secteurs circulaires** ».

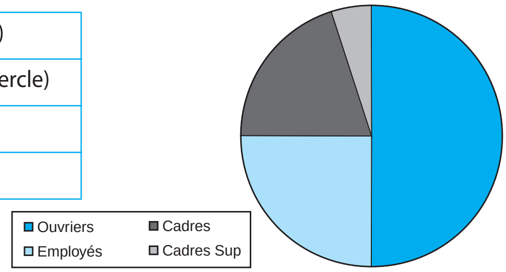
Dans l'exemple ci-dessus, les données chiffrées sont repérées par leurs couleurs dans la légende. On peut aussi, quand il n'y a pas d'effectifs trop faibles, les inscrire dans les secteurs. On peut aussi **agréger les tout petits secteurs** pour équilibrer le dessin.

La construction de ces diagrammes est simple : on multiplie chaque fréquence par 360 (un cercle, c'est 360 degrés). Ainsi, sur l'exemple de l'entreprise X précédent, il vient :

Construction d'un diagramme à secteurs circulaires

Figure 3.8 Répartition du personnel salarié de l'entreprise X par type d'activité au 1er janvier 2007

50 % d'ouvriers, soit 0,5 × 360 = 180 degrés (demi-cercle)
25 % d'employés, soit 0,25 × 360 = 90 degrés (quart de cercle)
20 % de cadres, soit 0,20 × 360 = 72 degrés
5 % de cadres supérieurs, soit 0,05 × 360 = 18 degrés



Les trois diagrammes à secteurs circulaires que nous venons de voir restent *à une dimension*.

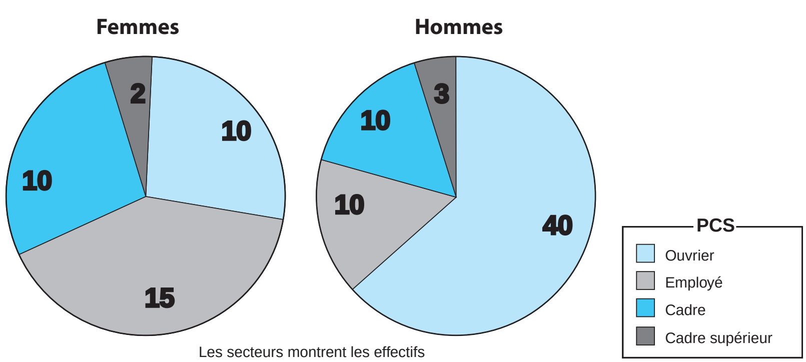
Comment passer à deux dimensions ?

1) La première méthode consiste à aligner plusieurs cercles d'égales surfaces, les uns à côté des autres, comme le montre le diagramme ci-dessous :

Tableau à deux dimensions et double diagramme à secteurs circulaires

Figure 3.9 Répartition du personnel salarié de l'entreprise X par type de PCS et par sexe au 1er janvier 2007

Effectif		Sexe: femme	Sexe: homme	Total
PCS	**ouvrier**	10	40	50
	employé	15	10	25
	cadre	10	10	20
	cadre supérieur	2	3	5
Total		37	63	100



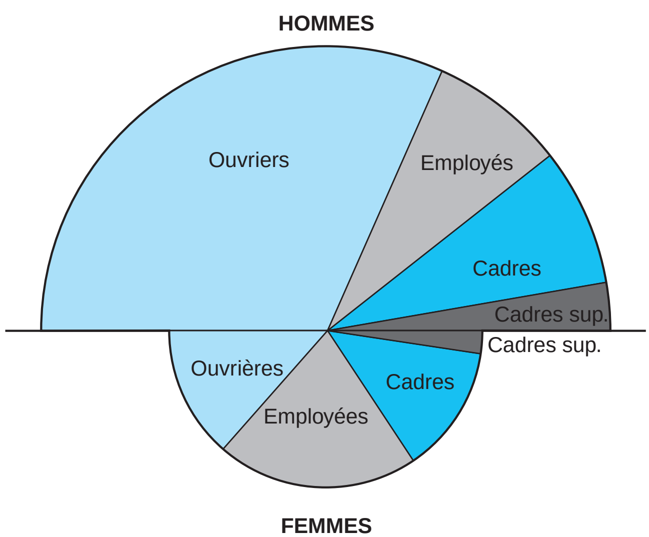
2) La deuxième méthode est de fabriquer des **diagrammes à secteurs semi-circulaires** : toujours dans le cas de l'entreprise X, on peut ajouter la dimension « sexe » en réservant 180 degrés pour les hommes et 180 degrés pour les femmes. Au lieu de multiplier les fréquences par 360, on multipliera par 180.

De plus, la surface de chaque demi-cercle étant proportionnelle à l'effectif, ce type de diagramme permet de **bien visualiser les différences d'effectifs** : on voit bien ici qu'il y a plus d'hommes que de femmes.

Il vient :

Diagramme à secteurs semi-circulaires

Figure 3.10 Répartition du personnel salarié de l'entreprise X par type d'activité et par sexe au 1^{er} janvier 2007 (sur mêmes données que figure 3.9)



Le *diagramme semi-circulaire* est une technique très intéressante pour représenter *deux dimensions* sur un diagramme circulaire.

Notons que ce type de diagramme offre directement la possibilité de *rajouter une information visuelle supplémentaire* (celle de différencier les effectifs) et qu'il est **regrettable** que la plupart des logiciels informatiques ne le proposent pas.

L'application pratique de l'exercice 4, page 109, constitue un intéressant exemple de choix et de construction par étapes de tels diagrammes, dans le domaine qualitatif.

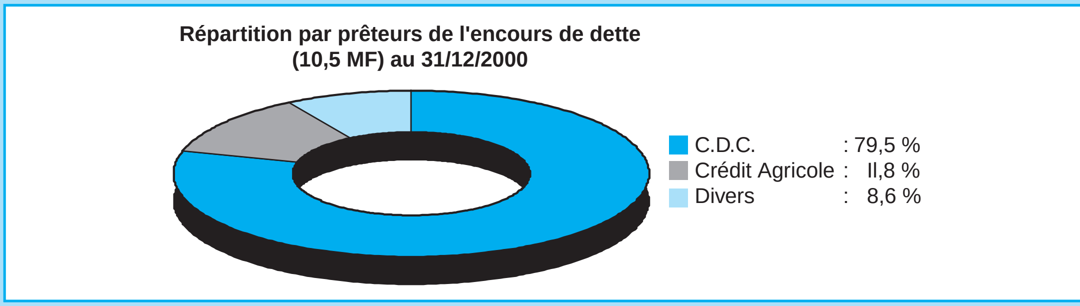
Il existe un autre type de diagramme angulaire, le « **diagramme en anneaux**[1] », qui peut être à une ou à plusieurs dimensions, mais qui est surtout adapté à un petit nombre de modalités : l'anneau, qui représente 100 % des observations, est divisé en secteurs. Chaque secteur possède la proportion de chaque modalité.

1. *Ring chart* ou *Doughnut chart* en anglais.

Voici un diagramme à un seul anneau (trois modalités du caractère « *prêteurs de l'encours de la dette* » dont l'anneau du prêteur majoritaire apparaît nettement) :

Diagramme en anneaux à une dimension

Figure 3.11 Source : *Rapport financier 2000,* Conseil général des Bouches-du-Rhône.



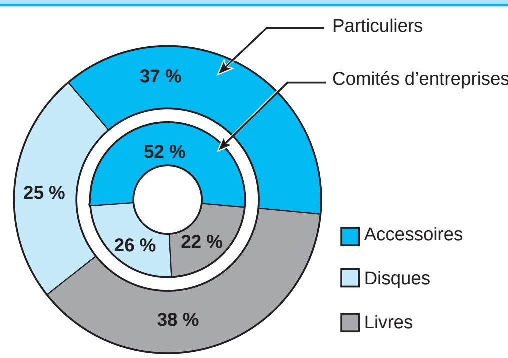
Ce type de diagramme s'adapte au cas de **deux dimensions**, mais n'a d'intérêt que s'il fait ressortir d'assez fortes variations entre les situations de chaque anneau.

Observons un **diagramme à deux anneaux,** qui fait apparaître deux dimensions : il s'agit de chiffres fictifs concernant les ventes *de trois types* d'articles de bureau, en pourcentage du total des ventes, pour un magasin dont la clientèle est essentiellement composée de *deux groupes* d'acheteurs, les *particuliers* et les *comités d'entreprise*.

Diagramme en anneaux à deux dimensions

Figure 3.12 Ventes d'articles de bureau

Vente / Articles	Pour comités d'entreprises	Pour particuliers
Accessoires	52 %	37 %
Livres	22 %	38 %
Disques	26 %	25 %



Les anneaux sont ici *séparés*. Ils peuvent aussi être « accolés » mais, dans ce dernier cas, les couleurs qui se retrouvent juxtaposées doivent être bien contrastées pour que l'on puisse apprécier les différences ; cela n'est pas toujours évident, selon l'endroit où « tombent » les juxtapositions.

Par ailleurs, à plus de trois anneaux, la figure devient pratiquement illisible.

D – Les diagrammes de surface

On arrive au cas *continu*.

Rappelons qu'un caractère **quantitatif** est dit « **continu** » et prend dès lors le nom de « **variable** » (voir chapitre 2) s'il existe une infinité de valeurs intermédiaires, et donc que l'on peut pousser **sa mesure à un certain nombre de décimales**. C'est le cas de la variable « **âge** » par exemple, lorsqu'on la décline par **classes juxtaposées** comme dans les tableaux des page 45 et suivantes du chapitre 2.

C'est aussi le cas des variables de mesure physique (poids, taille, âge), des variables de durée (temps, périodes), des variables d'espace (longueurs, hauteurs, aires), de monnaie (chiffres d'affaires, salaires), par exemple.

Les variables sont regroupées en « classes semi-ouvertes à droite[1] ». Les classes sont *contiguës* : par exemple, pour les âges, la classe [0, 10 ans[est attenante à la classe [10, 15 ans[, qui est elle-même attenante à la classe [15, 40 ans[, si l'on a affaire à des classes *d'amplitudes inégales*.

Le diagramme adapté à ces situations s'appelle « **histogramme** ». Il *ressemble* aux « tuyaux d'orgues » (rectangles), mais le message visuel est complètement différent et il ne faut **surtout pas confondre** les deux supports : l'histogramme donne une image pour laquelle **l'œil humain repère la surface des rectangles**, et non leurs *hauteurs*, comme pour les bâtons ou les tuyaux d'orgues.

Les **classes se jouxtant** obligatoirement, la dernière observation d'une classe est la même (à 2 ou 3 chiffres après la virgule), que la première de la classe suivante. La figure géométrique finale, formée par toutes les classes, se lit comme **LA SURFACE TOTALE** de l'ensemble des rectangles. Elle représente donc l'ensemble des cas possibles que peut prendre le phénomène, c'est-à-dire, en fréquences, **100 %**.

Construisons un histogramme, à partir des données du tableau suivant :

Tableau 3.1 Âge des 110 élèves inscrits dans les filières professionnelles des « Métiers de la Communication » de l'école « Com.C.Chic » durant l'année 2006 (Données fictives)

Âges (x_i)	Nombre (n_i)
[15, 20[	10
[20, 25[	30
[25, 30[	40
[30, 35[	20
[35, 40[	7
[40, 45[	3
	$\Sigma = 110$

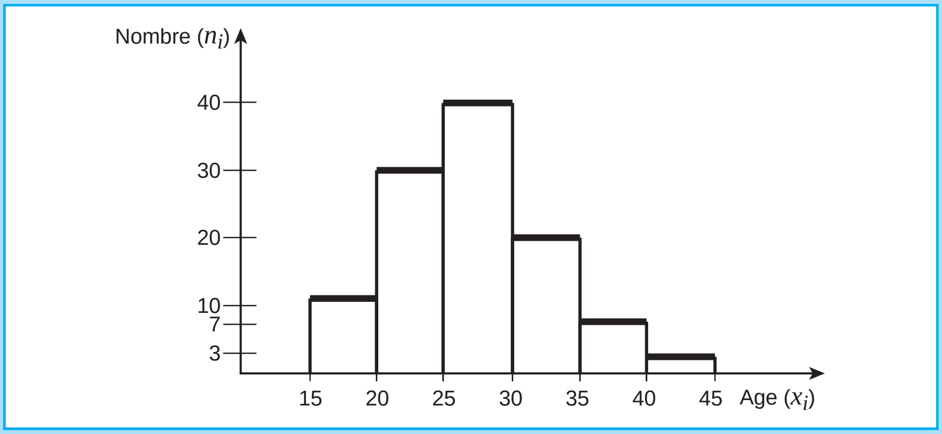
1. Comme on l'a vu dans le chapitre 2, la borne de droite est exclue, de telle manière que l'individu âgé par exemple de 20 ans, ne fait pas partie de la classe [10, 20[, mais de la classe [20, 30[.

- Tout d'abord, nous remarquons que **les classes sont** d'amplitudes **égales**.
- Nous choisissons les échelles des deux axes de coordonnées et portons les valeurs correspondantes. On peut voir que l'étendue de la série est assez large (de [15, 20[à [40, 45[), car il s'agit d'une école « professionnelle », comprenant des filières de formation continue.
- Nous construisons les rectangles de manière à ce que leurs hauteurs soient celles de la colonne « effectifs ».

Il vient :

Histogramme

Figure 3.13 Âge des 110 élèves inscrits dans les filières professionnelles des « Métiers de la Communication » de l'école « Com.C.Chic » durant l'année 2006 (Données fictives)



L'histogramme **est la surface totale**, visuellement repérée, formée à partir des rectangles de même largeur puisque les classes sont d'amplitudes égales.

On en déduit que la surface des trois premiers rectangles correspond à l'idée de *cumuler* les effectifs des trois premières classes (10 + 30 + 40), et l'on pourra dire : « 80 élèves sur 110, soit 73 %, ont **moins de** 30 ans ». On appelle ce résultat une « **fréquence cumulée ascendante**[1] ».

On comprend que la notion « **plus de** » correspondra à une « **fréquence cumulée descendante** » (il n'y a que 10 élèves qui ont plus de 35 ans, par exemple).

Le grand problème en matière d'histogramme réside dans la « **correction des classes inégales** ».

On va saisir toute la puissance de lecture en termes de surfaces en se prêtant à **l'exercice suivant** : imaginons que nous voulions **regrouper les deux dernières classes** ; c'est assez légitime puisqu'elles ont des effectifs très faibles.

On aura dans le tableau *un effectif de 10 élèves (7 + 3)* pour la dernière classe ainsi regroupée.

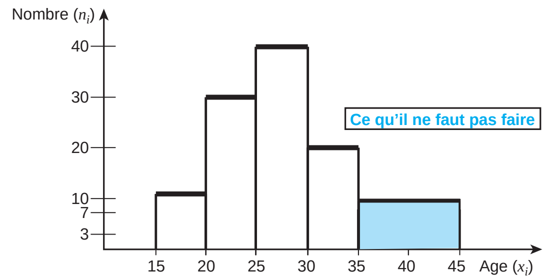
1. On retrouve cette notion de *fréquence cumulée* dans le chapitre 4 pour le calcul de la médiane.

1 – Voici d'abord ce qu'il ne faut pas faire, tout simplement parce que ce serait *faux* !

Histogramme mal corrigé

Figure 3.14 Âge des 110 élèves inscrits dans les filières professionnelles des « Métiers de la Communication » de l'école « Com.C.Chic » durant l'année 2006 (Données fictives)

Âges (x_i)	Nombre (n_i)
[15, 20[	10
[20, 25[	30
[25, 30[	40
[30, 35[	20
[35, 45[	**10**
	$\Sigma = 110$



Pourquoi ce diagramme est-il *faux* ? Parce que l'on n'a pas respecté les surfaces : on a **reporté les hauteurs** (7 + 3 = 10), et, évidemment, *à bases égales*, cela a **augmenté la surface des deux dernières classes** (surfaces colorées). L'histogramme n'a plus le même équilibre (plus la même surface) ; l'œil ne voit plus la même chose : le message visuel est faussé et, par là même, la communication aussi.

C'est un piège dans lequel tombent bien des utilisateurs, d'autant plus facilement que certains logiciels informatiques font également la même erreur (bien vérifier, dans la pratique…).

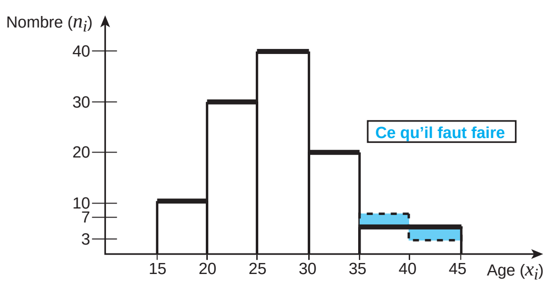
2 – Voici maintenant ce qu'il faut faire :

Puisqu'on a doublé les classes (plus exactement multiplié par 2 « l'amplitude » unitaire des classes), il faut diviser par 2 la hauteur correspondante (10/2 = 5). **La hauteur du dernier rectangle doit donc être de 5**, comme le montre la figure 3.15. Le trait en pointillés reprend la forme de l'histogramme originel et montre que l'équilibre est maintenant conservé (on dit qu'il y a *compensation des aires*).

Histogramme bien corrigé

Figure 3.15 Âge des 110 élèves inscrits dans les filières professionnelles des « Métiers de la Communication » de l'école « Com.C.Chic » durant l'année 2006 (Données fictives)

Âges (x_i)	Nombre (n_i)	Hauteur
[15,20[	10	10
[20,25[	30	30
[25,30[	40	40
[30,35[	20	20
[35,45[	10	**5**
	$\Sigma = 110$	



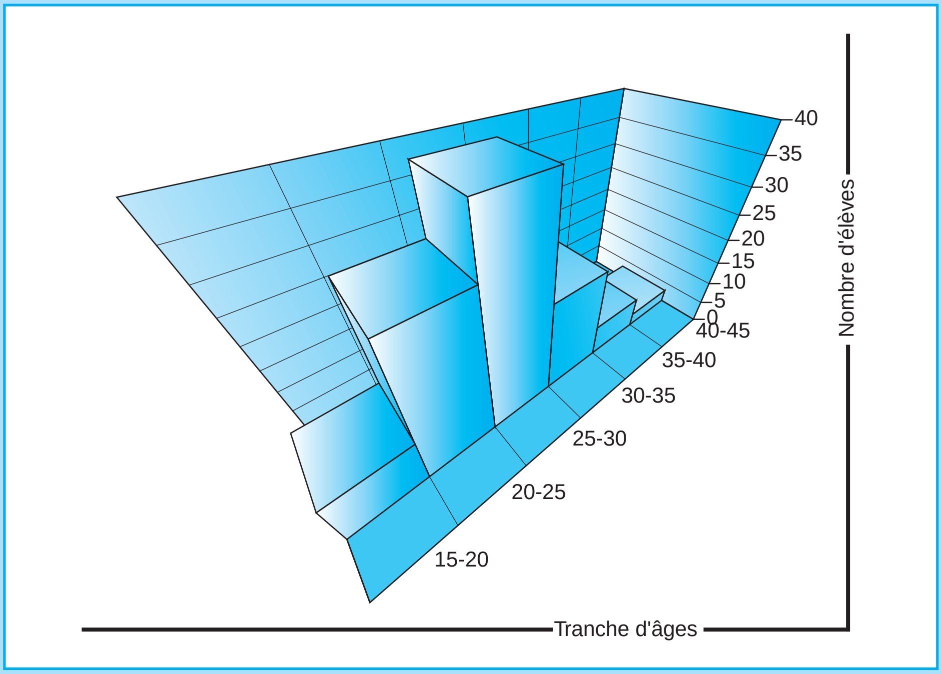
Par ailleurs, au niveau purement *esthétique*, les logiciels informatiques permettent de donner libre cours à notre imagination pour présenter des diagrammes très « **design** » : on trouve une grande variété de dessins très originaux, depuis les formes *à plat*, ou *obliques*, ou *à secteurs éclatés*, jusqu'aux *effets 3D*.

Les résultats sont plus ou moins réussis au *niveau de la communication*. N'oublions jamais qu'un diagramme doit faire passer un message simple, *compris en dix secondes une fois le titre lu*.

L'histogramme suivant, sur les données précédentes (avec classes égales), donne **un aperçu** de ce que l'on peut obtenir en 3D, avec un logiciel de base. Au-delà de l'aspect (relativement) *esthétique*, apprécié ou non, de ce genre de représentation, il faut être **extrêmement prudent**, surtout quand les classes sont **d'amplitudes inégales**, car les logiciels ne sont pas tous suffisamment performants pour effectuer les nécessaires corrections que l'on a vues précédemment.

Histogramme en 3D

Figure 3.16 Âge des 110 élèves inscrits dans les filières professionnelles des « Métiers de la Communication » de l'école « Com.C.Chic » durant l'année 2006 (Données fictives)

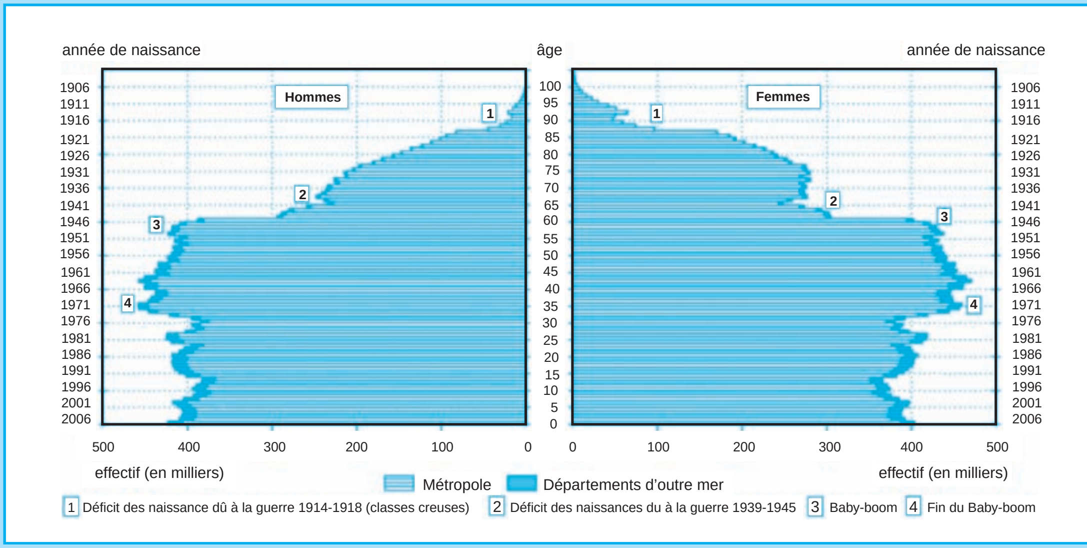


Notons enfin (voir page suivante) que les « **pyramides des âges** », que l'on voit à peu près partout disposées selon une figure traditionnelle, ne sont en fait que des *doubles histogrammes renversés* : un histogramme pour les modalités « hommes » (renversé sur la gauche) et un autre, en face, pour les modalités « femmes » (renversé sur la droite). Rappelons qu'on appelle souvent ces rectangles renversés des « **barres** », tout comme dans le cas des tuyaux d'orgues, probablement pour s'accorder aux *Bar Charts* des Anglo-Saxons.

Dans l'exemple proposé ci-après, les « **barres** » sont très fines, vu l'étendue assez large (heureusement pour notre espérance de vie) de la série considérée.

Double histogramme renversé

Figure 3.17 Bilan démographique 2006 : un excédent naturel record



Source : *INSEE Première*, n° 1 118, janvier 2007.

E – Les diagrammes des séries chronologiques[1]

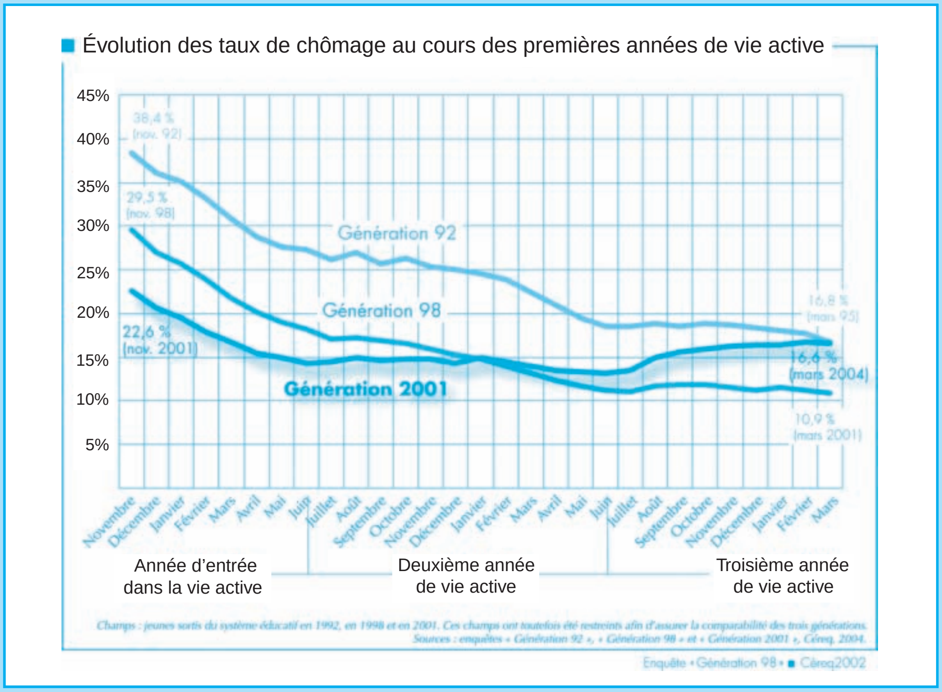
Les représentations de ces séries sont des sortes de diagrammes de hauteur un peu spéciales car, comme leur nom l'indique, les séries chronologiques (*chronos, le temps, en grec*) ne représentent que des **évolutions de phénomènes repérés dans le temps**. Ce temps, ici, est généralement divisé en années, trimestres, mois, ou jours.

Voici (figure 3.18) un exemple de série chronologique où le phénomène est repéré en mois :

1. Voir également chapitre 6, page 249 et suivantes.

Diagramme de séries chronologiques

Figure 3.18 Source : CEREQ, « Enquête génération », 2004.



Profitons de l'observation de ce diagramme pour *construire* une série chronologique :

L'axe des abscisses (horizontal, en bas) repère les échelles de temps. L'axe des ordonnées (vertical) est réservé au report des valeurs du phénomène étudié.

C'est donc **une hauteur que l'on reporte** mais, contrairement aux tuyaux d'orgues, ou aux bâtons, on ne trace pas de lignes verticales ; on indique **uniquement le point correspondant**, et l'on relie par **des segments de droite** (*jamais de courbes*) les points obtenus[1].

Dans la plupart des cas, les valeurs du phénomène (chiffre d'affaires, production industrielle, nombre d'étudiants, ou autres) ne commencent pas à zéro, mais bien plutôt à un chiffre assez important. On débute donc les graduations de l'axe vertical par un chiffre « rond » proche de la valeur observée la plus faible. Cette opération s'appelle la **coupure de l'axe.** Cette coupure doit en principe **être bien précisée** sur le graphique pour éviter les exagérations d'interprétation parfois dues aux effets d'optique. Les unités de mesure (milliers, millions, etc.) doivent également être précisées de manière claire.

1. Cela est la manière la plus traditionnelle de représenter des séries chronologiques. Rappelons qu'en figure 3.6, nous avions utilisé un graphique de type « radar » pour permettre des comparaisons temporelles. Cependant, ces radars ne sont « lisibles » que quand les valeurs mensuelles ou annuelles sont assez bien séparées visuellement, alors que les graphiques traditionnels ne souffrent pas de cet inconvénient.

Soit les données trimestrielles suivantes concernant la *fréquentation touristique*, sur quatre années, pour une région de montagne. Les unités sont des *nuitées d'hôtels*.

Tableau 3.2 Fréquentation touristique en nuitées

Trimestres	Année 2003	Année 2004	Année 2005	Année 2006
Premier	1 090	1 200	1 250	1 320
Deuxième	1 300	1 450	1 390	1 550
Troisième	920	980	1 100	1 110
Quatrième	1 000	1 070	1 160	1 220

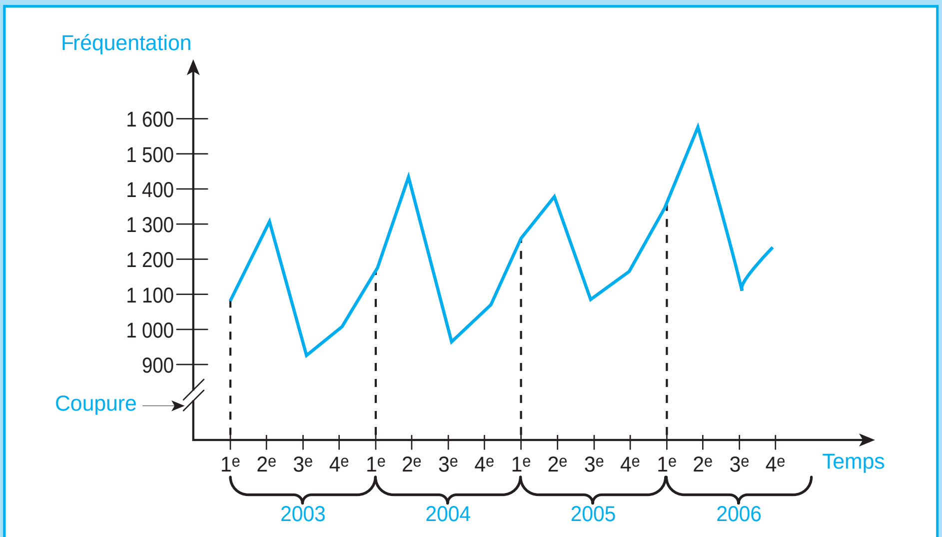
Construisons le diagramme de cette série chronologique. Le temps est partagé en trimestres sur quatre années : nous aurons donc 16 points à reporter et à joindre par des segments de droite (figure 3.19).

Les valeurs s'échelonnent de 920 à 1 550 selon les périodes. Pour ne pas donner une impression visuelle « **d'aplatissement** », il faut *étaler* le plus possible l'intervalle de variation des valeurs du phénomène. C'est le rôle de la coupure de l'axe et du choix de l'échelle des ordonnées : ici, nous repérerons la coupure aux alentours de la valeur 850, pour équilibrer le graphique. Cependant, notons que ce repérage est facile pour un graphique dont on connaît les valeurs comme ici, mais peut réserver des surprises pour un graphique où l'on porte les valeurs au jour le jour, ou mois par mois, en même temps que l'on découvre les données.

Le diagramme est le suivant :

Construction d'un diagramme de série chronologique

Figure 3.19 Fréquentation touristique en nuitées



Le graphique fait ressortir de façon visuelle, mieux que le *radar*, ici, les **variations répétitives**[1] d'année en année dues au rythme de fréquentation hivernale.

La réponse à la question 1 de l'exercice 1, page 250, (chapitre 6), permet de compléter les considérations précédentes sur un cas pratique, chiffré.

§ 2 – LES REPRÉSENTATIONS GRAPHIQUES À TRADUCTION INDIRECTE

On utilise, pour les constructions de ces diagrammes, des **outils de transcription** autres que les instruments que permet la géométrie de base.

- Pour les graphiques *semi-logarithmiques*, on utilise une échelle spéciale ou un *papier* « semi-log ».
- Pour les *diagrammes figuratifs,* on utilise des figurines (pictogrammes).
- Pour les *cartogrammes*, le support est une carte géographique.
- Pour les diagrammes factoriels, on utilise des méthodes de *maximisation de distances* sur des repères factoriels dont on ne connaît pas à l'avance le nom des axes ; ceux-ci étant révélés par l'analyse multidimensionnelle *a posteriori*.

A – Les graphiques semi-logarithmiques

Rassurons-nous, malgré le nom qu'ils portent, l'utilisation de ces graphiques n'impose en aucune manière de savoir calculer des logarithmes, ou d'être imbattable sur l'équation de la fonction « Log ».

Ici, il s'agit **d'un problème d'échelle** : au lieu d'utiliser les échelles arithmétiques, que l'on connaît depuis l'école maternelle, et où la distance entre la graduation 1 et la graduation 2 est la même que celle qui se trouve entre 2 et 3 et ainsi de suite, on utilise un *autre système de graduation*.

À la place de la proportionnalité précédente (**proportionnalité arithmétique**), on repère des graduations proportionnelles aux logarithmes des valeurs (**proportionnalité logarithmique**). Dans le premier cas, la distance entre les graduations 4 et 3, par exemple, donnera : **4 – 3 = 1 cm** sur une feuille de papier où l'échelle est arithmétique. Dans le deuxième cas (échelle logarithmique), cette distance sera : **log 4 – log 3 = 1 cm**. Comprenons bien que les *logarithmes* n'apparaîtront pas sur le graphique ; seules les **valeurs 3 et 4** apparaîtront.

Pourquoi se torturer ainsi ? Quel est l'avantage de remplacer des échelles arithmétiques que l'on connaît si bien ? La réponse tient en l'analyse de *deux inconvénients fondamentaux* des échelles habituelles :

1. Variations dites « saisonnières » ; voir chapitre 6.

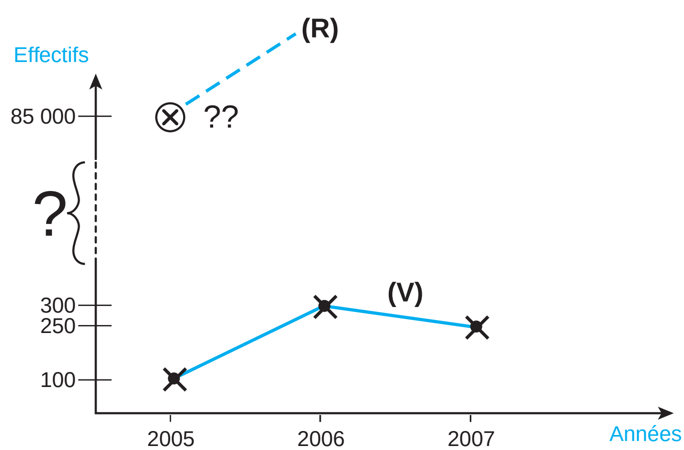
1 – Le cas de la « saturation des axes »

Prenons un exemple : voici les **effectifs des agents de l'État dans la ville** « **V** » **et dans la région** « **R** » sur trois années. Il est intéressant de se demander si ces effectifs augmentent ou diminuent de la même manière (proportionnellement ou pas) en même temps ; et quoi de plus « visuel » qu'un petit diagramme pour en juger ! Essayons de le construire sur une échelle habituelle (arithmétique) avec les données suivantes :

L'impossible représentation graphique des données

Figure 3.20 Effectifs des agents de l'État dans la ville V et dans la région R (Graphique impossible)

Dates	Nombre d'agents en V	Nombre d'agents en R
2005	100	90 000
2006	300	920 000
2007	250	85 000



On aura beau tenter toutes les combinaisons possibles de graduation, **les deux représentations visuelles ne « rentreront » jamais dans le même graphique** : si l'on essaye de faire une échelle verticale (axe des ordonnées) longue de plusieurs mètres… le résultat ne sera pas du tout pratique et n'aura pas un message visuellement efficace ! Si l'on change d'échelle au milieu de l'axe, ce qui est interdit par les règles mathématiques, les proportions ne seront pas respectées.

Quelle est la solution ? La solution, c'est le diagramme semi-logarithmique comme on le verra plus loin.

2 – Le cas des « parallèles non proportionnelles »

Un autre problème peut surgir quand on **visualise des comparaisons graphiques sur support arithmétique** : que peut-on, par exemple, conclure sur l'évolution dans le temps des productions de deux entreprises A et B ? On voit sur le graphique des **lignes parallèles.**

Le problème des variations relatives

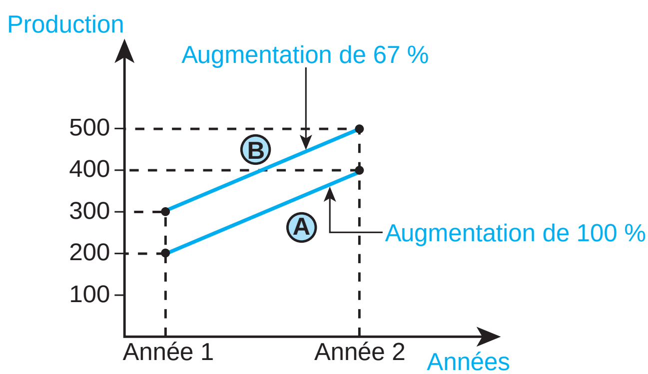
Figure 3.21 Production de deux entreprises A et B

L'entreprise A produit 200, puis 400
L'entreprise B produit 300, puis 500

Visuellement les deux évolutions apparaissent parallèles ; or, les productions ont augmenté dans des proportions bien différentes : 66,7 % et 100 %(doublement).

(500 : 300) – 1 = 0,6666, soit 66,7 %

(400 : 200) – 1 = 1, soit 100 %



Le message visuel est ici *trompeur* : on dirait (à première vue) que les deux productions progressent à la même vitesse. Or, il n'en est rien[1].

L**es échelles arithmétiques traduisent mal les variations relatives.** Quelle est la solution ? La solution, c'est encore le diagramme semi-logarithmique.

1. Un autre exemple de ce type, sur les taux de départ en vacances par catégories sociales, apparaît en correction de l'exercice 9, page 124 et suivantes. On y voit d'abord une évolution parallèle dans le schéma 3.14, qui devient non parallèle, *et donc incorrecte*, dans le schéma final, en échelle arithmétique.

3 – L'utilisation des supports et du papier semi-log

Pour bien comprendre

ÉCHELLES ARITHMÉTIQUES : *C'est le domaine de l'addition* (+)	ÉCHELLES LOGARITHMIQUES : *C'est le domaine de la multiplication* (×)
Elles décrivent des partitions : *Exemple :* l'enfant a mangé 20 % de la tarte aux pommes. S'il en reprend 5 %, il aura pris **deux parts** de tarte aux pommes 20 % + 5 % = 25 % *Il aura mangé un quart de la tarte.*	**Elles décrivent des évolutions :** *Exemple :* l'enfant a grandi de 10 % par an sur trois ans. Son **taux de croissance** annuel est de 0,10 et l'évolution aboutit à : $(1 + 0{,}10) \times (1 + 0{,}10) \times (1 + 0{,}10) = (1 + 0{,}10)^3$ **Sa taille de départ a été multipliée par 1,33**
ON ADDITIONNE OU ON SOUSTRAIT	ON MULTIPLIE OU ON DIVISE
Les échelles arithmétiques traduisent bien les écarts absolus : Si 1 cm vaut 10 €, 2 cm valent 20 €	**Les échelles logarithmiques traduisent bien les variations relatives :** 1/10 = 10/100 = 100/1000 = 10 %

On a vu que les échelles semi-log étaient construites selon le principe suivant : au lieu de reporter une même distance sur l'axe vertical comme c'est le cas pour les graphiques arithmétiques, on reporte une distance proportionnelle aux logarithmes :

10-------------------100---------------------1 000---------------------10 000

Comme ce sont des logarithmes décimaux, les distances physiques (sur le papier) entre les multiples de 10 sont égales.

Nota : il faut se représenter cet axe *en vertical* puisque c'est celui des ordonnées, c'est-à-dire celui de la quantification du phénomène ; l'axe des abscisses reste en échelle arithmétique.

Chaque intervalle entre deux graduations de 10 s'appelle un « **module** ». On voit donc déjà que l'on pourra placer sur la même feuille, par exemple « **à trois modules** » comme sur la figure 3.22, page 88, des quantités de l'ordre de 10, jusqu'à des quantités de l'ordre de 9 999 ; ce qui résout notre problème de **saturation des axes**, et l'exercice de la page 118 devient tout à fait faisable.

Par ailleurs, les propriétés mathématiques des logarithmes font que deux droites, ou deux courbes, qui se présentent de façon parallèle sur le papier **correspondent bien à des évolutions proportionnelles**. Ainsi le deuxième inconvénient décrit page 85 disparaît lui aussi.

Un dernier détail sur les distances : sur un papier arithmétique (papier millimétrique par exemple), si 3 – 4 donnent un centimètre, 5 – 4 aussi, et 145 – 144 aussi. Par contre, les différences entre les *log* ne donnent pas une proportionnalité constante :

- En effet : log 2 – log 1 = 0,30103, la distance pourra être de 3,01 millimètres sur l'échelle.
- De même : log 3 – log 2 = 0,176, la distance pourra être de 1,76 millimètre sur l'échelle.
- Et : log 4 – log 3 = 0,1249, la distance pourra être de 1,125 millimètre sur l'échelle.

On prend donc conscience que les graduations de chaque module sont **de plus en plus resserrées** de bas en haut sur l'axe vertical, ce qui ne pose pas de problème de lecture comme on le verra.

4 – Exercice : construction d'un diagramme semi-logarithmique[1]

Soit les données ci-après, représentant le nombre de visiteurs ayant fréquenté un complexe touristique, dont l'activité débute en 1998 et qui connaît un succès de fréquentation particulièrement remarquable.

Tableau 3.3 Visiteurs ayant fréquenté le complexe touristique « Montmer »

Années	Nombre de visiteurs	Années	Nombre de visiteurs
1998	350	2003	18 500
1999	850	2004	22 000
2000	1 200	2005	95 500
2001	2 000	2006	74 000
2002	20 000	2007	60 000

Tout d'abord essayons de présenter l'évolution sur un graphique habituel (arithmétique)… **On voit vite que c'est impossible**, vu l'écart de valeur entre 95 500 et 350 : l'échelle habituelle ne le permet donc pas.

Passons donc au graphique semi-logarithmique.

Utilisons un **papier** « **semi-log** » du type de ceux qui sont vendus dans le commerce (voir en page 88) pour faire cet exercice. Ensuite on pourra, *si l'on veut*, utiliser des logiciels spécialisés pour construire de tels diagrammes ; mais le but ici est surtout de **comprendre leur logique d'élaboration,** de pouvoir les reconnaître et les lire dans les revues ou ouvrages scientifiques.

Le papier semi-log de la page 88 comprend trois modules dont les ordres de grandeurs ne sont pas repérés. On a donc *le choix de la première graduation*. Dans la mesure où « log 0 » n'existe pas, chaque module commence par un 1.

Si le 1 du premier module représente la valeur 1, le 1 du deuxième module représentera la valeur 10, le 1 du troisième la valeur 100.

Si l'on commence à 100, le 1 du deuxième module sera 1 000 et le 1 du troisième 10 000.

Pour l'exercice, on va considérer que :

- Le 1 du premier module représente la valeur 100.
- Le 1 du deuxième module représentera donc la valeur 1 000.
- Le 1 du troisième, la valeur 10 000, et la dernière valeur de ce module est 99 999.

Ainsi, on peut être sûr que toutes les observations seront présentes dans le graphe.
L'axe des abscisses sera construit selon une logique arithmétique, par exemple en séparant **chaque année de référence par six graduations.**

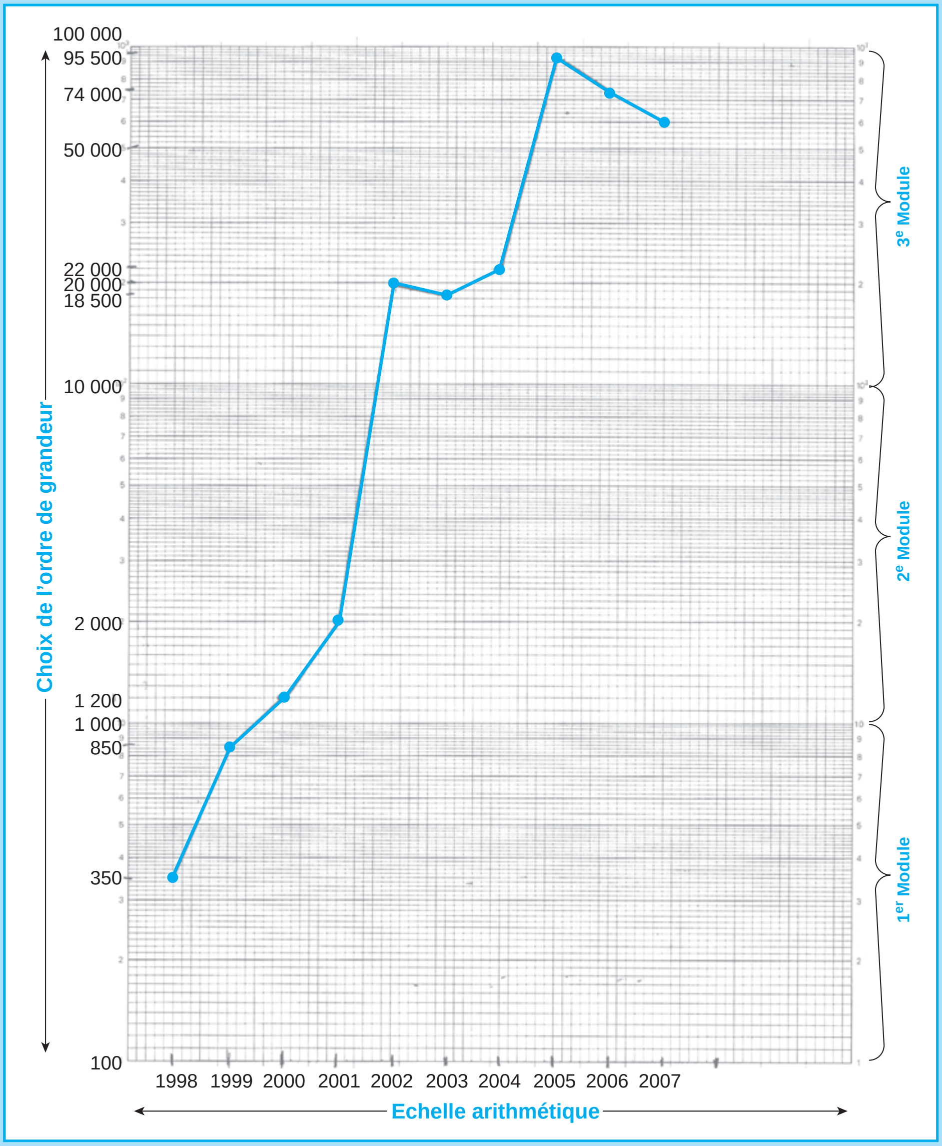
1. La résolution par Excel, sur un autre exemple chiffré, apparaît en fin de correction de l'exercice 7, page 121, « **Construction automatisée** ».

On reportera ensuite, **point par point**, les valeurs du phénomène étudié ; ici le nombre de visiteurs. Il faut prendre garde au deuxième module de ne pas confondre la valeur 1 200 avec la valeur 2 000.

On **écrira les valeurs** en face de chaque graduation choisie et l'on reliera par des segments de droite, et non des courbes, les points ainsi repérés, pour **obtenir l'allure finale du phénomène**, comme le montre le graphique ci-dessous.

Exercice de construction sur papier semi-logarithmique

Figure 3.22 Visiteurs ayant fréquenté le complexe touristique « Montmer »



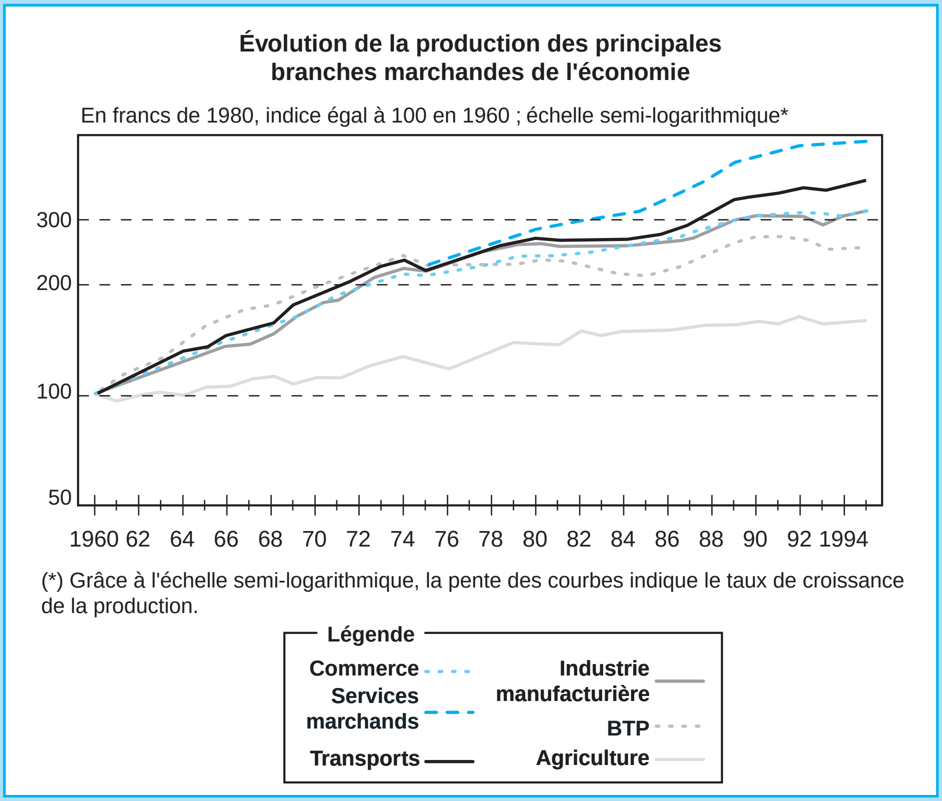
On comprend bien, d'une part, que l'utilisation de papiers semi-logarithmiques évite tout calcul de logarithme et, d'autre part, que **les valeurs se lisent directement sur le graphe comme sur une échelle habituelle**, sans que l'on soit particulièrement gêné par l'effet de resserrement des graduations. Les valeurs sont donc « **à leur place** », et c'est ça l'important.

Il faut savoir que le *passage* par les logarithmes entraîne un effet **géométrique** sur la courbe. Des longueurs égales sont représentées par des rapports égaux dans le cas du semi-logarithmique : **ce sont des variations relatives égales**. Dans le cas arithmétique, **ce sont des variations absolues égales**. La conséquence est que, par exemple, un phénomène dont **le taux de croissance annuel est constant** sera représenté par une **droite** en graphique semi-logarithmique, alors qu'il est représenté par une **courbe** en graphique arithmétique[1].

Ci-après, à titre *d'exemple de lecture*, un diagramme « semi-log » de l'INSEE.

Reconnaître un diagramme semi-logarithmique

Figure 3.23 Référence : « L'essor des services publics depuis les années 1960 », in *INSEE Première*, n 498, décembre 1996.



Source : INSEE, Comptes de la nation.

1. Pour de plus amples détails, et pour une « mise en situation » pratique de construction de ces diagrammes, se reporter à l'exercice 8, page 121, « Construire un diagramme semi-logarithmique ».

B – Les diagrammes figuratifs

Ici, nous sommes loin des applications géométriques simples, dans la mesure où ce sont **des illustrations** qui sont à la base de la transmission du message visuel. Ces types de diagrammes se décomposent en deux groupes, les **pictogrammes** et les **cartogrammes**.

1 – Les pictogrammes

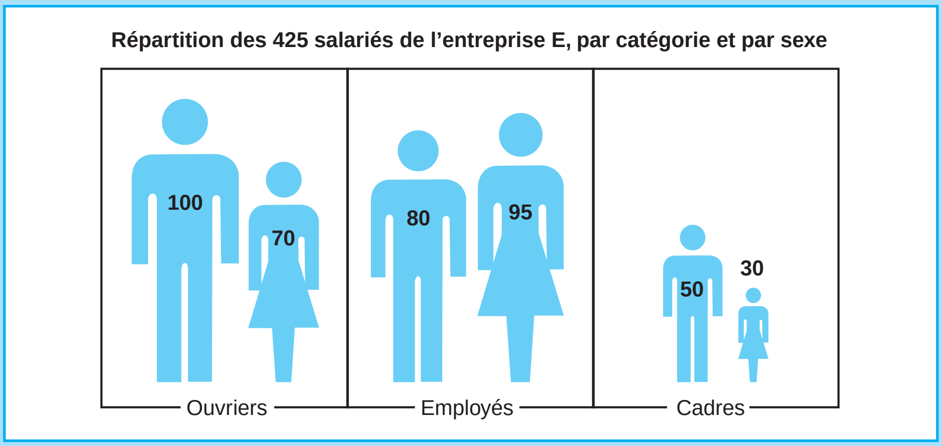
Ces diagrammes sont des **dessins colorés**, plus ou moins travaillés, à base d'images dont les plus ou moins grandes *corpulences*[1] rendent compte des quantités étudiées. Une méthode assez répandue, en ce domaine, est d'employer des **figurines** symbolisant des hommes et des femmes pour décrire des caractères tels que l'emploi, l'âge, le salaire, ou autre.

On obtient, ainsi, au moins deux dimensions : celle du caractère sexe, s'ajoutant à celle du caractère étudié.

La figure suivante simule une situation où les femmes sont moins employées que les hommes dans deux catégories socioprofessionnelles sur trois.

Diagramme de surface et corpulence

Figure 3.24



La rigueur statistique impose que, comme ici, **la surface, et non la hauteur des figurines** soit proportionnelle à l'effectif… Il est désolant de voir que, dans bien des études, cette norme n'est pas respectée.

Les symboles utilisés sont variés : depuis l'image reproduisant l'objet lui-même (automobiles, maisons, produits de consommation, etc.) jusqu'à l'image suggérant une identification en rapport avec le sujet traité dans l'étude, facile à repérer (des fûts de différentes tailles, pour représenter l'augmentation du prix de pétrole, des cheminées d'usine pour l'emploi industriel,

1. Nous employons le terme de « corpulence » pour spécifier qu'il s'agit de représentations de *surface* et non de *hauteur*.

etc.). Généralement, le chiffre représentant la quantité est inclus dans la figurine, ou immédiatement proche, s'il n'y a pas assez de place.

Parfois, des dessins quelque peu « torturés » rendent l'identification involontairement difficile : la signification est loin d'être immédiate et la communication passe moins bien. Ici encore, la règle de la lecture en *dix secondes une fois le titre lu* ne doit pas être oubliée.

2 – Les cartogrammes

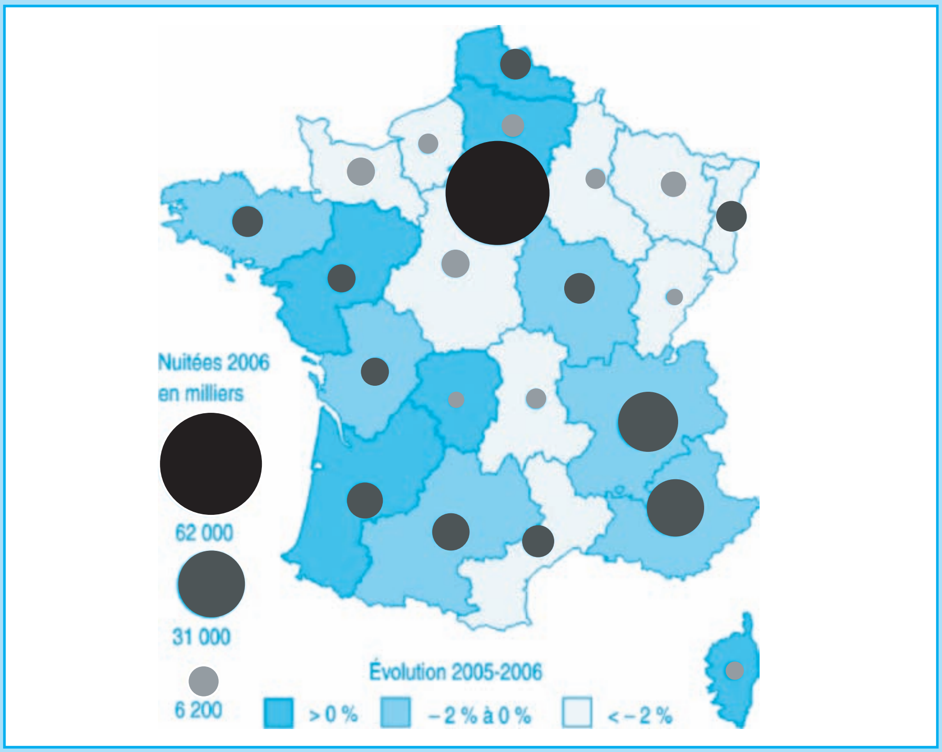
Les informations sont **repérées sur une carte**, soit par des **couleurs**, correspondant à des modalités bien indiquées dans la légende du diagramme, soit par des **cercles** de différents rayons. Dans ce dernier cas, la représentation est encore une **représentation de surface**, et non de hauteur.

La deuxième dimension est évidemment la **dimension spatiale** (villes, régions, pays, autres territoires).

Le cartogramme suivant, tiré de l'INSEE, *combine deux informations* (s'ajoutant à l'aspect régional) : **l'évolution annuelle** repérée par des couleurs, et **les volumes** du phénomène, repérés par des cercles :

Cartogramme de surface

Figure 3.25 Nuitées dans les hôtels en 2006

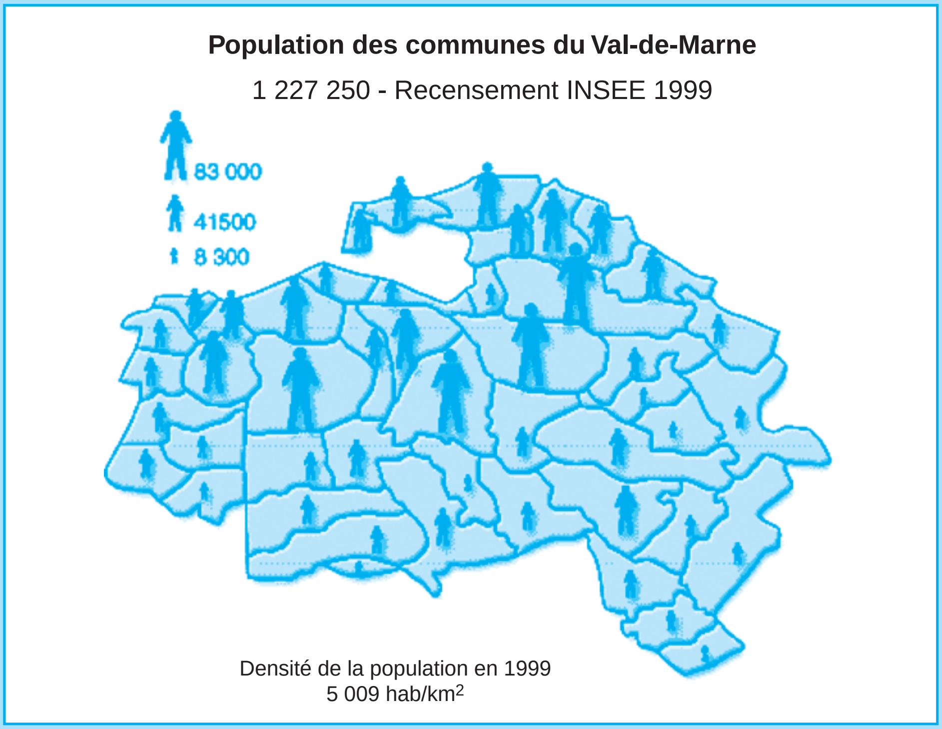


Source : INSEE, Direction du tourisme, partenaires régionaux, enquête de fréquentation hôtelière.
Référence : « L'hôtellerie et les campings en 2006 », in *INSEE Première*, n° 1 125, mars 2007.

Une représentation originale, « mixte » **cartogramme-pictogramme**, est donnée dans le diagramme suivant :

Cartogramme-pictogramme

Figure 3.26



Source : INSEE, Recensement RGP, 1999.

C – Les graphiques des analyses factorielles

Nous entrons ici dans le vaste domaine de « **l'analyse de données** » et de l'application de la *statistique descriptive mathématisée* aux méthodes dites « **multicritères** », ou « **multidimensionnelles** ». Ces méthodes aboutissent à des représentations graphiques modernes[1] et originales, qu'il faut savoir reconnaître quand on les rencontre dans des revues spécialisées ou des rapports. Dès lors, bien plus que l'aspect mathématique de construction, c'est **l'aboutissement graphique** qui nous intéresse dans ce paragraphe.

Les diagrammes que l'on a vus jusqu'à présent décrivent un phénomène par un ou deux caractères, très rarement plus. On parle de *dimensions,* comme on l'a vu dans le chapitre 2.

1. Bien que la base de ces méthodes fût connue depuis le début du XX[e] siècle (K. Pearson, C. Spearman, LL. Thurstone, H. Hotelling), il faut attendre les années 1965, et la puissance de calcul des ordinateurs, pour qu'elles s'enrichissent et deviennent applicables dans la pratique, notamment avec J. L. Benzécri et son équipe.

Par exemple, on peut statistiquement « décrire » le niveau d'un étudiant, par une dimension : ses *notes à l'examen*. Or, la réalité est bien plus complexe : un étudiant ne peut pas être résumé par une note, ou par un ensemble de notes ; il est bien autre chose (heureusement). On dit qu'il ne peut pas être « **réduit** » à **une** dimension. Son **profil** est plus large : il est « **pluri** » ou « **multidimensionnel** ». Ce qui est vrai pour l'étudiant est vrai pour bien des choses que l'on observe dans la réalité : un produit ne se réduit pas à un prix, un pays à sa production intérieure brute (PIB), un logement au nombre de ses pièces, un ouvrage littéraire à son nombre de pages, etc.

Le problème réside dans *l'appréhension* de ces multiples facettes qui constituent le « **profil** » de ce que l'on veut étudier.

Il est, dans la vie, un grand nombre de cas où l'observation statistique de deux individus, **repérés par un unique caractère**, apporte une vision d'éléments en opposition, alors que les profils sont proches : en matière d'élection présidentielle, par exemple, on remarque que certains électeurs d'extrême gauche ont un profil qui n'est pas très éloigné de ceux de certains électeurs d'extrême droite, alors que le caractère « position sur l'axe politique » les oppose au maximum ; de même, deux candidats à un emploi peuvent apparaître très proches d'un point de vue professionnel, et l'on aura du mal à les départager, alors que leurs profils, culturel, artistique ou de mode de vie, sont complètement différents. La vision statistique *unidimensionnelle* demande donc souvent à être dépassée.

Il faut donc synthétiser, et définir des « **critères** », les plus explicatifs et les plus objectifs possibles. Le choix des critères et des caractères qui les *mesurent* (dans le cas quantitatif), ou qui les *traduisent* (dans le cas qualitatif) se fait en fonction de la connaissance du terrain ou par enquête préalable.

Mais comment ne pas se perdre dans tous ces caractères ? C'est tout l'intérêt de ces *analyses de données* : on « **ne sélectionne pas** » tel caractère ou tel autre, mais on « **combine** » ces caractères entre eux, par des procédés algébriques[1], de manière à ce qu'ils rendent compte de leur dépendance ou indépendance. Ils font alors apparaître des « **facteurs communs** » ou au contraire des « **facteurs spécifiques** ». Le terme de *facteur* donne son nom à celui de l'analyse « **factorielle** » (titre du présent paragraphe).

Quelle **logique** dicte ces combinaisons ? C'est une logique purement « **visuelle** » à la base, on pourrait même dire « photographique » : en effet, si l'on veut prendre un cliché photographique d'un imposant édifice architectural, par exemple, on va se déplacer, se baisser, reculer, se hisser, afin d'obtenir l'image la moins *plate* possible. En d'autres termes, que fait-on ? On cherche à **maximiser les distances** entre les différents points de l'édifice, pour que *la photo présente le meilleur relief*.

C'est cette notion de *distances à maximiser* que l'on reprend dans les analyses factorielles. C'est moins direct et moins intuitif que dans le cas de la photo, car ce sont *des moyennes de carrés de distances entre des points de projection*, mais l'idée est la même : **étaler au maximum l'image[2] sur le diagramme**.

1. *Combinaisons linéaires simples* de type le plus souvent matriciel, dont le détail des calculs algébriques ne nous intéresse pas ici.
2. Étaler l'image revient à maximiser sa dispersion ; on dit que l'on maximise son « **inertie** ».

On parvient ainsi à *construire* de nouveaux caractères. Ces nouveaux caractères présentent quatre avantages :

D'abord, ils sont en nombre moins important que ceux que l'on avait choisis au début ; on a « **réduit** » les données.

Ensuite, ils « **synthétisent** » (ils résument) la réalité de départ en concentrant les informations, et par là même les caractères obtenus deviennent plus riches de sens, plus « **significatifs** ».

Ils sont parfois, nouveaux, non prévus, inattendus ; ce sont alors des caractères **cachés**, que l'analyse aura révélés.

Ils sont, enfin, **indépendants** les uns des autres, par construction, c'est-à-dire que chacun apporte une information *spécifique* et non influencée par les autres caractères.

Il existe plusieurs types d'analyses factorielles[1] ; la plus connue est « **l'analyse en composantes principales** » (**ACP**). La combinaison qui fait apparaître le maximum de facteurs communs s'appelle la « **première composante principale** », ou « **premier axe** » de la représentation graphique. On trouve ensuite la deuxième composante ou deuxième axe, et ainsi de suite.

Il faut bien comprendre que ces axes, révélés par la méthode, n'ont *pas encore de nom* ; tout au plus auraient-ils un nom commun à tous les caractères qui ont participé, à des degrés divers, à leur fabrication, ce qui risquerait de donner une appellation complètement incompréhensible. Il faut donc « **interpréter les axes** » : cette étape est à la fois la plus délicate et la plus pertinente de l'analyse.

L'*interprétation des axes factoriels* doit aboutir à une **signification concrète**, comprise par les utilisateurs sans aucune équivoque. On peut s'aider de calculs spécifiques[2] que fournit un grand nombre de programmes informatiques, mais cela ne permet, en aucune manière, de se passer de la connaissance du terrain : l'utilisateur y est le plus souvent le meilleur guide.

La méthode graphique consiste, ensuite, à **observer la disposition des points** sur le diagramme final, en opérant **des regroupements**, pour en tirer le profil le plus adapté à chaque axe, et c'est le point le plus intéressant de la lecture.

Prenons deux exemples : le premier **fictif** pour évoquer un *essai d'interprétation* des axes factoriels en ACP, le second sur **données réelles** pour revenir dans le concret, et savoir **reconnaître et lire** ce genre de diagramme.

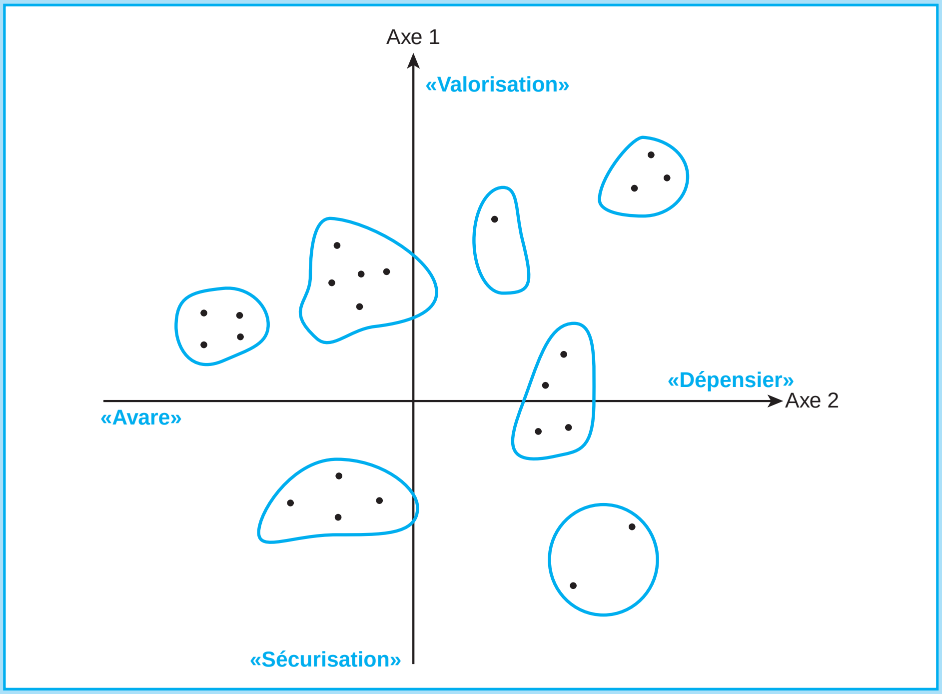
- Pour le premier, imaginons une chaîne de boutiques de vêtements de sport qui s'intéresse aux motifs subjectifs d'achat par types de clientèle. L'enquête portera sur des **critères** qui seront repérés par des **caractères à une dimension** : la catégorie socioprofessionnelle (classée en *supérieure, haute, médiane, basse, très basse*), l'âge, le nombre d'enfants, l'habitat et la distinction *urbain-rural*, le secteur d'activité (*agriculture, commerce, industrie, bâtiment, mode, théâtre, etc.*), le sport pratiqué, le goût pour les voyages (*organisés ou non, de type aventure ou non, courts séjours ou non, etc.*), les types de véhicules du ménage, les types de commerces fréquentés (*alimentaires, restauration, sorties*), et bien d'autres questions que l'on pourrait imaginer, sans que, bien sûr, le questionnaire ne tourne à l'interrogatoire policier.

1. Analyses : *canonique, discriminante, hiérarchique, des correspondances*.
2. Calculs de corrélation (voir ce mot au chapitre 7) entre chaque axe et ses composantes, les caractères initiaux.

On pourrait obtenir le diagramme suivant, en quatre quadrants, où **les regroupements** en « pavés » représentent les *proximités* de profil des groupes recensés, et les axes représentent les « socio-styles » d'achat. Les axes sont échelonnés **par opposition** (un comportement en haut, son contraire en bas).

Type de regroupements graphiques en analyse factorielle

Figure 3.27 Motifs subjectifs d'achat par types de clientèle



Selon les résultats de l'analyse, le premier axe (ici, l'axe vertical) **pourrait être interprété** comme repérant le caractère **subjectif** de la volonté d'achat : on oppose *la sécurisation* (« achetez ce pull, il est parfait, bien fabriqué, avec des laines de qualité irréprochable ») à la *valorisation* (« achetez ce pull, cela marquera votre appartenance à une élite »).

Le deuxième axe (ici l'axe horizontal) pourrait concerner, non pas le pouvoir d'achat réel du client, mais son **comportement** en la matière, en opposant le *style économe* (« avare » dans le graphique) au *style dépensier*.

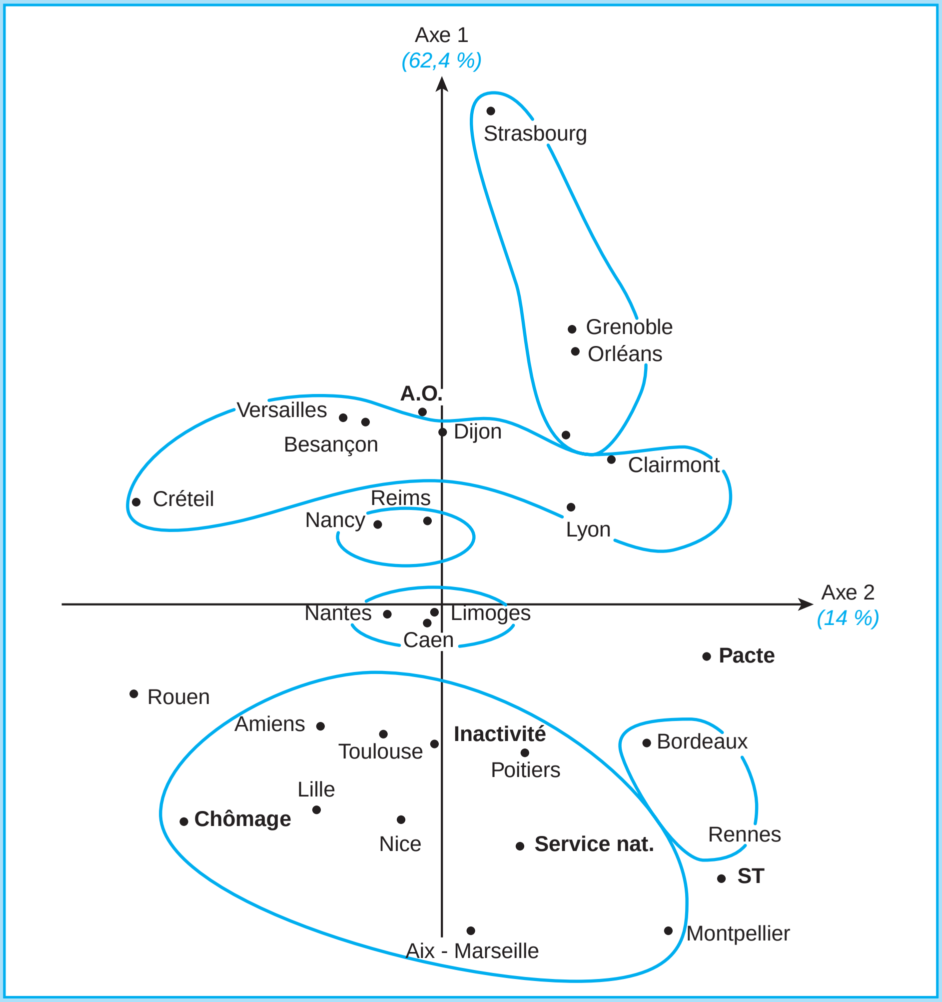
- Le deuxième exemple est tiré d'une étude du CEREQ[1], dans laquelle l'auteur, sur la base d'une enquête passée en 1980, analyse les critères de **sortie des jeunes de l'appareil éducatif et leur insertion dans les différentes régions** de France. Les régions sont ici repérées en

1. CEREQ, « Quand les jeunes formés au niveau CAP-BEP abordent la vie active dans les régions », in *Formation Emploi* n° 8, par Jean Biret, octobre-décembre 1984.

25 académies. Les groupes concernés (*garçons*, dans cet exemple) sont déclinés selon les niveaux de formation, le statut d'actif (ayant effectivement un emploi, ou chômeur), le statut d'inactif (en stage, ou sans activité déclarée), le statut d'occupé (en emploi formation dans le cadre des contrats du 3e pacte pour l'emploi, ou en dehors de la réglementation de ce pacte), le statut de militaire du service national. Le diagramme de type ACP[1] est présenté ci-dessous (figure 3.28) :

Un exemple réel d'analyse en composantes principales

Figure 3.28 Insertion des jeunes sortis de l'appareil éducatif dans différentes régions de France



Source : CEREQ, *Formation Emploi* n° 8, octobre-décembre 1984

1. C'est, plus exactement, une « **analyse des correspondances** », de type Benzécri, qui peut être considérée comme une ACP sur des modalités de caractères qualitatifs, codées en valeurs entières numérisées.

Les regroupements par proximités de points concernent les régions (plus exactement les 25 académies). L'interprétation du premier axe factoriel (ici vertical) oppose, en haut l'activité occupée en dehors du pacte (**AO**) et, en bas, le service national (**Service nat.**).

Le deuxième axe oppose les formes **pacte** et **chômage** de façon assez nette. L'auteur ajoute que « l'**inactivité** et les formes de type stage (**ST**) ne contribuent quasiment pas à la formation des axes 1 et 2 ». On voit bien, par exemple, dans le pavé du bas, **Lille proche de Nice**, mettant en relief que ce n'est pas le degré de latitude géographique qui explique la relative faiblesse des taux d'occupation masculins, hors pacte.

Loin des conclusions et résultats chiffrés, nous ne présentons ici cet exemple pris dans la réalité que dans un but uniquement **pédagogique** : celui de **nous familiariser** avec la lecture des représentations graphiques d'analyses factorielles.

Hors l'exemple du CEREQ, ajoutons un mot sur les limites de ces analyses factorielles : si la projection des points et le calcul des axes obéissent à une *indiscutable* rigueur mathématique, il n'en va pas toujours de même de *l'utilisation* qui en est faite ensuite : les axes demandent à être « interprétés » et, dans bien des cas, leur « signification » n'est pas directement évidente, ce qui risque de conduire à des approximations discutables. Il vaut mieux alors choisir d'autres modes d'investigation. De plus, quand on regroupe les points les plus proches, il est bon de *vérifier* que les deux axes rendent compte suffisamment bien de la *réalité du terrain*. À défaut, on risque de considérer comme voisins des points risquant d'être très éloignés sur le troisième axe, voire le quatrième. Un peu comme quand les anciens ont regroupé dans chaque constellation les étoiles qui leur paraissaient proches, vues de la terre, mais que l'on a découvert en fait très éloignées les unes des autres dans l'espace sidéral.

Exercices sur les graphiques, diagrammes et autres messages visuels

***Nota* important :** Les résolutions de calculs par Excel 2007, selon une démarche « pas à pas », sont proposées à la fin des réponses concernées.

Le lecteur non débutant, ayant une connaissance suffisante des logiciels, pourra aisément se passer d'un certain nombre de détails que nous développons ici, et trouver facilement des raccourcis.

Exercice 1 : Cas d'une variable quantitative discrète : choisir le message visuel adapté

Énoncé :

Reprenons les résultats de l'exercice n° 1 page 53 précédent, concernant le nombre d'enfants par familles salariées de l'entreprise Machin. Le tableau final que nous avions construit était le suivant :

Nombre d'enfants par famille des salariés de l'établissement Machin

Nombre d'enfants par famille (x_i)	Nombre de familles (n_i)
0	9
1	7
2	10
3	0
4	4
	TOTAL 30

Source : *Recensement réalisé par le comité d'entreprise de l'établissement Machin durant le mois de décembre 2009, auprès du personnel salarié.*

Question 1 Construisez le graphique le plus adapté à cette situation.

Question 2 Si l'on bénéficiait de résultats quantitatifs supplémentaires donnant la distance habitat – travail (en quatre classes kilométriques) des familles enquêtées, pourrait-on les faire apparaître sur ce graphique ? Choisissez deux méthodes de représentation et tracez les graphes en inventant des valeurs cohérentes. Justifiez votre réponse.

Question 3 Au vu de ces résultats cohérents, présentez le tableau à deux dimensions dont les valeurs vous ont servi à fabriquer les deux diagrammes précédents.

Corrigé de l'exercice 1

Réponse à la question 1

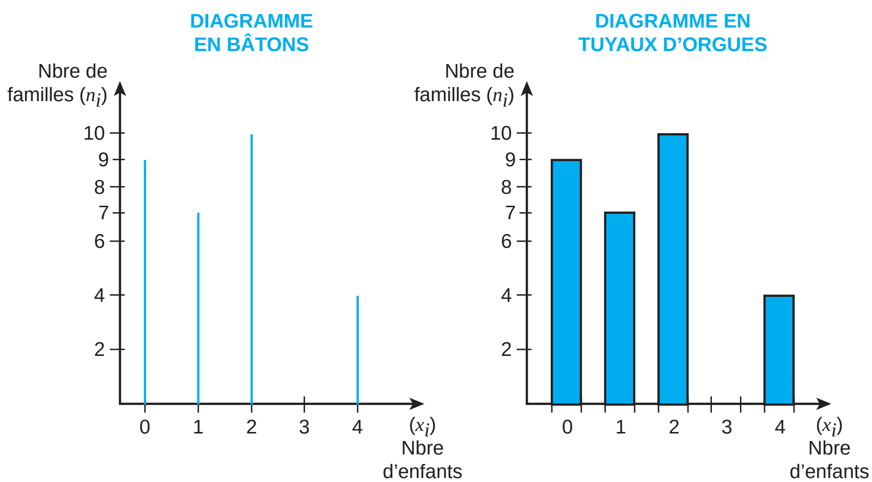
Le caractère est de type **quantitatif discret** (les modalités sont des nombres entiers, et l'on passe de 1 à 2, puis à 3, et ainsi de suite, de façon discontinue, sans chiffres après la virgule) ; rappelons que dans les cas quantitatifs, on assimile les deux termes *caractère* et *variable*.

Le diagramme de référence peut être un « diagramme en bâtons » ou « en tuyaux d'orgues »[1], permettant de reporter sur l'axe vertical **des hauteurs** ; ces hauteurs sont les valeurs des effectifs étudiés, c'est-à-dire le nombre de familles correspondant à chaque modalité.

Ces deux possibilités de graphique sont présentées dans le schéma 3.1 ci-dessous.

Schéma 3.1

Nombre d'enfants par famille des salariés de l'établissement Machin



Source : recensement réalisé par le comité d'entreprise de l'établissement Machin durant le mois de décembre 2009, auprès du personnel salarié.

Notons que le diagramme en tuyaux d'orgues est généralement préféré à celui en simples bâtons, car les bâtons, assez fins, ne se perçoivent pas toujours très bien visuellement.

1. Ou également en « barres » horizontales ; voir page 69.

Résolution par Excel

Rappelons que le lecteur ayant une connaissance suffisante des logiciels pourra aisément se passer d'un certain nombre de détails développés ici. Nous restons volontairement à un niveau « débutant ».

A) Saisir dans une colonne les nombres correspondant aux effectifs (9 ;7 ;10 ;0 ;4), puis sélectionner **Insertion** dans la barre des menus, puis **Graphique colonnes**, ensuite ce qu'Excel appelle **« histogramme » 2D groupé**[1]. Le diagramme apparaît.

B) Cliquer sur **Disposition** dans la barre des menus, puis, dans la barre d'outils (au-dessous), choisir **Titre du graphique**, puis **Au-dessus de graphique**. Remplacer « titre » par *Nombre d'enfants par famille* dans la zone du graphe, et cliquer hors du graphe.

Revenir sur **Disposition**, et choisir dans la barre d'outils **Titre des axes**. Cliquer sur **Titre de l'axe principal**, puis sur **Titre en dessous de l'axe**.

Dans la zone du graphe, déplacer (en faisant apparaître le curseur) « titre de l'axe » vers la droite, et remplacer ces mots par **Nombre d'enfants**.

Cliquer à nouveau sur **Disposition**, **Titre des axes**, et répéter la procédure pour l'axe vertical, dont le titre est **Nombre de familles**, en élargissant la case si nécessaire.

Cliquer hors du graphe et continuer si besoin par des choix sur les **couleurs**, les **styles** et autres détails plus ou moins personnels.

1. Qui n'est pas un histogramme, mais un diagramme en tuyaux d'orgues.

Réponse à la question 2

Si l'on bénéficiait de ces résultats supplémentaires, on doublerait l'information, et on passerait donc à un diagramme **à deux dimensions.**

La première méthode consisterait à *partitionner* les « tuyaux » en autant de segments verticaux que de classes d'âge fournies, et à utiliser des couleurs ou des trames diverses pour les différencier visuellement (voir le schéma 3.2 page suivante).

Si l'on a par exemple les quatre classes de distances suivantes **[de 0 à 5 km [, [de 5 à 10 km [, [de 10 à 20 km [, [20 km et plus [** pour chaque famille, on peut déterminer la part de chaque classe de distances afin de partitionner chaque tuyau d'orgue **proportionnellement**, comme on le voit dans le diagramme page suivante.

Résolution par Excel

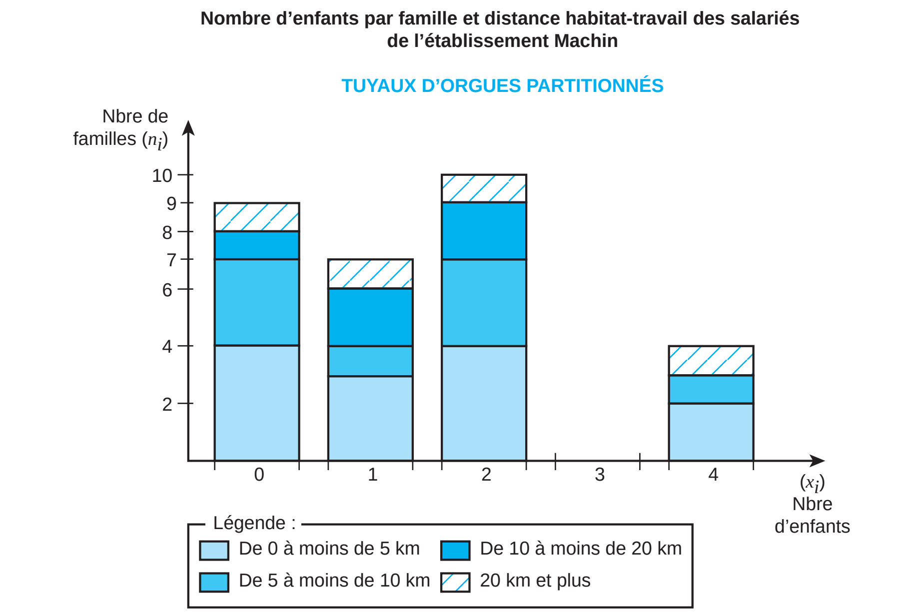
A) Saisir dans une première colonne le titre [0 ;5[km et les nombres d'effectifs suivants, que l'énoncé ne donne pas mais que nous proposons[1] (4 ;3 ;4 ;0 ;2), puis dans une deuxième colonne, le titre [5 ;10[km et les nombres (3 ;1 ;3 ;0 ;1), puis dans une troisième colonne, le titre [10 ;20[et les nombres (1 ;2 ;2 ;0 ;0), et enfin dans une quatrième, le titre [20 et plus[et les nombres (1 ;1 ;1 ;0 ;1). **Sélectionner** la plage des 4 colonnes et cliquer sur **Insertion** dans la barre des menus, puis sur **Graphique colonnes**, ensuite sur ce qu'Excel appelle **« histogramme » empilé**[2]. Le diagramme apparaît.

Choisir le **style**, la **couleur**, et les différentes options que l'on désire.

B) Recommencer la procédure B) du paragraphe précédent pour les noms des axes.

1. Ce sont ceux de la réponse à la dernière question, du tableau pages 101 et 102, proposés pour respecter les totaux des données.
2. Qui n'est pas un histogramme, mais un diagramme en tuyaux d'orgues.

Schéma 3.2



Source : recensement réalisé par le comité d'entreprise de l'établissement Machin durant le mois de décembre 2009, auprès du personnel salarié

La deuxième méthode consiste à « accoler » les classes de hauteurs *proportionnelles aux effectifs partiels*, comme le montre le schéma 3.3 (page suivante). Cela est possible ici car nous n'avons que quatre types de classes de distances ; à plus de quatre, le diagramme serait trop chargé.

Attention, il faut que les *hauteurs* dans ce graphe soient en cohérence avec les *partitions* choisies précédemment

Ainsi, par exemple, pour les familles sans enfants, le schéma précédent indiquait que 4 familles habitaient entre 0 et 5 km du lieu de travail : il faut donc que la hauteur du premier tuyau dans le diagramme ci-dessous soit égale à 4 unités, sinon le diagramme sera faux..

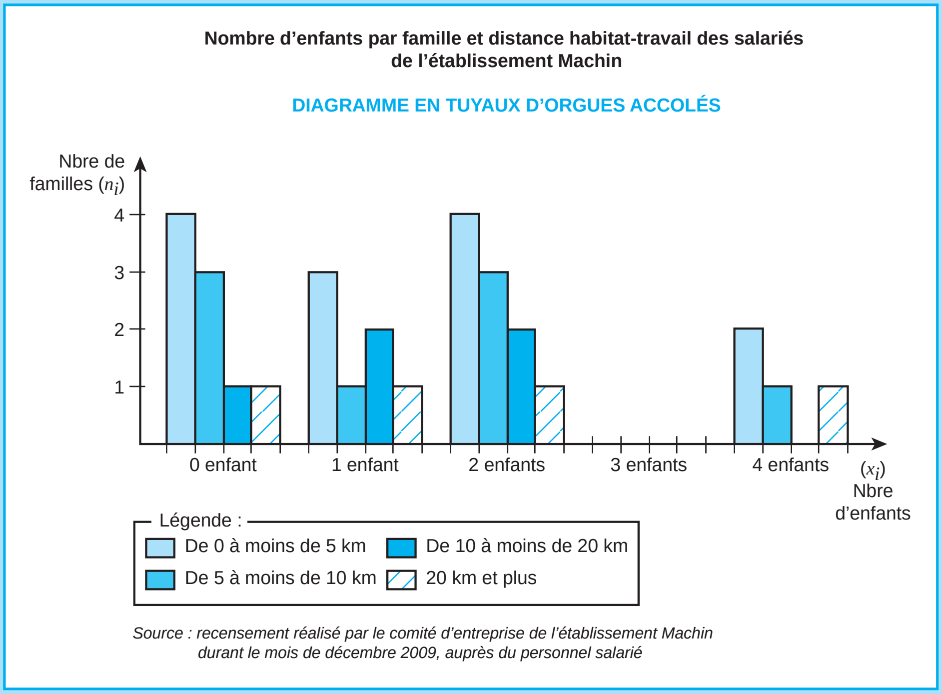
Résolution par Excel

A) Saisir dans une première colonne le titre [0 ;5[km et les nombres d'effectifs suivants, que l'énoncé ne donne pas mais que nous proposons[1] (4 ;3 ;4 ;0 ;2), puis dans une deuxième colonne, le titre [5 ;10[km et les nombres (3 ;1 ;3 ;0 ;1), puis dans une troisième, le titre [10 ;20[et les nombres (1 ;2 ;2 ;0 ;0), et enfin dans une quatrième, le titre [20 et plus[et les nombres (1 ;1 ;1 ;0 ;1). **Sélectionner** la plage des 4 colonnes et cliquer sur **Insertion** dans la barre des menus, puis sur **Graphique colonnes**, enfin sur ce qu'Excel appelle **« histogramme » groupé**[2]. Le diagramme apparaît. Choisir le **style**, la **couleur**, et les différentes options que l'on désire.

B) Recommencer la procédure du paragraphe B) précédent pour les noms des axes.

1. Ce sont ceux de la réponse à la dernière question, du tableau page 102, proposés pour respecter les totaux des données.
2. Qui n'est pas un histogramme, mais un diagramme en tuyaux d'orgues.

Schéma 3.3



Réponse à la question 3

Ici nous avons choisi, *à titre d'exemple*, les valeurs suivantes qui **garantissent la cohérence** entre les deux diagrammes :

Km / Nb enfants	[0,5[	[5,10[	[10,20[	20 et plus	Total
0	4	3	1	1	**9**
1	3	1	2	1	**7**
2	4	3	2	1	**10**
3	0	0	0	0	**0**
4	2	1	0	1	**4**
Total	**13**	**8**	**5**	**4**	**30**

Exercice 2 : Cas d'une variable continue : choisir le message visuel adapté

Énoncé :

L'association de défense des consommateurs ADC teste la durée de vie d'appareils électroniques identiques de type TTX. Les résultats sont les suivants : on compte 9 appareils qui durent entre 300 et 400 heures ; 13 qui durent entre 400 et 500 heures ; 25, entre 500 et 600 heures ; 12, entre 600 et 700 heures ; 6, entre 700 et 800 heures.

Question 1 Quelle est la variable étudiée (nombre d'heures ou nombre d'appareils) ? S'agit-il d'un caractère qualitatif, quantitatif, discret, continu ? Dans ce cas, quel est le type de diagramme qui s'impose pour traduire correctement le message visuel ?

Question 2 Construisez à main levée le diagramme, en précisant : le nom des axes et l'effectif total ; s'agit-il d'un diagramme de hauteur ou de surface ?

Corrigé de l'exercice 2

Réponse à la question 1

La variable étudiée est *le nombre d'heures*. Il s'agit d'un caractère *quantitatif continu*, que l'on regroupera dans les cinq classes ouvertes à droite suivantes :

[300,400[; [400,500[; [500,600[; [600,700[; [700,800[.

Comme les bornes de classes sont adjacentes[1], les rectangles du graphe seront adjacents ; il n'y aura pas d'espace entre les barres verticales. Le diagramme est donc *un histogramme*, diagramme de référence dans le cas des variables continues.

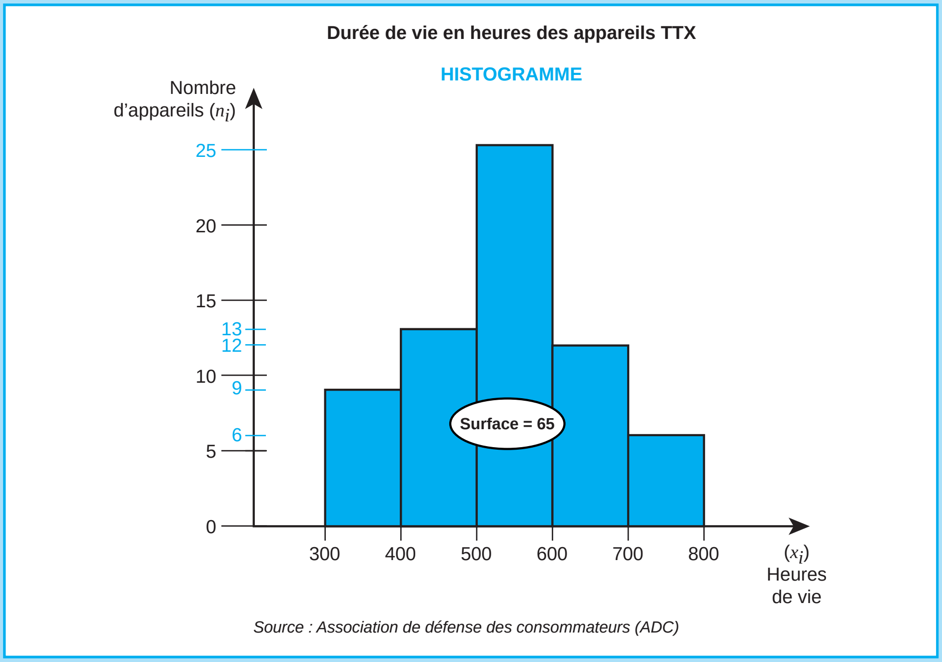
Réponse à la question 2

La construction de l'histogramme est assez simple dans ce cas, **car les classes sont d'amplitudes égales** (intervalles égaux de 100 heures). On placera les centaines d'heures en abscisses (axe horizontal) et les effectifs en ordonnées (axe vertical), en choisissant une échelle centimétrique assez large pour bien faire apparaître les différences de hauteur, comme indiqué dans le schéma 3.4.

Le diagramme est par définition un **diagramme de surface**. La surface totale dans l'histogramme représente 100 % du phénomène, c'est-à-dire la totalité des effectifs, soit 65 appareils testés. *L'effectif total* est donc égal à 65, comme on peut le voir dans le graphe.

1. On exclut les chiffres des bornes de droite de chaque modalité, pour reprendre le même chiffre en borne de gauche de la modalité suivante (voir page 77).

Schéma 3.4



Résolution par Excel

Rappelons que le lecteur ayant une connaissance suffisante des logiciels pourra aisément se passer d'un certain nombre de détails développés ici. Nous restons volontairement à un niveau « débutant ».

Saisir les données {9 ;13 ;25 ;12 ;6} en lignes ou en colonnes, et sélectionner la plage.

Cliquer sur **Insertion** dans la barre des menus, puis sur **Graphique colonnes**, ensuite sur ce qu'Excel appelle **« histogramme » groupé**[1]. Un diagramme provisoire apparaît.

Cliquer sur **Intervertir lignes/colonnes** dans la barre d'outils (troisième icône du menu **Création**), l'histogramme s'affiche.

Choisir le **style** et les **couleurs** et procéder comme précédemment pour les titres des axes.

1. Qui n'est pas un histogramme, mais un diagramme en tuyaux d'orgues.

Exercice 3 : Histogramme à classes d'amplitudes inégales, choix de « l'amplitude unitaire » et conservation des surfaces

Énoncé :

Votre travail dans l'entreprise Machin, qui vous avait accepté en stage d'essai lors de l'exercice n° 2 page 54 précédent, avait consisté à créer un tableau sur les salaires mensuels. Ce tableau est le suivant. La variable est continue. On voit que les amplitudes de classes sont très inégales, ce qui va certainement poser un problème à la construction du diagramme.

Répartition des salariés de l'entreprise E par tranches de salaires mensuels

Salaires en € (x_i)	Effectif (n_i)
[1 100, 1 300[	19
[1 300, 1 700[	26
[1 700, 1 900[	33
[1 900, 2 400[	45
[2 400, 2 900[	27
[2 900, 3 200[	24
[3 200, 3 800[	12
[3 800, 4 100[	8
[4 100 et plus[	6
	TOTAL 200

Source : *Service comptabilité DRH de l'entreprise E.*

Question 1 Donnez le nom du diagramme adapté à cette situation.

Question 2 Tracez le graphe en faisant attention aux classes inégales. Montrez à quoi aurait ressemblé le diagramme si l'on avait, par erreur, oublié de corriger les hauteurs.

Question 3 Quelle est la part des salariés recevant moins de 3 200 € ? Quelle est celle recevant plus de 3 200 € ? Exprimez ces parts sur le graphe.

Corrigé de l'exercice 3

Réponse à la question 1

La série est constituée de classes de salaires (variable continue). Comme dans l'exercice précédent, le diagramme est donc *un histogramme*, message visuel de référence dans le cas des variables continues.

Réponse à la question 2

Les classes de salaires *ne sont pas d'amplitudes égales* ; dès lors, on se trouve dans le cas de la nécessaire **correction des hauteurs**, comme on l'a vu page 78 et suivantes du présent chapitre.

En effet, une classe qui serait par exemple *deux fois plus large* que la précédente induirait à la construction, si l'on ne faisait pas attention, un rectangle dont la base serait *deux fois plus importante* que pour le rectangle précédent. Dès lors, ce rectangle aurait une surface double de ce qu'il devrait avoir, et le diagramme de surface serait alors faux[1].

Il faut donc, pour satisfaire à la **condition de conservation des surfaces**, prendre une **amplitude unitaire** (ici par exemple **200 €**) et diviser par deux la hauteur si l'on a multiplié par deux l'amplitude de la classe ; de même, il faut diviser par trois la hauteur si l'on a multiplié par trois l'amplitude de classe, etc. On aurait pu tout aussi bien choisir une amplitude unitaire de **100 €**, et appliquer les proportions de la même manière[2].

Il faut bien comprendre que, **comme dans une carte géographique**, c'est l'*échelle* adoptée qui rend compte du message visuel.

Dans une carte géographique au 1/100 millième, par exemple, il va de soi que chaque distance sur la carte représente une distance proportionnelle à ce qu'elle est dans la réalité ; si l'on veut « faire un zoom » sur une partie de la carte, on change alors d'échelle, et donc d'unité.

Ici, c'est la même chose : *l'amplitude unitaire*, une fois choisie, détermine **l'échelle de l'histogramme**, de telle façon qu'un millimètre carré sur le papier représente un pourcentage constant de l'effectif total.

La surface totale formée par l'ensemble des rectangles représente 100 % de l'effectif.

Dans un histogramme, ce sont les surfaces et non pas les hauteurs qui sont proportionnelles aux réalités du phénomène étudié : un histogramme est un **diagramme de surface**.

Le tableau préalable à la construction nécessite deux colonnes supplémentaires : l'une (a_i) repérant les amplitudes de chaque classe, l'autre (h_i) indiquant les *hauteurs corrigées proportionnellement*, comme on le voit ci-dessous :

Salaires en € (x_i)	Effectif (n_i)	Amplitudes (a_i)	Hauteurs (h_i)	Fréquences
[1 100, 1 300[	19	**200**	**19**	9,5 %
[1 300, 1 700[	26	**400**	**13**	13 %
[1 700, 1 900[	33	**200**	**33**	16,5 %
[1 900, 2 400[	45	**500**	**18**	22,5 %
[2 400, 2 900[	27	**500**	**10,8**	13,5 %
[2 900, 3 200[	24	**300**	**16**	12 %
[3 200, 3 800[	12	**600**	**4**	6 %
[3 800, 4 100[	8	**300**	**5,33**	4 %
[4 100 et plus[	6	**400**	**3**	3 %
	TOTAL 200			100 %

1. Attention à certains logiciels qui ne respectent pas cette logique mathématique et qui, par là même, confondant histogramme et diagramme en tuyaux d'orgues, rendent des schémas faux.
2. Il est quelquefois suggéré, mais pas forcément indispensable, de choisir pour amplitude unitaire le *PGCD des amplitudes* (pour ceux qui se souviennent de leurs cours de collège), **ici 100**, ce qui aboutit aux mêmes résultats.

Notons que la dernière classe se présente en « **amplitude indéfinie** ». Dans ce cas, l'observateur choisit une borne de droite *guidée par le bon sens*, ou par sa *connaissance du terrain*. Nous choisissons ici une amplitude de *400 euros*.

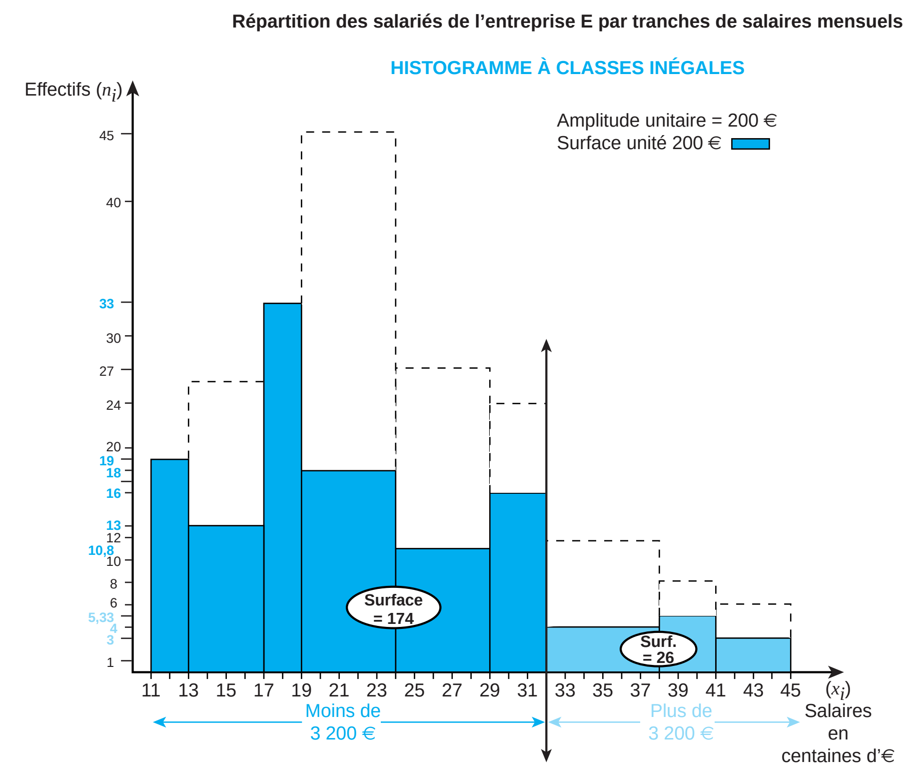
Le choix de l'amplitude unitaire ayant été *déterminé à 200*, et la colonne (a_i) étant remplie, il suffit de **diviser les effectifs par le rapport des amplitudes** et de reporter ces valeurs dans la colonne (h_i) des hauteurs corrigées. Ainsi, à la deuxième ligne, on divise l'effectif de 26 par le rapport d'amplitude (400/200 = 2) pour obtenir la hauteur corrigée du rectangle qui sera reportée sur l'axe vertical du graphe. À la deuxième ligne, on peut donc inscrire 13, et ainsi de suite pour toutes les autres lignes.

L'histogramme – *surfaces par unités d'amplitudes* – apparaît en couleurs dans le schéma 3.5.

On a reporté *en pointillés* ce à quoi aurait ressemblé le diagramme si l'on avait, *par erreur*, oublié de corriger les hauteurs : la non-proportionnalité et le déséquilibre des surfaces sautent aux yeux.

En effet, on aurait ainsi raisonné *hors unités d'amplitude*, et dès lors on aurait pu croire que la classe de salaires contenant le maximum de salariés était la classe [1 900,2 400[.

Schéma 3.5



* Rappel : les traits en pointillés montrent "le piège du diagramme erroné" (ne conservant pas les surfaces)
Source : Service comptabilité DRH de l'entreprise E

Cela aurait traduit le message visuel suivant : un groupe de 45 personnes touche entre 1 900 et 2 100 €, et encore un autre groupe de 45 personnes touche entre 2 100 et 2 300 €, et encore un demi-groupe de 45 personnes touche entre 2 300 et 2 400 euros, alors qu'il n'y a **qu'un seul groupe de 45 personnes** dans cette classe.

On voit bien que c'est la classe précédente [1 700,1 900[qui comprend le maximum de personnes – 33 par unité d'amplitude –, et qui correspond au rectangle le plus haut.

Résolution par Excel

Rappelons que le lecteur ayant une connaissance suffisante des logiciels pourra aisément se passer d'un certain nombre de détails développés ici. Nous restons volontairement à un niveau « débutant ».

Le logiciel Excel n'est *pas idéal* pour corriger des classes d'amplitudes inégales. Par contre, si l'on reproduit comme dans l'exercice 2 précédent les **valeurs corrigées** du tableau, en répétant autant de fois qu'il le faut le nombre correspondant à l'amplitude unité[1], Excel retournera un diagramme exact.

Dès lors, les nombres **à saisir en colonnes** sont :

{19 ;19 ;13 ;13 ;13 ;13 ;33 ;33 ;18 ;18 ;18 ;18 ;18 ;10,8 ; 10,8 ; 10,8 ; 10,8 ; 10,8 ; 16 ;16 ;16 ;4 ; 4 ; 4 ; 4 ; 4 ; 4 ;5,33 ; 5,33 ; 5,33 ;3 ;3 ;3 ;3}

Cliquer sur **Insertion** dans la barre des menus, puis sur **Graphique colonnes**, ensuite sur ce qu'Excel appelle **« histogramme » groupé**[2]. Un diagramme provisoire apparaît.

Cliquer sur **Intervertir lignes/colonnes** dans la barre d'outils (troisième icône) ; l'histogramme s'affiche.

Choisir le **style** et les **couleurs**, et procéder comme précédemment pour les **titres des axes**.

1. PGCD des amplitudes (plus grand commun diviseur).
2. Qui n'est pas un histogramme, mais un diagramme en tuyaux d'orgues.

Réponse à la question 3

Combien de salariés touchent « moins de 3 200 € » ? Ceux qui touchent de 2 900 à 3 200 € (borne exclue à droite), plus ceux qui touchent de 2 400 à 2 900 €, plus ceux qui touchent de 1 900 à 2 400 €, et ainsi de suite jusqu'à la première classe.

Il faut donc produire les **effectifs cumulés croissants** (tout comme on avait produit les *fréquences cumulées ascendantes* dans le chapitre précédent). Mais ce qui est intéressant ici c'est **de traduire cette réalité sur le graphe, de façon visuelle immédiate**.

Le tableau se complète de la façon suivante :

Répartition des salariés de l'entreprise E par tranches de salaires mensuels

Salaires en € (x_i)	Effectif (n_i)	Effectifs cumulés croissants	Effectifs cumulés décroissants
[1 100, 1 300[	19	19	200
[1 300, 1 700[	26	45	181
[1 700, 1 900[	33	78	155
[1 900, 2 400[	45	123	122
[2 400, 2 900[	27	150	77
[2 900, 3 200[	24	**174**	50
[3 200, 3 800[	12	186	**26**
[3 800, 4 100[	8	194	14
[4 100 et plus[	6	200	6
	TOTAL 200		

Source : Service comptabilité DRH de l'entreprise E.

On lit sur le tableau que **174 salariés** gagnent strictement **moins de 3 200 €**. On a sommé les effectifs **de haut en bas**, et les valeurs croissent à la lecture de haut en bas.

Sur l'histogramme, cette part est représentée visuellement par l'ensemble des rectangles colorés en bleu foncé, c'est-à-dire par la **surface située à gauche du point d'abscisse 3 200**.

Les salariés qui gagnent « **plus de 3 200 €** » sont aussi ceux qui gagnent « au moins 3 200 € » (borne comprise) : ce sont ceux de la classe [3 200,3 800[, plus ceux de la classe de la ligne suivante, etc.

Il faut donc, dans une autre colonne, sommer les effectifs **de bas en haut** : c'est la colonne des *effectifs cumulés décroissants*. Les valeurs croissent de bas en haut à la lecture.

Il y a **26 salariés** qui gagnent plus de 3 200 €, borne comprise.

Cette part est visuellement représentée dans le diagramme par l'ensemble des surfaces des rectangles situées **à droite** du point d'abscisse 3 200, coloré en bleu clair.

Exercice 4 : Construire un diagramme pour des données qualitatives peu nombreuses

Énoncé :

Dans l'université UUU, on compte 39 % d'étudiants en droit, 20 % d'étudiants en économie-gestion, 13 % en physique, et 28 % en biologie.

Question 1 Quel type de diagramme est le plus adapté dans ce cas pour transmettre le message visuel ? Rappelez-en ses hypothèses de construction et de lecture.

Question 2 Construisez le diagramme.

Question 3 Le secrétariat des inscriptions nous donne maintenant la répartition des étudiants par sexe : il y a 5 826 filles sur 10 000 étudiants au total. La part des filles dans chaque discipline est respectivement : 73 %, 55 %, 11 % et 62 %. Construisez le diagramme adapté.

Question 4 Pourrait-on utiliser un message visuel plus complet en conservant la logique angulaire ?

Corrigé de l'exercice 4

Réponse à la question 1

On pourrait *a priori* choisir un *diagramme en tuyaux d'orgues*, puisque le caractère est qualitatif, mais on ne le fera pas pour deux motifs :

- D'abord, parce qu'il n'y a aucune raison pour placer en premier lieu telle ou telle discipline sur l'axe horizontal, ce qui semblerait, à tort, la privilégier.
- Ensuite, parce que les modalités ne sont pas trop nombreuses (quatre seulement).

Dès lors **on préférera un diagramme à secteurs circulaires**.

Rappelons que dans un diagramme à secteurs circulaires, c'est **l'ouverture de l'angle de chaque secteur** qui traduit visuellement l'intensité du phénomène.

Un cercle comprend 360° ; un angle ouvert à 90° par exemple signifie que le secteur comprend le quart des observations. Cette idée de *quart*, donc de 25 %, renvoie à la notion de fréquence relative (f_i), et « l'angle au centre » est donc égal à : $\boldsymbol{\alpha_i = 360° \times f_i}$.

Réponse à la question 2

On peut faire rapidement un tableau de travail indiquant les ouvertures des angles, si l'on a quelques craintes à élaborer directement le diagramme. Il vient :

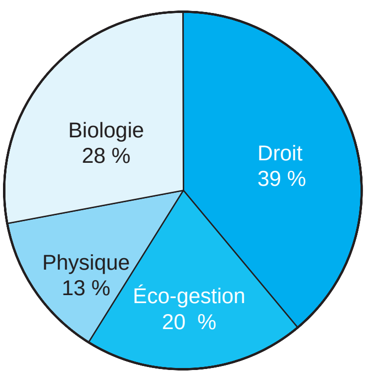
Modalités (x_i)	**Fréquences (f_i)%**	**Angles au centre α_i**
Droit	39 %	$360 \times 0{,}39 =$ **140,4°**
Éco-gestion	20 %	$360 \times 0{,}20 =$ **72°**
Physique	13 %	$360 \times 0{,}13 =$ **46,8°**
Biologie	28 %	$360 \times 0{,}28 =$ **100,8°**
TOTAL	**100 %**	**360°**

Le diagramme est donc le suivant :

Schéma 3.6

Part des disciplines étudiées dans l'université UUU

DIAGRAMME À SECTEURS CIRCULAIRES



Source : secrétariat des inscriptions

La lecture est assez évidente lorsqu'on se reporte à une légende correctement établie. Il est souvent opportun de mettre les intitulés de légende dans les secteurs correspondants.

Résolution par Excel

Rappelons que le lecteur ayant une connaissance suffisante des logiciels pourra aisément se passer d'un certain nombre de détails développés ici. Nous restons volontairement à un niveau « débutant ».

Écrire les quatre intitulés des disciplines enseignées dans une première colonne, et les valeurs d'effectifs (nombres entiers ou décimaux) en face, dans la colonne suivante. **Sélectionner** la plage. Dans la barre des menus, cliquer sur **Insertion**, puis, en dessous, dans **Secteurs**, choisir par exemple *le premier type offert*. Le diagramme avec sa légende apparaît. Choisir le **style** et la **couleur**.

Faire un clic droit sur le diagramme, puis sélectionner **Ajouter les étiquettes de données** : les valeurs numériques apparaissent dans les secteurs. On peut choisir également de modifier les étiquettes de légende, de mettre un titre, etc.

Réponse à la question 3

On introduit donc une dimension supplémentaire – le sexe –, et l'on sait que les diagrammes à secteurs ne sont pas très conviviaux dans le cas des graphiques à deux dimensions (ou plus), car il faut présenter deux (ou plus) **diagrammes à secteurs circulaires côte à côte**, ce qui ne facilite pas la lecture.

Néanmoins, construisons pour l'instant deux diagrammes correspondant aux données ; nous verrons par la suite si on peut améliorer la présentation.

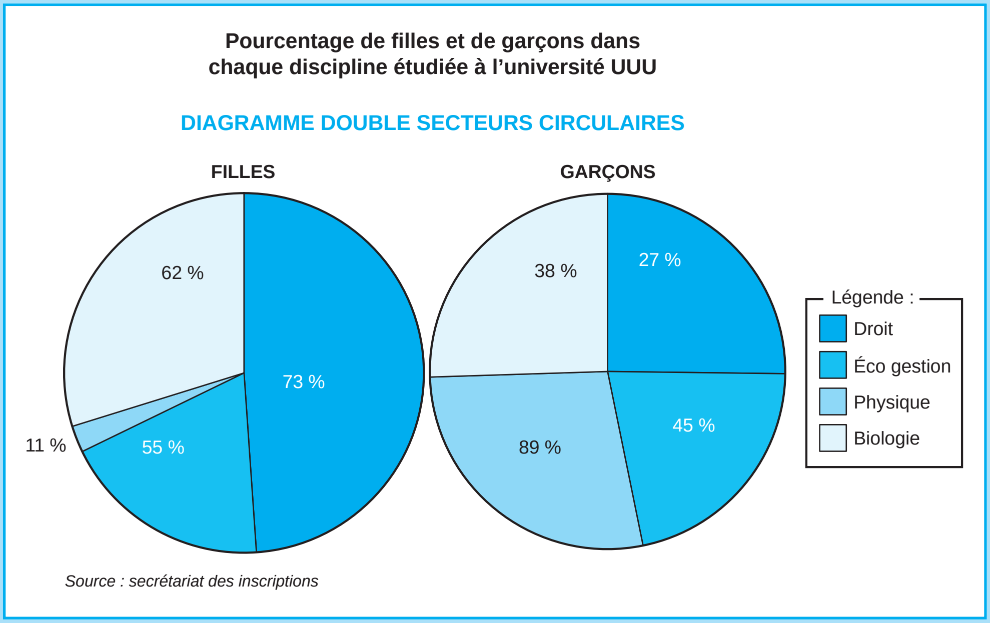
Le tableau à deux dimensions, préalable à la construction, se construit de la manière suivante : puisqu'il y a 5 826 filles, il y a 4 174 garçons. La « colonne total » est obtenue par les fréquences relatives données à la question 1. Les effectifs par sexe se remplissent en proportion des totaux de chaque ligne.

On construit les graphes en se servant directement des effectifs, ou en calculant les pourcentages verticaux. On conserve **la même aire pour chacun des diagrammes**, puisque le message visuel consiste à comparer (en lignes) les parts de chaque sexe, et non pas à montrer qu'il y a plus de filles que de garçons.

Sexe / Discipline	Filles		Garçons		Total	
Droit	2 847	73 %	1 053	27 %	3 900	100 %
Éco-gestion	1 100	55 %	900	45 %	2 000	100 %
Physique	143	11 %	1 157	89 %	1 300	100 %
Biologie	1 736	62 %	1 064	38 %	2 800	100 %

Le diagramme double à secteurs circulaires devient :

Schéma 3.7



Résolution par Excel

La procédure est identique à celle que nous avons vue précédemment. Il suffit de faire deux diagrammes, et de calculer les pourcentages colonnes.

Réponse à la question 4

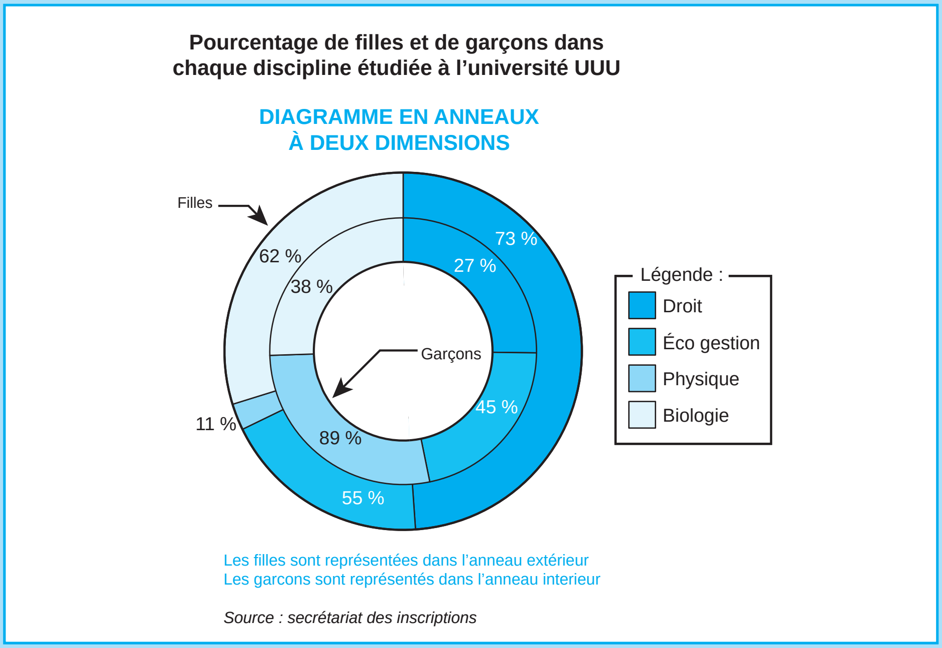
Oui, on peut améliorer l'information tout en conservant la logique angulaire :

- soit en utilisant un **diagramme en anneaux à deux dimensions** (solution 1) ;
- soit en construisant un **diagramme à secteurs semi-circulaires** (solution 2).

La solution 2 est bien meilleure, car elle est largement plus lisible ; en outre, elle fournit un renseignement supplémentaire sur les volumes d'effectifs (ici, le nombre de filles et de garçons), comme nous l'avions souligné à la page 74.

Commençons par la **solution 1** : le graphe suivant montre l'anneau à deux dimensions obtenu sur les données de l'exercice.

Schéma 3.8



Résolution par Excel

Excel ne propose pas directement ni facilement une construction de diagramme en anneaux à deux dimensions[1].

Pour une seule dimension, c'est-à-dire pour *un seul anneau*, la marche à suivre est la suivante : **écrire** les quatre intitulés des disciplines enseignées et, en face, **saisir** les valeurs (nombres entiers ou décimaux) en colonnes, puis **sélectionner la plage**. Dans la barre des menus, cliquer sur **Insertion**, puis, en dessous, sur **Autres graphiques**, et **Anneau**. Le diagramme apparaît avec sa légende. Choisir le **style** et la **couleur**.

Faire un clic droit sur l'un des secteurs de l'anneau, et sélectionner **Ajouter les étiquettes de données** pour obtenir les effectifs dans chaque secteur d'anneau. On peut choisir de mettre un titre, ainsi que différentes options de modification.

1. L'intérêt de ce chapitre n'est pas uniquement de fabriquer des diagrammes mais aussi, tout simplement, de connaître leur existence, de savoir les interpréter. C'est pourquoi nous les présentons.

La solution 2 est bien meilleure, et beaucoup plus lisible, mais la plupart des logiciels ne la proposent pas, ce qui est regrettable. Cela nécessite donc un rapide calcul supplémentaire, celui du **rapport des surfaces**.

Pour trouver *les rayons* des demi-cercles, il faut faire d'abord le *rapport des effectifs* : il y a 5 826 filles et 4 174 garçons, soit un rapport d'effectif de 1,4 qui devra correspondre au rapport des surfaces. Ce rapport des surfaces se convertit en *rapport des rayons* en prenant les racines carrées, soit 1,18. Dès lors, si nous choisissons un rayon de 2 cm pour les garçons, il faudra prendre un rayon de 2,36 cm pour les filles.

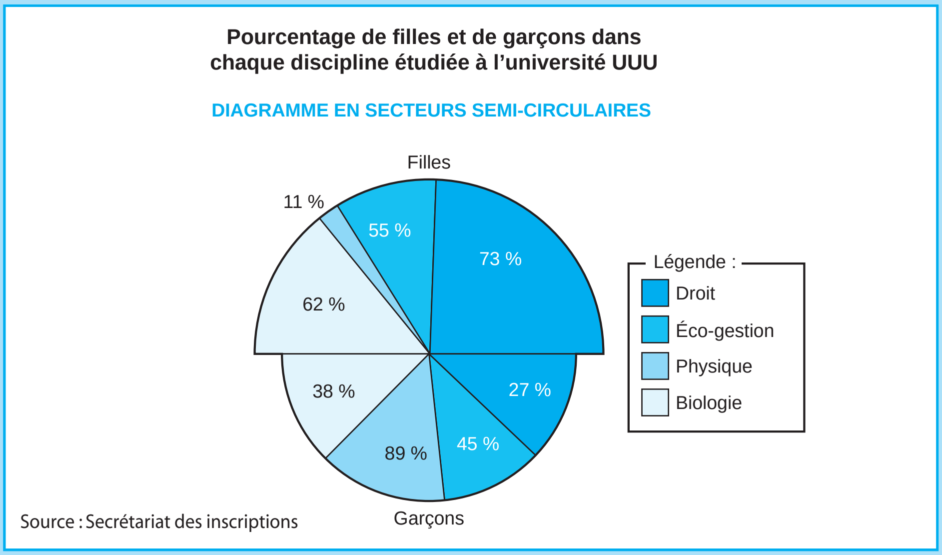
Le principe est de déterminer les angles au centre non plus en multipliant les fréquences par 360° mais tout simplement en **les multipliant par 180°**.

Le tableau de construction devient :

Discipline \ Sexe	Filles		Garçons		Angles au centre	
	Nb	(f_i)	Nb	(f_i)	Filles	Garçons
Droit	2 847	**48,9 %**	1 053	**25,2 %**	180° × 0,489 = **88°**	180° × 0,252 = **45°**
Éco-gestion	1 100	**18,9 %**	900	**21,6 %**	180° × 0,189 = **34°**	180° × 0,216 = **39°**
Physique	143	**2,4 %**	1 157	**27,7 %**	180° × 0,024 = **5°**	180° × 0,277 = **50°**
Biologie	1 736	**29,8 %**	1 064	**25,5 %**	180° × 0,298 = **53°**	180° × 0,255 = **46°**
Total	**5 826**	**100 %**	**4 174**	**100 %**	**180°**	**180°**

En conséquence, le diagramme se présente sous la forme suivante :

Schéma 3.9



Résolution par Excel

Excel, comme bien d'autres logiciels, ne propose ni directement ni facilement la construction de diagrammes semi-circulaires, ce qui est regrettable vu la grande opérationnalité de tels messages visuels.

Exercice 5 : Diagramme à axes multiples : série temporelle en coordonnées polaires (ou toile d'araignée)

Énoncé :

Les précipitations en millimètres relevées mensuellement par le SOSM (Service d'observation statistique météorologique) dans la capitale d'un pays imaginaire sont données pour trois années dans le tableau suivant :

Mois \ Années	2008	2009	2010
Janvier	256	267	234
Février	277	299	226
Mars	167	199	135
Avril	199	234	150
Mai	143	97	65
Juin	64	122	70
Juillet	85	102	85
Août	50	65	42
Septembre	67	86	106
Octobre	123	156	256
Novembre	157	186	244
Décembre	243	287	273

Question 1 Rappelez rapidement les principes de base des diagrammes à 12 axes.

Question 2 Construisez le diagramme en « toile d'araignée » correspondant aux données de l'exercice.

Corrigé de l'exercice 5

Réponse à la question 1

Ce type de diagramme est appelé aussi « en coordonnées polaires » ou « radar » (ici à 12 axes) dans certains logiciels.

C'est un **diagramme de hauteur** dans la mesure où chaque axe est **gradué de façon arithmétique** à partir de zéro. Sa construction à partir de logiciels traditionnels est remarquablement facile.

Il est souvent employé dans le cas de **séries temporelles mensuelles**, bien que la *graduation arithmétique*, contrairement à la *graduation logarithmique*, ne se prête pas toujours bien à des observations que l'on cherche à comparer dans le temps.

Le grand avantage réside dans le fait qu'il est possible de comparer les différents scores, mois par mois, sur les années d'observations : c'est-à-dire de comparer tous les mois de janvier entre eux, tous les mois de février entre eux, et ainsi de suite, révélant la lecture du phénomène de saisonnalité[1].

1. Voir page 241 et suivantes.

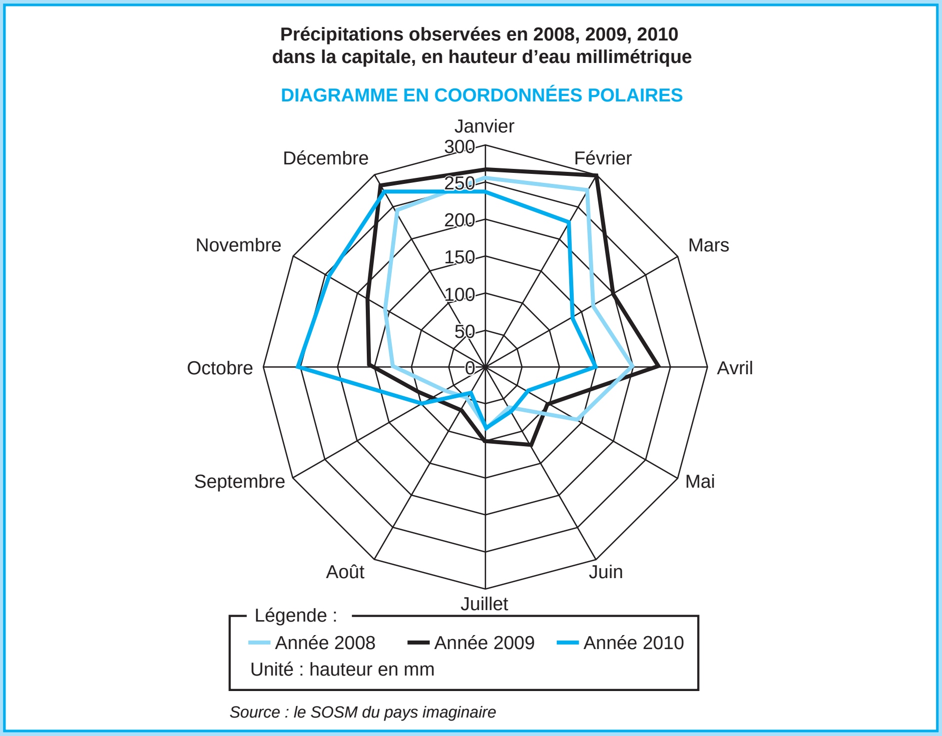
Il s'agit d'une **distribution à deux caractères**, et donc d'un diagramme **à deux dimensions** : en effet, le premier caractère est le « temps » observé en **12 modalités annuelles** ; le second traduit le phénomène étudié : ici les précipitations observées en millimètres.

La lecture, selon l'organisation des données, et si l'on multiplie les années d'observations, n'est pas toujours facile, surtout si les couleurs ne sont pas bien tranchées.

Réponse à la question 2

Le diagramme est le suivant :

Schéma 3.10



Résolution par Excel

Rappelons que le lecteur ayant une connaissance suffisante des logiciels pourra aisément se passer d'un certain nombre de détails développés ici. Nous restons volontairement à un niveau « débutant ».

Saisir les valeurs sur quatre colonnes. Pour la première colonne, saisir **Janvier**, cliquer sur le coin bas à droite, faire apparaître le curseur « + », et faire défiler jusqu'à **Décembre**.

Saisir **Année 2008** en titre de la colonne deux, **saisir** les nombres en vertical, et ainsi de suite jusqu'à la quatrième colonne. **Sélectionner** la plage totale. Dans la barre des menus, cliquer sur **Insertion**, puis, en dessous, sur **Autres graphiques** et **Radar**, puis choisir par exemple *le premier type offert*. Le diagramme complet, avec l'échelle, les étiquettes de mois et d'années, apparaît. Choisir le **style** et la **couleur**.

Faire un clic droit sur le diagramme, si des modifications sont nécessaires.

Exercice 6 : Diagramme en « radar » à quatre axes

Énoncé :

Dans l'exercice n° 4, page 109, questions 3 et 4, nous avions construit des diagrammes circulaires. Est-il judicieux de représenter le même message visuel au moyen d'un diagramme à axes multiples ?

Corrigé de l'exercice 6

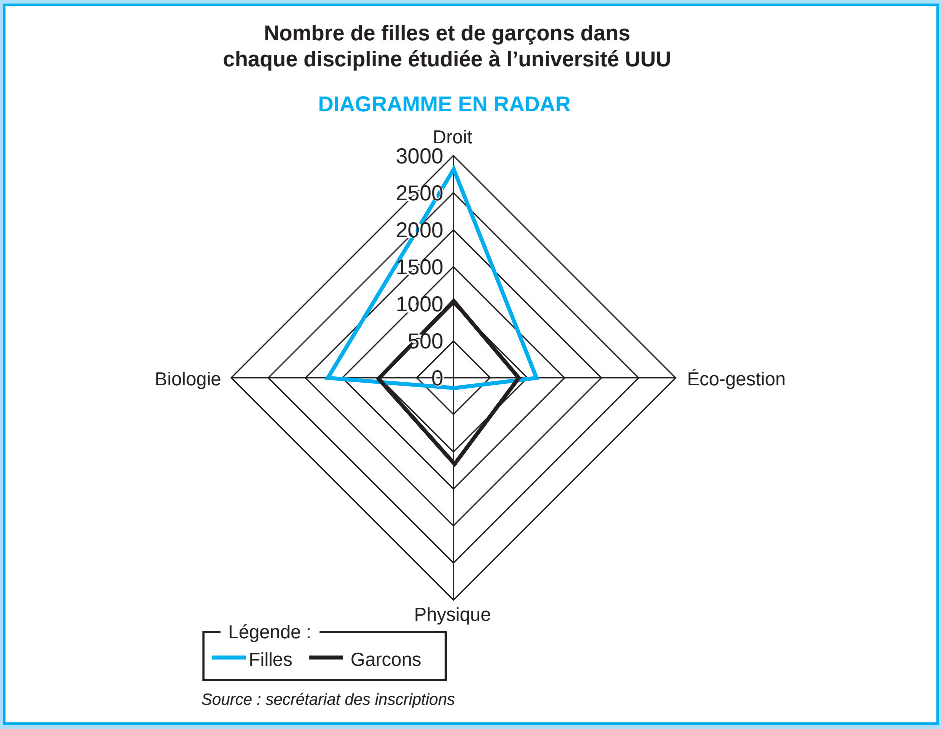
La distribution est **à deux caractères**, puisqu'on croise la répartition des disciplines enseignées et le sexe.

Un *double diagramme circulaire*, comme on l'a vu à la question 3 de l'exercice concerné, n'est pas très convivial, car le regard de l'observateur ne cesse de passer d'un secteur du premier cercle à un secteur du second cercle, ce qui n'est pas pratique et même un peu lassant.

Les autres types de diagrammes, vus à la question 4, sont bien meilleurs au niveau du message visuel, d'autant plus que les données ne sont pas en grand nombre : il n'y a que *quatre modalités au maximum*.

Dès lors, si l'on veut présenter un **diagramme à axes multiples**, on aura uniquement **quatre axes**, ce qui n'est pas excessif pour un tel diagramme qui peut en supporter beaucoup plus.

Schéma 3.11



Un tel diagramme est rapidement construit, même par un logiciel assez « basique », et c'est un avantage ; cependant, la lecture demande une *adaptation à l'échelle* qui n'est pas toujours immédiate.

Dans le cas présent, où l'un des deux caractères est binaire (garçon – fille) et l'autre ne comprend pas trop de modalités (ici quatre), la représentation en secteurs semi-circulaires reste la plus judicieuse.

Résolution par Excel

Rappelons que le lecteur ayant une connaissance suffisante des logiciels pourra aisément se passer d'un certain nombre de détails développés ici. Nous restons volontairement à un niveau « débutant ».

Saisir les valeurs sur trois colonnes. Pour la première colonne, écrire **Droit, Éco-gestion**, **Physique** et **Biologie**. Pour les deux autres, en face, **saisir les données** ainsi que les titres des colonnes : **Filles** et **Garçons**.

Sélectionner la plage totale. Dans la barre des menus, cliquer sur **Insertion**, puis, dans la barre d'outils, sur **Autres graphiques** et **Radar**, ensuite choisir par exemple *le premier type offert*. Le diagramme complet, avec l'échelle, les étiquettes et la légende, apparaît. Choisir le **style** et la **couleur**.

Faire un clic droit sur le diagramme, pour le modifier si besoin.

Exercice 7 : Réflexion sur un graphique semi-logarithmique simple, et construction automatisée

Énoncé :

On veut comparer sur un même graphique l'évolution des rejets de déchets industriels, mesurés dans la même unité, dans la ville « V » et dans la région « R » sur quatre années : 250 et 22 580 pour la première année ; 401 et 36 218 ensuite ; 998 et 90 139 ensuite ; 650 et 58 708 enfin.

Faites un « tableau de travail », c'est-à-dire sans le protocole obligatoire pour la diffusion (titre, source, unités), afin simplement de déterminer pourquoi ce message visuel ne peut se faire sur un graphique traditionnel (ou « arithmétique »).

Dites combien de modules d'un diagramme, dont vous préciserez le nom, il faudra employer pour rendre compte de ces évolutions dans le temps. Vous disposez d'un ordinateur : construisez le graphique à l'aide d'un logiciel ordinaire.

Corrigé de l'exercice 7

Réponses

Le tableau de travail (sans titre), permettant de résumer les données, est le suivant :

Tableau de travail

Années	Ville V	Région R
1	250	22 580
2	401	36 218
3	998	90 139
4	650	58 708

La comparaison graphique de ces séries pose *deux types de problèmes* :

a) Tout d'abord, si l'on voulait reproduire la première série (V) sur un graphique « arithmétique », c'est-à-dire en utilisant sur l'axe vertical une échelle graduée en millimètres, comme on le fait de façon courante, *l'étendue des graduations* resterait dans **l'intervalle des centaines, de 100 à 999**, pour rendre compte des quatre valeurs : 250, 401, 650 et 998.

Jusque-là tout irait bien, mais c'est lorsqu'on essaierait de reproduire l'autre série (R) sur le même graphique que l'on s'apercevrait que c'est *complètement impossible*, les valeurs se situant en effet dans **l'intervalle des 10 000 à 99 999**.

Dès lors, même si l'on fabriquait une échelle verticale graduée en millimètres de 10 mètres de long, le message visuel ne serait pas opérant, montrant une courbe non fluctuante tout en bas et une autre tout en haut, à plusieurs mètres. En outre, les règles mathématiques des repères normés nous *interdisent* de changer la proportionnalité dans les graduations de l'échelle.

Ce premier problème est connu sous le nom de « **saturation des axes** ».

b) Ensuite, pour visualiser correctement les évolutions des deux séries au cours de différentes années, la tendance légitime des observateurs est de repérer des lignes plus ou moins parallèles et de dire alors que les **évolutions temporelles sont elles aussi parallèles**, ce qui est complètement inadéquat sur un repère arithmétique (voir le petit exemple en page 85).

Les comparaisons visuelles en seraient faussées ; ce deuxième problème est connu sous le nom de « **non-proportionnalité des parallèles** ».

Il faut donc utiliser **un graphique semi-logarithmique** (voir les règles de construction et d'utilisation page 86 et suivantes, ainsi que dans les exercices qui suivent) qui permet une étendue de valeurs beaucoup plus large, ainsi qu'une exacte conservation des proportions.

Le *moyen mnémotechnique* pour se souvenir des différences est le suivant :

Échelles arithmétiques : ON ADDITIONNE OU ON SOUSTRAIT	**Échelles logarithmiques :** ON MULTIPLIE OU ON DIVISE
Les échelles arithmétiques traduisent bien les écarts absolus : Si 1 cm vaut 10 €, 2 cm valent 20 €	**Les échelles logarithmiques traduisent bien les variations relatives :** 1/10 = 10/100 = 100/1000 = 10 %
C'est le domaine des parts, en statique.	**C'est le domaine des taux de croissance.**

Il faut donc dans le cas présent, vu les contraintes de saturation des axes et de non-proportionnalité, passer à un message visuel sous forme de **graphique semi-logarithmique**.

On pourra, à cet effet, utiliser une feuille graduée de type « papier semi-logarithmique », ou bien calculer les graduations de l'axe, comme cela est indiqué dans l'exercice suivant, ou encore convertir l'échelle en utilisant l'option que la plupart des logiciels informatiques de base proposent.

Le nombre de modules nécessaires dans cet exemple, vu que la plus petite valeur du phénomène étudié est **250** et que la plus grande est **90 139**, sera de **trois modules** :

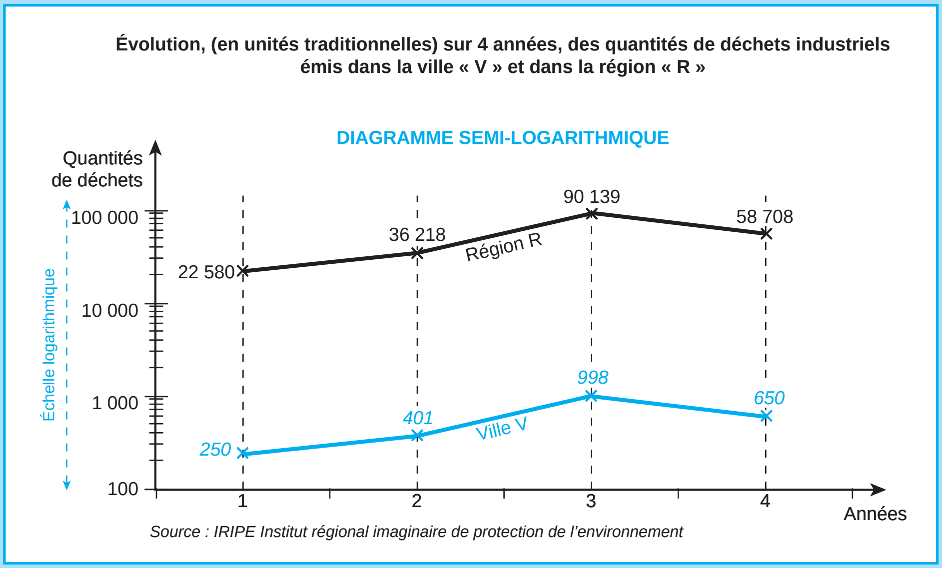
– Le **1** du premier module représentera la valeur **100**.
– Le **1** du deuxième module représentera la valeur **1 000**.
– Le **1** du troisième module représentera la valeur **10 000**.

C'est vers le haut du troisième module que l'on pourra repérer la valeur la plus grande : **90 139**.

Les points obtenus seront reliés par des **segments de droite**, et non par des courbes, puisqu'on ne connaît pas les fluctuations possibles entre les dates de référence.

Les chiffres du tableau seront indiqués en face des graduations correspondantes, de façon qu'on puisse aisément lire les **valeurs exactes du phénomène** sur l'axe vertical.

Schéma 3.12



On voit sur le schéma 3.12 une très nette **variation parallèle des segments** des deux lignes brisées, ce qui signifie que les deux tendances évoluent de la même manière dans le même temps. Quand la ville connaît une augmentation d'émission de ses déchets, la région subit **la même évolution proportionnellement**.

Il n'y a donc pas, en la matière, de *spécificité* de la ville par rapport à l'espace plus grand qu'est la région.

Notons que les plus ou moins fortes pentes des segments de droite indiquent des plus ou moins forts taux de croissance annuels.

Rappelons que le lecteur ayant une connaissance suffisante des logiciels pourra aisément se passer d'un certain nombre de détails développés ici. Nous restons volontairement à un niveau « débutant ».

Saisir ou copier-coller le tableau de travail, avec les titres, sur deux colonnes. **Sélectionner** la plage. Dans la barre des menus, sélectionner **Insertion**, puis, dans la barre d'outils, **Graphique ligne**, puis **Ligne courbe 2D**. Un graphique apparaît *en échelle arithmétique* (on constate d'ailleurs visuellement que le diagramme n'est pas convivial du tout). Faire un c**lic droit sur l'axe vertical**, et choisir **Mise en forme de l'axe**. Dans la boîte de dialogue **Format de l'axe** qui apparaît, cocher *« Option d'axe »* **échelle logarithmique base 10**, et *Graduations principales* et *secondaires* **sur l'axe**.

Fermer la fenêtre : le graphique semi-logarithmique avec sa légende apparaît. On peut choisir également de modifier le style, les couleurs, les étiquettes, de mettre un titre, etc.

Résolution par Excel

Exercice 8 : Construire un diagramme semi-logarithmique

Énoncé :

Vous disposez d'une machine à calculer mais pas d'un ordinateur, et vous n'avez pas sous la main de feuille graduée « en semi-log ». Comment faites-vous pour dessiner sur une feuille ordinaire un repère dont l'axe vertical est une échelle logarithmique ?

Nota : Il s'agit d'une construction sur support papier lorsqu'on ne dispose pas d'un ordinateur. L'exercice ne concerne donc pas un travail sur Excel.

Corrigé de l'exercice 8

Réponses

Il n'est *nullement nécessaire* de savoir ce qu'est un logarithme ou une fonction puissance pour **fabriquer** une échelle logarithmique ou pour **lire** un diagramme en semi-log.

Le terme « semi-log » signifie que **seul l'axe vertical du repère est gradué en logarithmes**, l'axe horizontal restant en graduations arithmétiques.

Les calculettes les plus basiques vendues dans le commerce, mais où l'on peut repérer quelque part sur le clavier les touches $\boxed{lg}$ log décimaux, ou $\boxed{ln}$ log népériens, sont dites *scientifiques*.

On peut y lire par exemple que :

Le log de zéro n'existe pas.

a) Le log *décimal* de **1** est égal à **zéro**.

b) Le log *décimal* de **2** est égal à **0,30103**, que nous pouvons « approcher » par **0,30**.

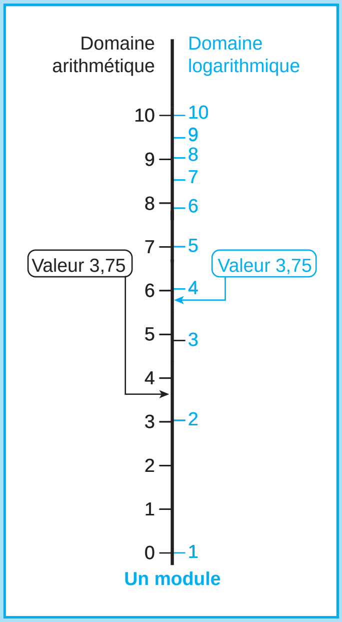
c) Le log *décimal* de **3** est égal à **0,47712**, que nous pouvons « approcher » par **0,48**.

d) Le log *décimal* de **4** est égal à **0,60206**, que nous pouvons « approcher » par **0,60**.

e) Et ainsi de suite jusqu'au log de **10** qui est tout simplement égal à **1**. Ce nombre 10 est alors appelé « **base** » des logarithmes décimaux.

Il suffit de reporter les **chiffres après virgule** de ces logarithmes décimaux jusqu'à 10 sur une échelle arithmétique graduée régulièrement de 0 à 10 pour obtenir *une conversion* en échelle logarithmique, comme ci-dessous (si on n'a pas un logiciel disponible) :

Schéma 3.12 A.



Nous venons de construire **le premier module** d'un graphique semi-logarithmique : il est gradué de 1 à 10. On voit que les **graduations sont de plus en plus resserrées** à mesure qu'elles augmentent. Cela vient du fait que, contrairement à l'échelle arithmétique où la *différence entre deux graduations* est constamment égale à l'unité, en échelle logarithmique, cette différence est *variable*[1].

Si nous passons de 10 à 11, puis à 12, et ainsi de suite jusqu'à 100, nous obtiendrons **exactement les mêmes chiffres après la virgule** que ceux obtenus précédemment[2]. Ainsi, par exemple, log 4 = 0,60206, mais log 40 = 1,60206 ; *les parties fractionnaires se conservent* en base 10.

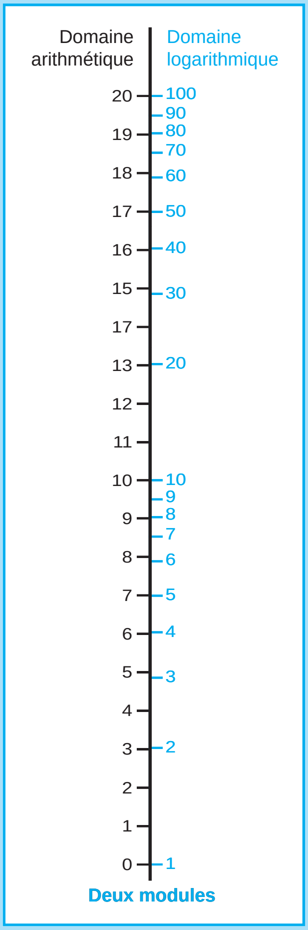
Il en va de même pour toutes les puissances de 10 suivantes : log 400 = 2,60206, et log 4 000 = 3,60206.

Dès lors, nous pouvons **continuer la graduation** de l'échelle logarithmique pour les valeurs allant **de 10 à 99**, en reportant les **chiffres après virgule** exactement comme nous l'avons fait pour graduer le premier module, et obtenir *une échelle à deux modules* ; et ainsi de suite pour le **troisième module**, repérant les graduations de **100 à 999**, voire le

1. C'est une conséquence de la concavité de la fonction « log », et l'on voit facilement par exemple que : log 3 – log 2 est de l'ordre de **0,18**, alors que log 8 – log 7 n'est que de l'ordre de **0,06** ; mais, évidemment, 3 – 2 donne **1**, et 8 – 7 donne toujours **1**.
2. Conséquence de la propriété de la base 10 dans la définition de la fonction « log décimal ».

quatrième module, si l'on a assez de place en vertical dans la feuille, ou si les données de la distribution que l'on étudie l'imposent.

Schéma 3.12 B.



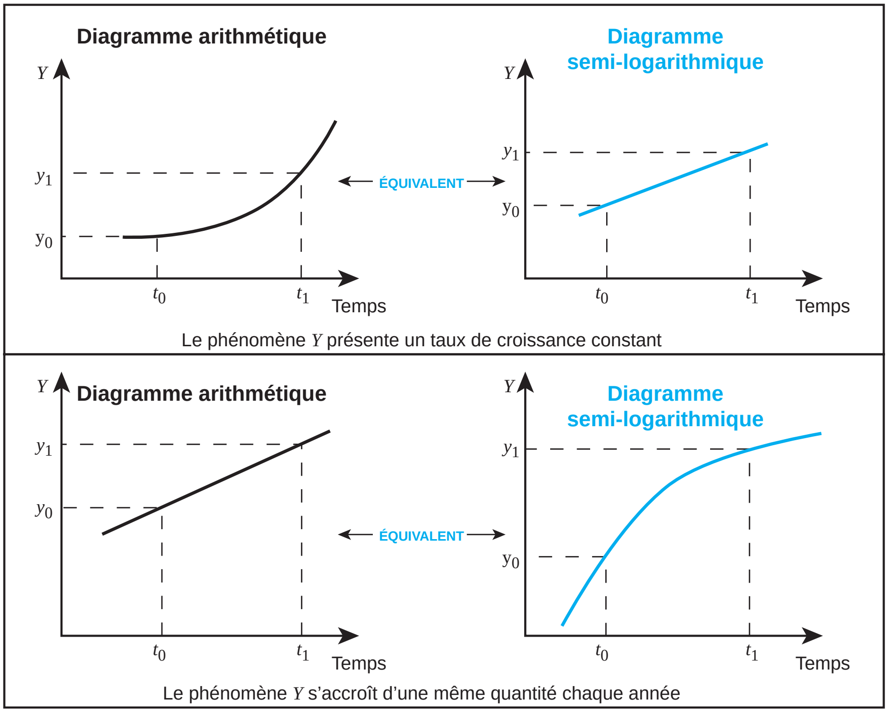
Selon les ordres de grandeur des valeurs à reporter sur l'axe, on pourra choisir par exemple :

- un premier module qui commencera à 100 et qui finira donc à 1 000 ;
- un deuxième module qui indiquera des graduations de 1 000 à 10 000, etc.

Les **valeurs exactes du phénomène étudié seront reportées à leurs places respectives** dans cette nouvelle échelle, facilitant ainsi une *lecture directe des données.*

Notons bien que dans un diagramme semi-log (voir le *moyen mnémotechnique* dans le bas du tableau de l'exercice précédent, page 119), **des rapports égaux à 10 %** par exemple seront **alignés** sur un segment de droite, alors qu'ils auraient été représentés par une **exponentielle** sur un diagramme arithmétique, comme le montre le schéma 3.13, page suivante.

Schéma 3.13



Exercice 9 : Repérer visuellement des évolutions sensiblement parallèles

Énoncé :

Le service officiel de statistique de Boustrasie[1] publie l'évolution sur les six dernières années des taux (en pourcentages) de départ en vacances d'hiver des Boustrasiens, selon la catégorie sociale du chef de ménage, dans le tableau simplifié suivant :

Années \ PCS	Cadres supérieurs	Cadres moyens	Employés	Ouvriers
2005	50 %	39 %	21 %	9 %
2006	69 %	54 %	30 %	13 %
2007	48 %	37 %	21 %	8 %
2008	75 %	58 %	33 %	14 %
2009	56 %	43 %	24 %	10 %
2010	73 %	56 %	31 %	14 %

1. Pays imaginaire.

Pourquoi le type de diagramme « semi-logarithmique » est-il adapté dans ce cas particulier ? Construisez le diagramme sur une feuille graduée en semi-log. Combien de modules allez-vous utiliser ?

Nota : L'exercice porte sur une construction sur support papier gradué et non par Excel.

Corrigé de l'exercice 9

Réponses

On sait que si l'on obtient des droites ou courbes « sensiblement parallèles » sur un diagramme semi-log, c'est donc que les évolutions dans le temps des phénomènes sont « relativement identiques ». C'est dans ce but de comparaison que le type de diagramme semi-log paraît bien adapté.

Traçons les séries « point par point » sur un papier semi-logarithmique à plusieurs modules, comme on l'a fait page 88.

La valeur la plus basse, 8, nous impose de choisir l'échelle du premier module de 1 à 10. La valeur la plus haute ne dépassant évidemment pas 100 %, nous resterons sur un **papier à deux modules**, le second module étant gradué de 10 à 100.

Le graphique apparaît page suivante.

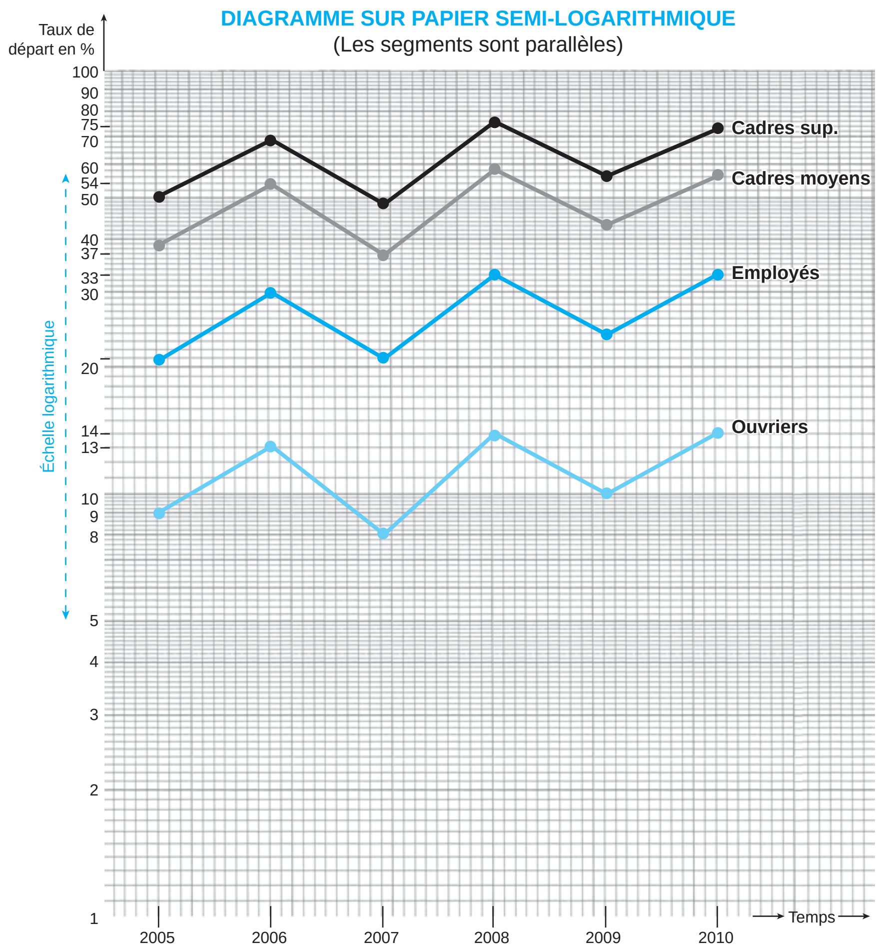
Pour sa construction, on commence par choisir les graduations de **l'axe horizontal**[1] représentant les six années, en optant pour un étalement correct des valeurs afin de garantir la lisibilité. Ensuite, on reporte les valeurs données des pourcentages sur **l'axe vertical**[2], de manière que l'observateur puisse les lire sans se préoccuper des logarithmes.

Il apparaît clairement que **les évolutions sont sensiblement parallèles**, ce qui signifie que les Boustrasiens ont adopté des comportements de départ en vacances d'hiver identiques : quand le taux de départ pour une catégorie baisse, il baisse dans la même proportion pour toutes les autres catégories.

1. Axe qui est gradué selon une échelle arithmétique.
2. Axe gradué selon une échelle logarithmique.

Schéma 3.14

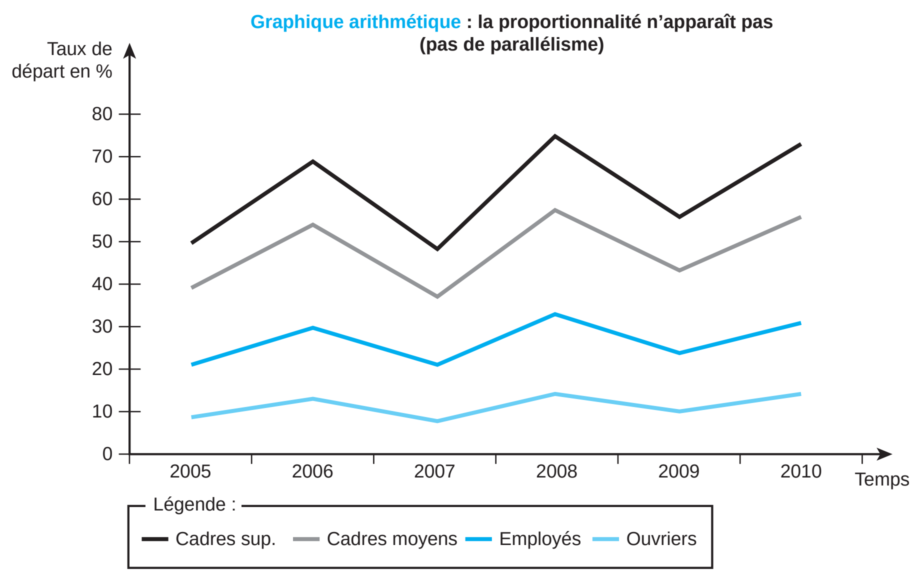
Évolution, de 2005 à 2010, des taux de départ en vacances d'hiver des boustrasiens, selon la catégorie sociale du chef de ménage



Source : SOSB service officiel de statistique de Boustrasie

Un graphique en coordonnées *arithmétiques* n'aurait **pas donné le même message visuel**, comme on peut en juger ci-dessous, surtout en comparant la ligne la plus haute et la ligne la plus basse[1].

Schéma 3.15



Source : SOSB service officiel de statistique de Boustrasie

1. On dirait que le taux de départ en vacances des cadres supérieurs, la dernière année, augmente nettement plus que celui des ouvriers, ce qui est faux.

Exercice 10 : Questions diverses sur les autres diagrammes à traduction indirecte

Énoncé :

a) Un cartogramme est-il un diagramme à une ou deux dimensions ?

b) Un pictogramme est-il un diagramme de hauteur ou de surface ?

c) Quel type d'analyse prend en compte les profils multicritères des phénomènes que l'on étudie ?

d) Sur quelle logique de construction graphique s'appuie l'analyse en composantes principales ACP ?

e) Peut-on fabriquer un diagramme qui combine les deux méthodes « pictogramme » et « cartogramme » ?

Corrigé de l'exercice 10

Réponses

a) C'est un type de diagramme à **deux dimensions** : la première est celle du caractère étudié ; la deuxième dimension concerne *l'espace géographique* de la carte.

b) La rigueur statistique impose que ce soit bien **la surface** et non la hauteur des figurines ou des symboles qui transmette le message visuel. On parle de « **corpulence** » des figurines.

c) Les phénomènes de la vie ne peuvent pas être toujours étudiés selon *un seul critère* (le prix, le poids, la note, etc.). Les **analyses multidimensionnelles ou multicritères** permettent de décrire des « profils » à multiples dimensions (voir page 92 et suivantes). Les méthodes d'analyse employées dans ces cas et qui font apparaître **des facteurs de dépendance ou d'indépendance** (communs ou spécifiques) sont connues sous le nom **d'analyses factorielles**[1].

d) Parmi les analyses factorielles, l'une est l'ACP, **analyse en composantes principales**, qui se situe dans le champ de la statistique mathématique (voir page 94). Sa logique de représentation graphique s'appuie sur la maximisation de certaines distances entre caractéristiques mesurant les dispersions[2], et fait apparaître dans le graphe **des regroupements typiques**, parfois faciles à lire, sur des repères dont les axes ne sont pas donnés *a priori* mais au contraire interprétés par l'analyste.

e) Oui, comme le montre le schéma du CEREQ reproduit page 96, mais leur construction ainsi que leur lecture ne sont pas toujours faciles.

1. Dans ces analyses, on cherche à définir des **plans factoriels** qui permettent de conserver le « **maximum** » d'informations. Cette *maximisation* (ou minimisation) **quantifiée** est à la base de la logique de la plupart de ces méthodes.
2. Variances et covariances.