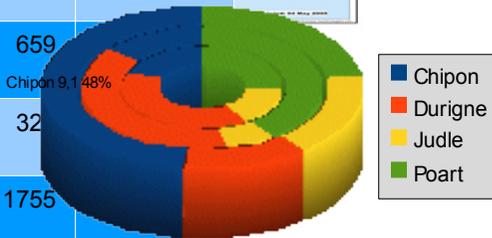


Cod	Nom	Pop00	Hommes 00	Femmes 00	Masculinité %
101 01	CENTRO URBANO	131769	61902	69867	88.599768
101 02	YOTALA	5982	2905	3077	94.410140
101 03	ARABATE	1310	651	659	
101 04	CHAUNACA	665	336	329	
101 05	CHUQUI-CHUQUI	3522	1767	1755	
101 06	HUANIFAYA	1648	824	824	100



- Fiche 1 Bibliographie
- Fiche 2a,b Population, élément et caractères
- Fiche 3 Fréquences
- Fiche 4 Du tableau complet au condensé
- Fiche 5 Indicateurs statistiques de position
- Fiche 6 Indicateurs statistiques de dispersion
- Fiche 7 Discrétisation
- Fiche 8 Discrétisation et cartographie
- Fiche 9a,b,c Discrétisation et distribution
- Fiche 10 La forme de la distribution
- Fiche 11 Le nombre de classes
- Fiche 12a,b,c,d,e,f Méthodes de discrétisation
- Fiche 13a,b,c Discrétisation et logiciels
- Fiche 14 Discrétisation et comparaison
- Fiche 15a,b,c,d Représentations graphiques
- Fiche 16 Liste des formules open office V3

Statistiques et traitement des données

Fredéric Fortunel est maître de conférences à l'Université du Maine dans le laboratoire de recherche Eso-Le Mans (UMR CNRS 6590).

Frederic[point]fortunel[at]univ-lemans[point]fr

<http://eso-gregum.univ-lemans.fr/spip.php?article32>

Enseignements :

Sémiologie graphique et cartographie thématique, SIG, Problématique du développement, Asie du Sud-Est, Enjeux agricoles des Suds, Identités.

Recherches :

Axe 1 - Territoires agricoles : paysanneries, filières, foncier.

Les recherches développées ont pour axe principal les territoires agricoles. A partir de cas sud-est asiatiques (Viêt Nam, Laos, Cambodge), il s'agit de saisir comment s'organisent des bassins productifs orientés vers les cultures commerciales pérennes en soulignant les interactions entre la gestion des ressources –foncier notamment-, les systèmes sociaux –paysanneries- et les logiques économiques –acteurs de la filière d'amont et d'aval.

Des collaborations avec l'Université Royale d'Agriculture (Cambodge), l'Université d'Agroforesterie (Viêt Nam) et l'IHEID (Institut des Hautes études internationales et du Développement –Suisse) sont engagées dans le cadre d'un financement par le Programme de Coopération Scientifique Inter-universitaire (PCSI de l'Agence Universitaire de la Francophonie) : « Les hévéacultures familiales vietnamo-cambodgiennes et leur intégration dans la région du Mékong ».

Axe 2 – Identités : autochtonies et marginalités

En relation avec le premier axe, les recherches envisagent la question identitaire des populations « marginales ». Situés en périphéries spatiale, sociale et politique, des groupes ethniques localisés dans les bassins de production agricole (plateau des Boloven au Laos, Mondulkiri au Cambodge, Binh Phuoc au Viet Nam) se trouvent confrontés à des phénomènes d'exclusion au profit de migrants venus conquérir des terres. La question posée est de savoir quelles stratégies ces groupes minorés et minoritaires mobilisent pour transformer leurs identités.

Bibliographie indicative :

FORTUNEL Frédéric, « From collectivization to poverty: the indigenous people and the State-owned enterprises in Highlands of Vietnam », in BOURDIER Frédéric (eds.), *Development and Dominion, Indigenous peoples of Cambodia, Vietnam and Laos*, Bangkok, White lotus press, 2010, pp. 397-409.

FORTUNEL Frédéric, « 10 ans d'application de la réforme foncière au Viêt Nam », in COLIN Jean-Philippe, LE MEUR Pierre-Yves, LÉONARD Éric (eds.), *Les politiques d'enregistrement des droits fonciers, du cadre légal aux pratiques locales*, IRD-Karthala, Paris, 2010, pp. 477-496.

FORTUNEL Frédéric, « Les plateaux méridionaux d'Asie du Sud-Est continentale, de la marginalité à l'interconnexion », *Péninsule*, n°57, 2008, pp. 19-41.

FORTUNEL Frédéric, « Trajectoires foncières de minorités ethniques au Viêt Nam. Collectivisation, réformes et dénationalisation des terres », *Études rurales*, n° 181, 2008, pp. 103-114.

FORTUNEL Frédéric, « Le plateau des Boloven et la culture du café, entre division interne et intégration régionale », *L'espace géographique*, n°3, 2007, pp. 215-228.

DE KONINCK Rodolphe, DURAND Frédéric, FORTUNEL Frédéric (dir.), *Agriculture, environnement et sociétés sur les Hautes terres du Viêt Nam*, IRASEC-Arkuiris, Bangkok-Toulouse, 2005, 224 p.

FORTUNEL Frédéric, *Le café au Viêt Nam, de la colonisation à l'essor contemporain d'un grand producteur mondial*, Harmattan, coll. points sur l'Asie, 2000, 172 p.

FICHE 1 BIBLIOGRAPHIE

BEGUIN M., PUMAIN D. *La représentation des données géographiques : Statistique et cartographie*.
Collection Cursus, Édition Armand Colin, Paris. 192p. (Deuxième édition 2000)

LAHOUSSE Ph., PIEDANNA V., 1998, *L'outil statistique en géographie, Tome I, Les distributions à une dimension*, Série " Synthèse Géographie ", Armand Colin, Paris, 96p.

LAHOUSSE Ph., PIEDANNA V., 1999, *L'outil statistique en géographie, Tome II : L'analyse bivariée*, Série "Synthèse Géographie ", Armand Colin, Paris, 96p.

SANDERS L., 1989, *L'analyse des données appliquées à la géographie*, Montpellier, RECLUS, Coll. « Alidade ».

Ces fiches ont été réalisées à partir de plusieurs sources dont :

EADATER, *Représentation cartographique*, IGN, Paris, 2001.

MINVIELLE E, SOUIAH Sid-A., *L'analyse statistique et spatiale*, Éditions du temps, Paris, 2003.

DUMOLARD P., DUBUS N., CHARLEUX L., *Les statistiques en géographie*, Belin, coll. Atouts, Paris, 2003.

SOURIS M., DEMORAES F., *Formation SIG et santé*, IRD.

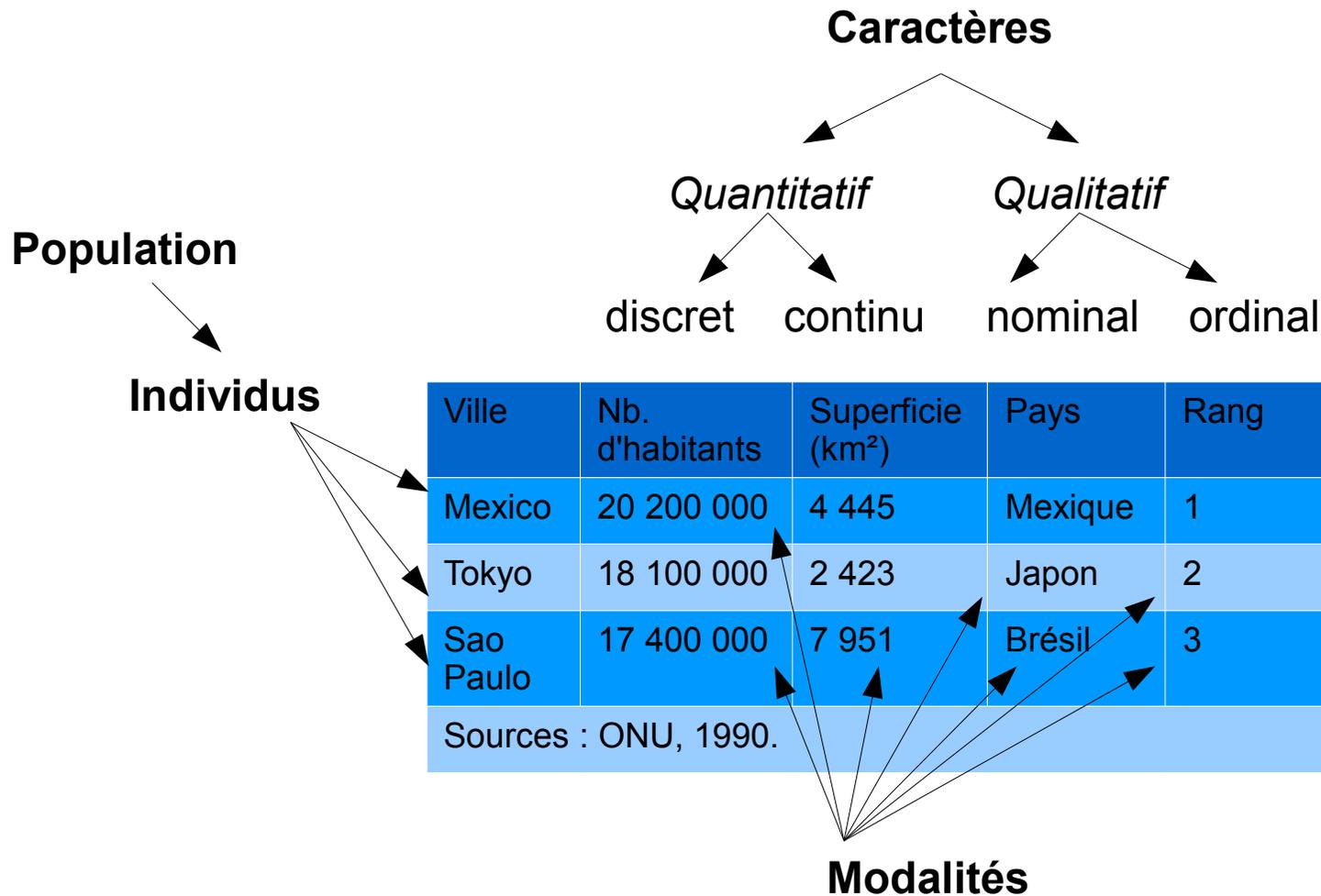
MATHIAN H., *Méthodes de discrétisation*, École d'été « statistiques cartographies et analyse spatiale », Yaoundé 2006.

FICHE 2a POPULATION, ÉLÉMENT et CARACTÈRES

Les données observées sont traduites par un vocabulaire spécifique :

- Une **population** est l'ensemble des éléments observés.
Il s'agit d'une notion "collective". Ex: des données sur un groupe humain, sur les jours de l'année, sur les accidents de la route, sur des lieux.
- Un **individu, élément** ou unité statistique.
Il s'agit de tout élément d'une population, la notion est élémentaire. Le nombre d'individu que compte une population est nommé **effectif total** de cette population : n .
- Le **caractère** est une propriété commune à tous les individus d'une population observée d'une manière homogène. Chaque caractère présente des modalités sous forme de valeur (Q_x) ou de catégories (Q_l).
 - *Caractère quantitatif* : traduit une quantité issue de mesures ou de dénombrement.
 - quantitatif discret : seulement quelques valeurs isolées dans l'intervalle de variation (dénombrement).
 - quantitatif continu : peut prendre toutes les valeurs à l'intérieur de l'intervalle de variation (mesure).
 - *Caractère qualitatif* : traduit une qualité. Essentiellement une catégorisation.
 - qualitatif ordinal : organisées selon un ordre hiérarchique.
 - qualitatif nominal : organisées selon un ordre non hiérarchique.

FICHE 2b POPULATION, ÉLÉMENT et CARACTÈRES



Toute *population* est un ensemble d'*individus*.
Chaque *caractère* est un ensemble de *modalités*.

A chaque *individu* est associé une *modalité* et une seule.

Tout *individu* de la *population* présente l'une des *modalités* du *caractère*.

Les *modalités* forment une partition du *caractère*.

Une **variable statistique** est un *caractère* dont les *modalités* sont connues, *individu par individu*.

FICHE 3 INDICATEURS DE CONCENTRATION : LES FRÉQUENCES

Les fréquences, simples ou cumulées, servent à cerner le poids respectif de chaque modalité dans une population.

De manière essentielle une fréquence est un effectif relatif.

Elle se présente sous la forme de nombres compris entre 0 et 1.

Une fréquence peut s'exprimer sous forme de pourcentage mais attention tout pourcentage n'est pas une fréquence. Ex: le taux de croissance d'une commune entre deux recensements s'exprime en % mais n'est pas un effectif relatif car inférieur à 0 et supérieur à 1.

$$f_j = n_j / n$$

f_j est la fréquence simple, n_j et l'effectif partiel, n est l'effectif total

Le calcul des fréquences simples est utile si plusieurs individus partagent au moins à une reprise la même modalité de caractère ou la même classe de valeurs.

Les fréquences cumulées permettent de traduire l'évolution des fréquences d'une série.

Elles sont ascendantes ou descendantes et se calculent, une fois les modalités ordonnées, par addition successive des fréquences simples.

Voir fiche suivante.

FICHE 4 DU TABLEAU COMPLET AU CONDENSÉ

Qualitatif nominal

A chaque modalité est affecté le nombre d'individu

Population	Couleur	Couleur	Nj	Fj
Paul	bleu	bleu	2	0,6
Pierre	gris	gris	1	0,3
Jeanne	bleu			

Tableau complet

Couleur	Nj	Fj
bleu	2	0,6
gris	1	0,3

Tableau condensé

Quantitatif discret

A chaque valeur du caractère on affecte l'effectif partiel.

Foyers	Nb de véhicule	Nb de véhicule /foyer	Nj	Fj	Fj croiss	Fj décroiss
M1	2	0	467	0,46	0,46	1
M2	0	1	243	0,24	0,68	0,53
.../..		2	250	0,25	0,93	0,29
Mx	3	3	40	0,04	1	0,04
Mx'	0					

Qualitatif ordinal

On ordonne les modalités du caractère. Il est possible de calculer la F croiss. et décroiss.

Population	Mention	Mention	Nj	Fj	Fj croiss	Fj décroiss
Paul	TB	TB	1	0,2	1	0,2
Pierre	P	B	1	0,2	0,8	0,4
Jeanne	P	P	3	0,6	0,6	1
Hélène	B					
Audrey	P					

Ce tableau doit être lu ainsi :

Fj croissante : 80% des élèves ont une note inférieure ou égale à la mention Bien.

Fj décroissante : 40 % des élèves ont une note supérieure à la mention Bien.

Quantitatif continu

Il est nécessaire de procéder à des classes de valeurs. Il s'agit de la discrétisation (*voir fiche spécifique*).

FICHE 5 INDICATEURS STATISTIQUES DE POSITION

LA MOYENNE - M

La moyenne est le centre de gravité d'une distribution (barycentre). De ce fait, elle est très sensible aux valeurs extrêmes.

En cas de distribution dissymétrique, elle décrit mal la population.

La somme de toutes les valeurs divisée par le nombre d'individus dans la série.

$$M_x = (X_1 + X_2 + X \dots + X_n) / n \quad \text{OU} \quad = \text{MOYENNE}(\text{MATRICE})$$

LA MEDIANE - Q₂

La médiane est la valeur de la série qui partage la distribution en deux sous-ensembles d'égale effectifs c.a.d qu'une moitié de la série est inférieure à la médiane et l'autre moitié est supérieure à la médiane.

Une médiane peut se calculer dans une série continue ou discrète (à partir d'une série en classes).

Une médiane est déterminée par le classement des valeurs et non les valeurs elles-mêmes.

Il faut avant toute chose classer la série par ordre croissant.

$$Q_2 = (n+1) / 2 \quad \text{OU} \quad = \text{MEDIANE}(\text{MATRICE})$$

si n impair la médiane est présente dans la série, si n pair, la médiane correspond à la moyenne des 2 valeurs qui encadrent ce rang.

En sciences sociales, il est important de faire la distinction entre moyenne et médiane. Par exemple le salaire médian désigne le palier divisant l'ensemble des salariés en deux parties égales : 50% des salariés gagnent moins que le salaire médian et l'autre moitié gagne plus. Comme les hauts salaires s'éloignent plus de la médiane que les bas salaires, le salaire moyen est supérieur au salaire médian.

FICHE 6 INDICATEURS STATISTIQUES DE DISPERSION

Contrairement à l'étendue et aux quartiles, la variance permet de combiner toutes les valeurs à l'intérieur d'un ensemble de données afin d'obtenir la mesure de dispersion. La variance (symbolisée par S^2) et l'écart type (la racine carrée de la variance, symbolisée par S) sont les mesures de dispersion les plus couramment utilisées.

VARIANCE

En statistique et probabilité, la variance est une mesure arbitraire servant à caractériser la dispersion d'une distribution ou d'un échantillon.

La variance est considérée comme la moyenne des carrés des écarts à la moyenne et caractérise la dispersion des valeurs par rapport à la moyenne.

Le fait que l'on prenne le carré de ces écarts à la moyenne évite que des écarts positifs et négatifs ne s'annulent.

$$\text{Var}(X) = \sigma^2 \quad \text{ou} \quad =\text{VAR}(\text{MATRICE})$$

La variance est toujours positive ou nulle.

Lorsque la variance est nulle, cela signifie que la variable aléatoire correspond à une constante (toutes les réalisations sont identiques).

Le fait de déplacer simplement une distribution (ajouter +b) ne modifie pas sa variance. Par contre, changer l'échelle (multiplier par a) modifie la variance. Cette propriété confirme le fait que la variance d'une constante est nulle.

COVARIANCE

$$\text{COV}(X,Y) = \sigma(Xy) \quad \text{OU} \quad =\text{COVARIANCE}(\text{matrice1};\text{matrice2})$$

En statistiques, la covariance est un nombre permettant d'évaluer le sens de variation de deux variables et, par là, de qualifier l'indépendance de ces variables.

Deux variables ayant une covariance non nulle sont dites dépendantes : par exemple, dans une population donnée, le poids et la taille sont des variables dépendantes. Cependant, elles ne sont pas corrélées : la corrélation est une relation linéaire, or le poids ne varie généralement pas proportionnellement à la taille.

La connaissance des covariances est le plus souvent indispensable dans les fonctions d'estimation, de filtrage et de lissage.

En sociolinguistique, la covariance désigne la correspondance entre l'appartenance à une certaine classe sociale et un certain parler inhérent à cette condition sociale.

FICHE 6b INDICATEURS STATISTIQUES DE DISPERSION

ECART-TYPE - σ (signe *sigma* minuscule)

L'écart-type est la racine carrée de la variance.

Il est un indicateur de la dispersion des valeurs autour de la moyenne d'une valeur quantitative.

En bref, il est l'écart moyen des valeurs à leur moyenne arithmétique.

L'écart type est la mesure de dispersion, ou étalement, la plus couramment utilisée en statistique lorsqu'on emploie la moyenne pour calculer une tendance centrale. Il mesure donc la dispersion autour de la moyenne. En raison de ses liens étroits avec la moyenne, l'écart type peut être grandement influencé si cette dernière donne une mauvaise mesure de tendance centrale.

L'unité de mesure de l'écart-type est la même que celle utilisée pour les valeurs de la série statistique. il est toujours positif.

Dans Excel : **ECARTYPE.P (MATRICE)** pour la population entière, **ECARTYPE (MATRICE)** sur un échantillon.

L'écart-type sert à mesurer la dispersion d'un ensemble de données, par exemple la répartition des notes d'une classe. Dans ce cas, plus *l'écart-type est faible, plus la classe est homogène.*

À l'inverse, on peut souhaiter avoir un écart type le plus large possible pour éviter que les notes soient trop resserrées (exemple classique du professeur qui note de 8 à 13). Dans le cas d'une notation de 0 à 20, l'écart type minimum est 0 (si tous les élèves/étudiants ont la même note), et jusqu'à environ 10 si la moitié à 0/20 et l'autre moitié 20/20.

FICHE 7 DISCRÉTISATION

Comme son nom l'indique la discrétisation est une procédure destinée à rendre un caractère continu en caractère discret... explications.

La discrétisation est avant tout une simplification de caractères de manière à obtenir d'un tableau complet un tableau condensé. Les valeurs appartenant à une même classe sont alors considérées comme identiques.

$$[b_{\text{inf}} - b_{\text{sup}} [$$

b est la borne b_{inf} est la borne inférieure, b_{sup} est la borne supérieure

Constituer des classes, au delà des méthodes de discrétisation, implique de faire deux choix :

- Observation de la distribution de la série → prochaine fiche.
Selon la forme que prend une série, il faudra en déduire une/des méthodes de discrétisation adaptée.
- Choix du nombre de classes → fiche suivante.
Le choix de nombre de classes est conditionné par les possibilités de la *représentation cartographique* (au delà de 6 classes, la distinction visuelle est délicate), *l'effectif total* de la population étudiée et de *l'allure de la variable*.
- Choix des bornes, de l'amplitude des classes et du type de discrétisation → fiches suivantes.
Ce choix ne peut être fait de manière arbitraire. Il doit révéler d'une signification particulière quant au domaine d'étude.

FICHE 8 DISCRÉTISATION et CARTOGRAPHIE

Cartographier une série quantitative nécessite de discrétiser les valeurs... c.a.d faire une carte exige souvent de découper une série statistique en classes.

Contraintes logiques :

Les classes doivent couvrir l'ensemble de la distribution et doivent être contiguës (jointives).
Les classes ne doivent pas être vides.
Implique un degré de généralisation.

Contraintes techniques :

Liées à la méthode de discrétisation (ex: nb pair et impair de classes).

Contraintes visuelles :

L'enjeu est de faire un *compromis* entre la fidélité aux données numérique et une représentation visuelle satisfaisante.

→ afin de résumer au mieux la distribution (conserver l'hétérogénéité)

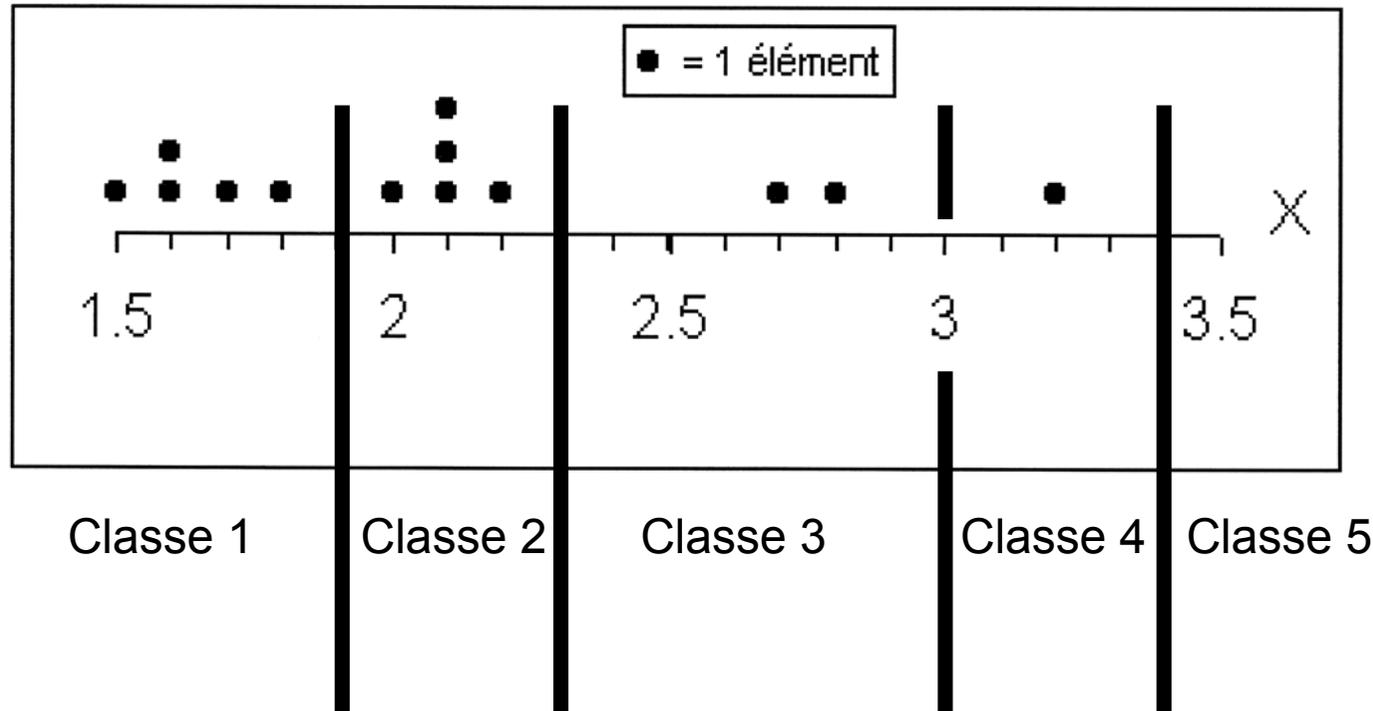
→ construire une carte efficace.

Le compromis se situe entre un grand nombre de classe afin de répondre au critère 1 et un petit nombre de classe pour répondre au critère 2.

plus la distribution est respectée plus le nombre de variables visuelles sera important, plus la carte sera illisible.

FICHE 9a DISCRÉTISATION et DISTRIBUTION

Le scalogramme -ou matrice ordonnée- est l'outil premier pour déterminer une classe.



Marche à suivre :

- Ranger les données par ordre croissant (ou décroissant) avec le scalogramme
- Choisir le nombre de classes et leurs bornes
- Construire le tableau condensé avec
 - 1ere colonne: les classes
 - 2eme colonne : les effectifs
 - 3eme colonne : fréquences simples
 - 4eme colonne : fréquences croiss. et décroiss.

FICHE 9b DISCRÉTISATION et DISTRIBUTION

Partition en classes, deux règles à respecter :

Exhaustivité : la réunion de l'ensemble des classes doit recouvrir au moins l'ensemble du domaine de variation du caractère (du minimum au maximum observé) mais qui peut être plus large (minimum et maximum possibles).

Disjonction : les classes sont deux à deux disjointes (leur intersection est nulle).
[X-X2 [

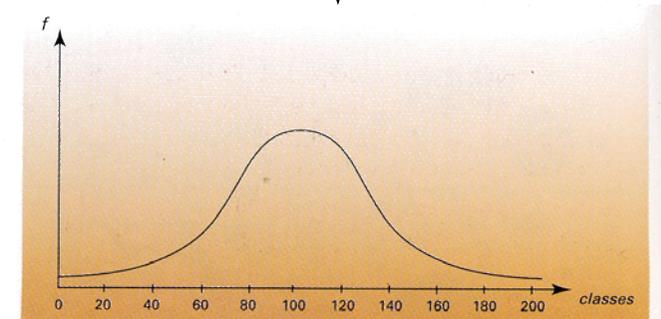
Exercices : la production de café en 2007 ; les eaux minérales.

FICHE 10 LA FORME DE LA DISTRIBUTION

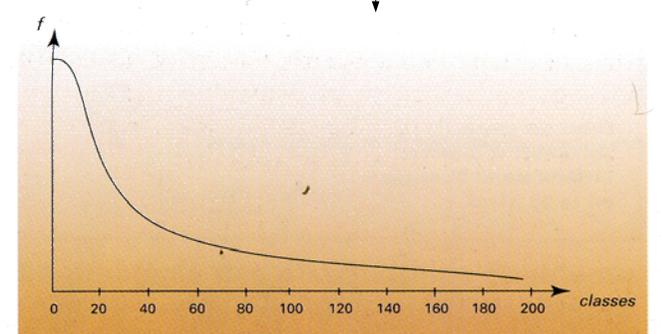
- **Distribution uniforme** : Les données sont réparties uniformément sur toute l'amplitude de la série. La méthode des seuils d'égale étendue est la mieux adaptée à ce type de distribution. Ce type de distribution est peu courant.
- **Distribution symétrique** : Le diagramme de distribution a l'allure exacte d'une courbe de Gauss (en "cloche") avec une concentration des données autour de la moyenne. Il faudra dans ce cas retenir un mode de discrétisation qui favorise la représentation des valeurs moyennes. La méthode des classes standardisées selon moyenne et écart type convient le mieux à ce type de distribution.
- **Distribution dissymétrique** : Il y a concentration des données sur les valeurs faibles. Il faudra donc choisir un type de discrétisation favorisant la représentation des valeurs faibles ; c'est le cas de la méthode des classes selon progression géométrique.
- **Distribution dissymétrique avec plusieurs "pics"** : Sous peine d'obtenir une carte vide d'information, il faudra recourir à la méthode des classes d'effectifs égaux (quantiles), ou bien choisir la méthode des seuils observés. Celle-ci a l'avantage de permettre de déterminer les limites de classes visuellement, en s'appuyant sur les ruptures ou discontinuités de la série. La méthode des quantiles est également une méthode qui convient à ce type de distribution très fréquent.

ATTENTION : Lorsque l'on veut mettre des cartes en relation et se livrer à des comparaisons il est impératif de retenir la méthode des classes d'effectifs égaux.

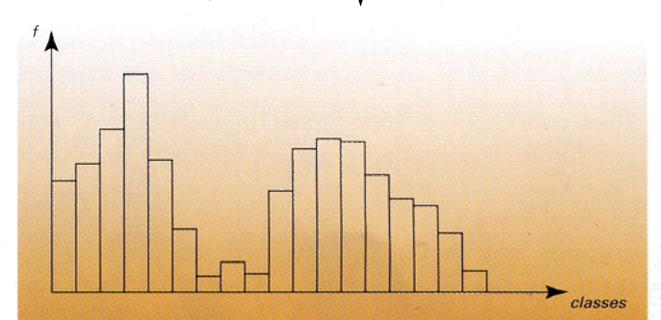
Distribution symétrique↓



Distribution asymétrique↓



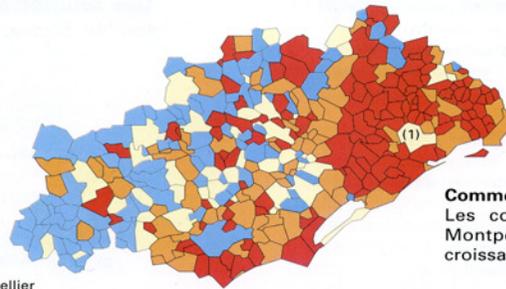
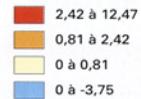
Distribution plurimodale↓



FICHE 11 LE NOMBRE DE CLASSES

4 classes : des regroupements importants.

Taux d'évolution annuelle
de la population (1982-1990)
en %

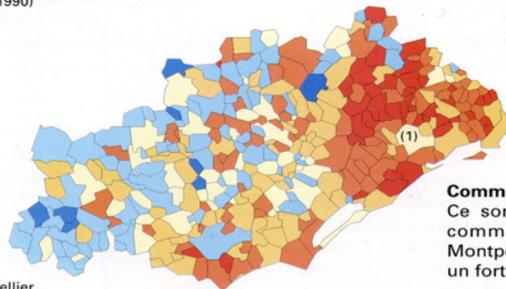
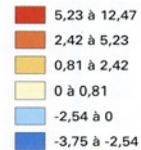


(1) commune de Montpellier

Commentaire possible :
Les communes autour de
Montpellier sont en forte
croissance.

6 classes : une discrétisation plus adaptée.

Taux d'évolution annuelle
de la population (1982-1990)
en %

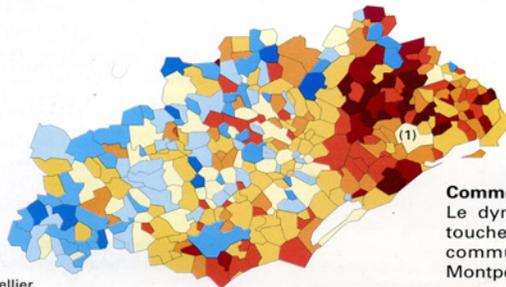
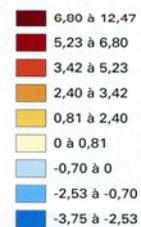


(1) commune de Montpellier

Commentaire possible :
Ce sont principalement les
communes autour de
Montpellier qui connaissent
un fort dynamisme.

9 classes : affinement des structures spatiales.
(solution mathématique d'après Huntsberger)

Taux d'évolution annuelle
de la population (1982-1990)
en %



(1) commune de Montpellier

Commentaire possible :
Le dynamisme le plus fort
touche essentiellement les
communes au Nord de
Montpellier.

FICHE 12a METHODES DE DISCRETISATION

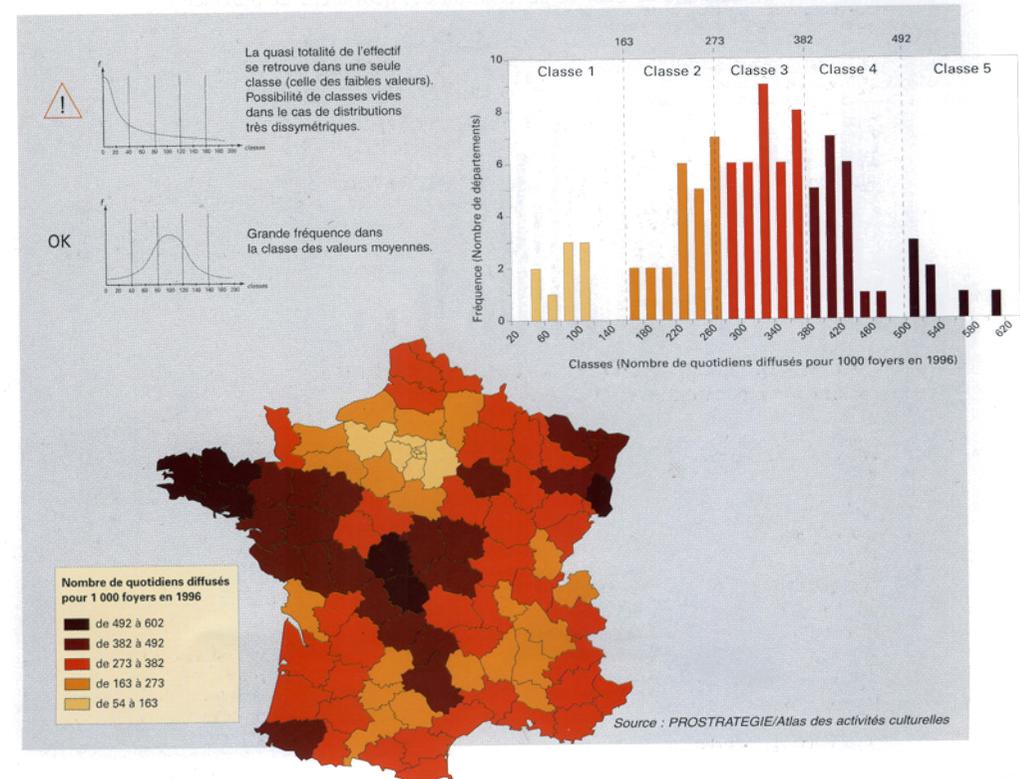
A) Par équivalence ou amplitudes égales

Réalisation de classes de même amplitude. Pour cela, il faut calculer l'étendue de la série :
Étendue = Max – Min.

On divise ensuite l'étendue par n classes afin d'obtenir l'amplitude de chaque classe $a = (\text{max}-\text{Min}) / k$
Si les valeurs extrêmes s'écartent de la série, cette méthode est peu pertinente.



La discrétisation : classes d'égale amplitude,
exemple du taux de pénétration des quotidiens de province



FICHE 12b METHODES DE DISCRETISATION

B) Les quantiles ou effectifs égaux

Cette méthode implique une répartition équilibrée des individus.

Effectif d'une classe = N/k

N est l'effectif total, k est le nombre de classes.

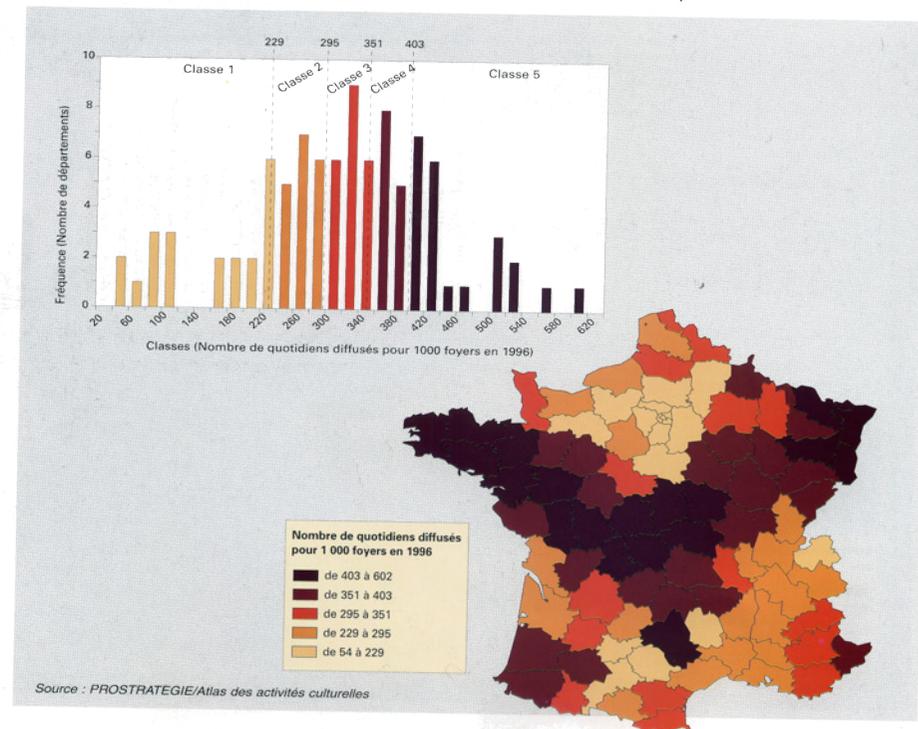
Pour $k = 4$, on obtient des quartiles comme limites de classes.

S'applique à toutes les distributions statistiques. Si cette méthode assure une bonne comparaison entre séries statistiques différentes, en revanche elle gomme les amplitudes inégales : on perd toutes les informations sur la distribution de la série.



25%	25%	25%	25%
cl1	cl2	cl3	cl4

La discrétisation : classes d'égal effectif,
exemple du taux de pénétration des quotidiens de province



FICHE 12c METHODES DE DISCRÉTISATION

C) Les quartiles

6 classes uniquement avec isolement des deux classes extrêmes en séparant les 5% des valeurs les plus petites et les 5 % des plus grandes.



5%	22,5%	22,5%	22,5%	22,5%	5%
cl1	cl2	cl3	cl4	cl5	cl6

D) La progression arithmétique

Étendue des classes augmente selon une progression arithmétique : la 2^o classe a une étendue double de la 1^o, et la 3^o classe cumule l'étendue de la 1^o et de la 2^o...

L'intérêt est de mieux étaler la répartition dans les faibles valeurs plus finement représentées, par contre les fortes valeurs se retrouvent regroupées dans la dernière classe. Les classes sont donc d'amplitude croissante.

A utiliser avec précaution car ne s'applique qu'à des formes précises de distribution dont essentiellement les séries "dissymétriques" (maximum de données vers les faibles valeurs) avec "asymétrie vers la gauche".

E) La progression géométrique

Classes découpées selon une progression géométrique

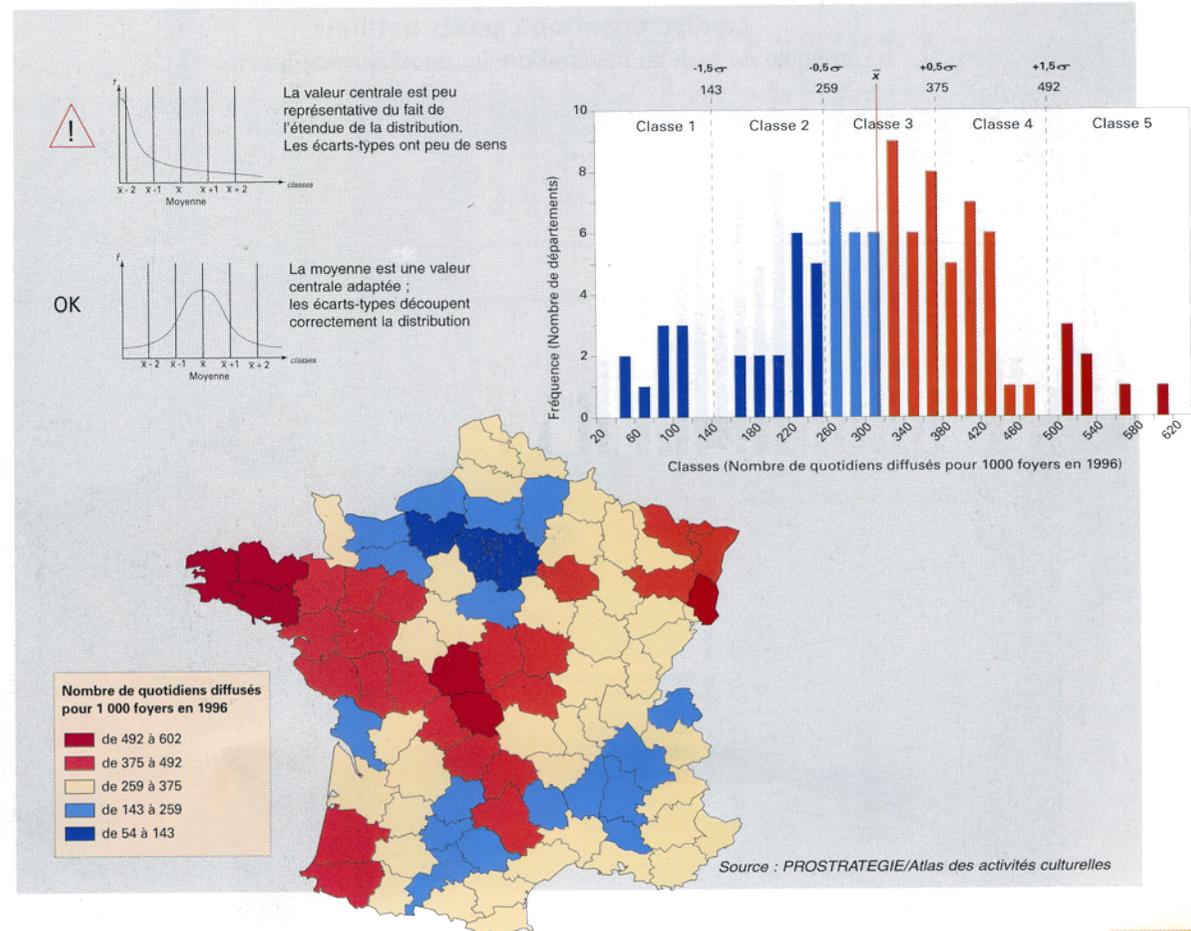
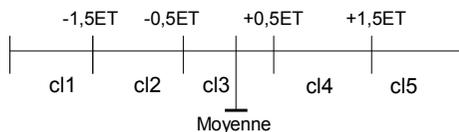
FICHE 12d METHODES DE DISCRETISATION

F) déviation standard ou Écart-type



Chaque classe est déterminée selon une fraction d'écart-type par rapport à la moyenne. Cette méthode présente un avantage majeur : elle permet es comparaison indépendamment des problèmes liés à la taille de la variable puisque cela équivaut à avoir une origine commune (la moyenne) et une unité de mesure identique (écart-type).

La discrétisation : classes selon l'écart-type (standard),
exemple du taux de pénétration des quotidiens de province



FICHE 12e METHODES DE DISCRETISATION

G) Méthode des « seuils naturels »

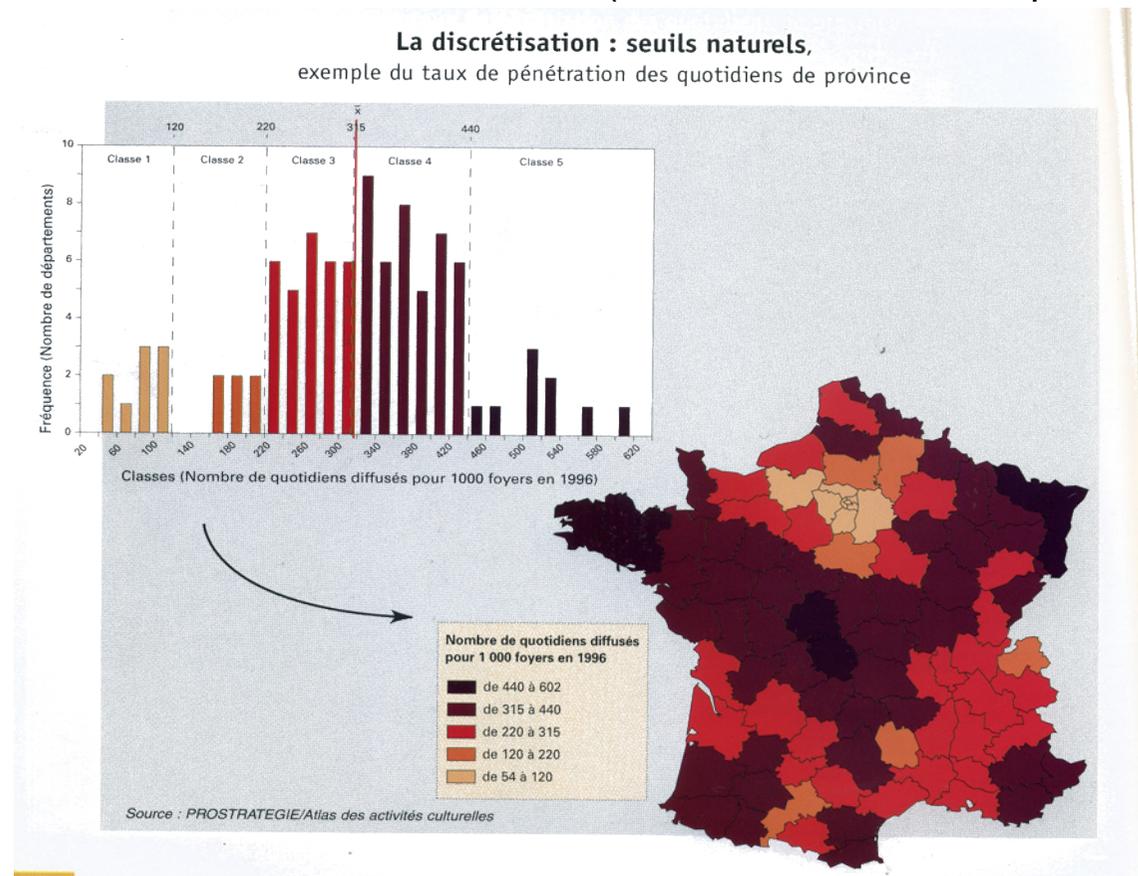


Cette méthode est liée à l'observation d'un diagramme en bâton, d'un histogramme, d'une courbe de fréquences ou de fréquences cumulées

Elle permet de tenir compte des discontinuités observables mais elle n'est acceptable et intéressante que si les discontinuités existent réellement.

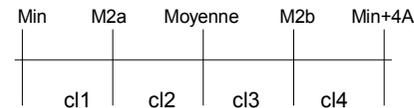
Plutôt que de seuils naturels il serait préférable de parler de seuils observables.

Cette méthode peut venir en complément d'un autre mode de discrétisation (mais attention aux comparaisons).



FICHE 12f METHODES DE DISCRÉTISATION

H) Les moyennes emboîtées



Cette méthode repose sur la notion de moyenne considérée comme le centre de gravité de la variable. La moyenne divise la série en deux groupes afin de construire deux classes. A leur tour, les moyennes des deux sous-groupes permettent un nouveau découpage en 4 classes et ainsi de suite. Cette méthode est un compromis entre la méthode de progression géométrique (privilégie les valeurs élevées) et la méthode standard. Chaque borne de classes exprime une moyenne.

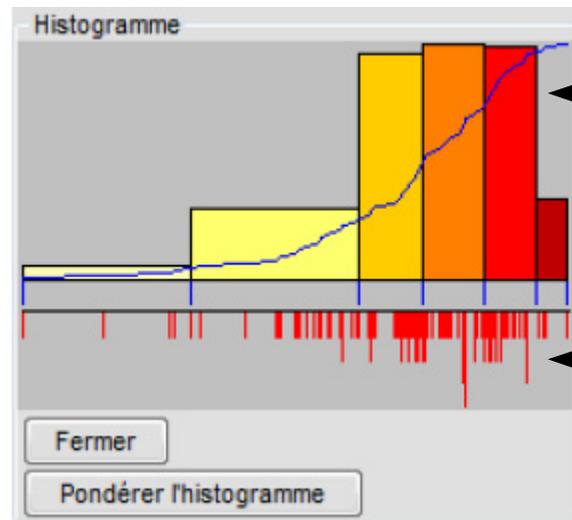
I) méthode de Jenks ou répartition automatique

Fondée sur la notion de variance. Le découpage en classes est déterminé de façon à minimiser la variance intra-classe et à maximiser la variance interclasses (variance : valeur moyenne du carré des écarts à la moyenne). Présente l'avantage de proposer un découpage où les individus d'une même classes sont es plus proches possibles (faible variance intra-classes) et de bien isoler les différents groupes (forte variance inter-classes). Cette méthode respecte bien la structure des données.



FICHE 13a DISCRÉTISATION & PHILCARTO 5

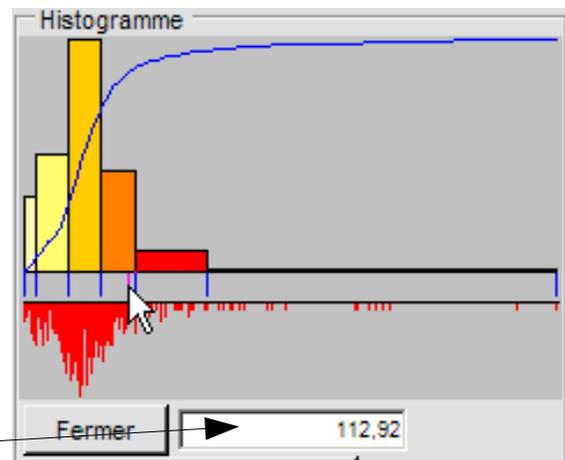
La discrétisation est paramétrable dans Philcarto en agissant directement sur l'histogramme.



La partie supérieure est occupée par l'histogramme (nb. de valeurs de chaque classe) et une courbe cumulative.

La partie inférieure est occupée par l'histogramme des fréquences permettant de repérer les éventuelles ruptures.

Les bornes bleues figurent les bornes de classes qui peuvent être paramétrables. Il suffit de cliquer gauche enfoncé sur la barre bleue et de la déplacer. La valeur de la nouvelle borne s'affiche dans le champ spécifique.

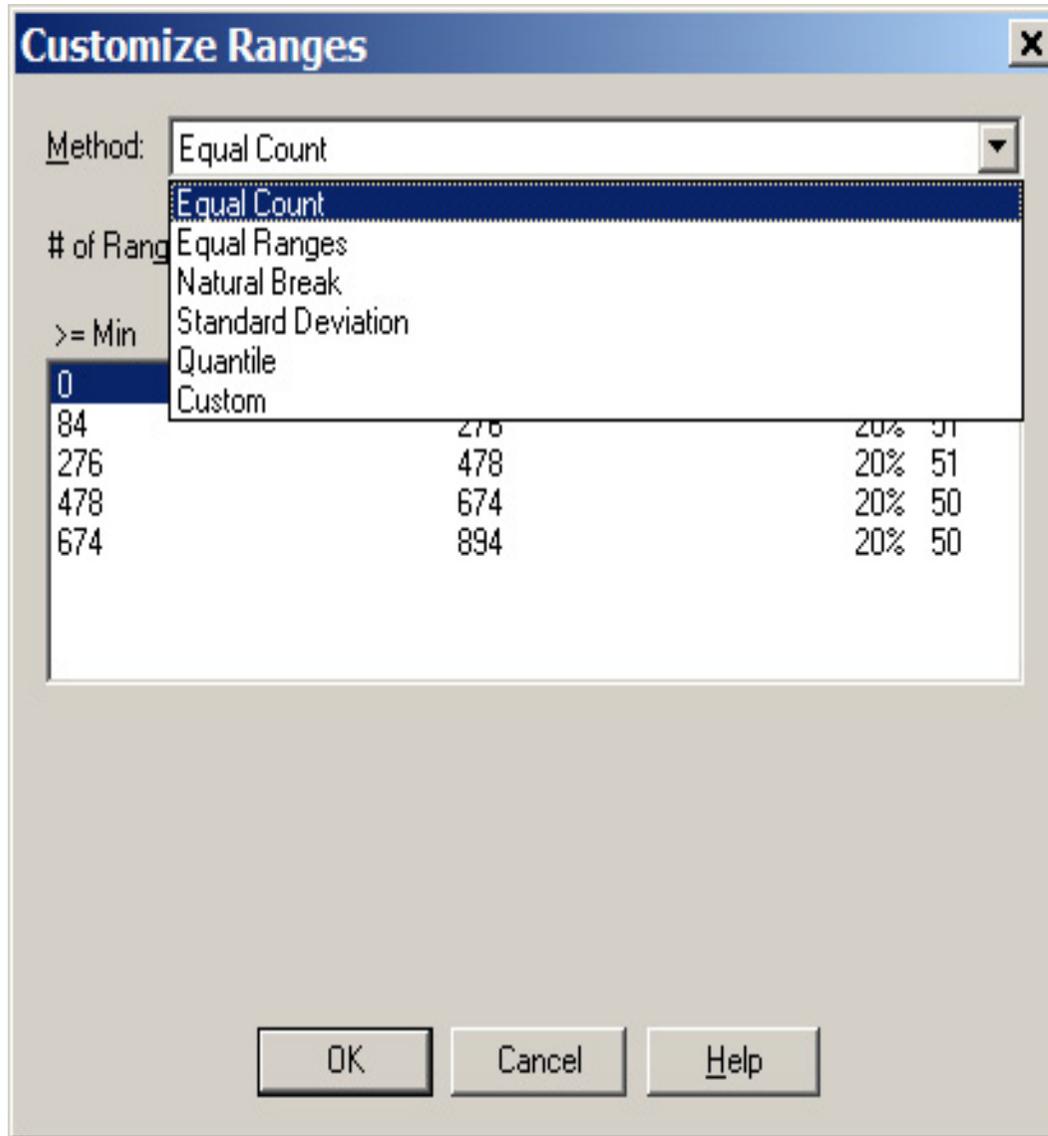


Pondérer l'histogramme

La pondération permet de sélectionner une nouvelle variable quantitative absolue. Seul l'histogramme subit une modification, la carte n'est pas modifiée.

FICHE 13b DISCRÉTISATION & MAPINFO

La discrétisation est paramétrable dans Mapinfo en agissant à l'étape 3 de la carte thématique.



Même nb d'enregistrements

Même amplitude : la différence entre max et min de chaque classe est la même.

Répartition auto : algorithme pour limiter les différences entre la moyenne des valeurs et la différence des valeurs pour chaque classe (Méthode de Jenks).

Écart-type : Écart moyen des valeurs à leur moyenne arithmétique.

Quantile : distribue une variable à travers une seconde variable (!)

Intervalle manuel : vous déterminez les classes

FICHE 13c DISCRÉTISATION & ARCMAP

La discrétisation est paramétrable dans Arcmap en agissant à l'étape 3 de la carte thématique.

Même nb d'enregistrements

Même amplitude : la différence entre max et min de chaque classe est la même.

Répartition auto : algorithme pour limiter les différences entre la moyenne des valeurs et la différence des valeurs pour chaque classe (Méthode de Jenks).

Écart-type : Écart moyen des valeurs à leur moyenne arithmétique.

Quantile : distribue une variable à travers une seconde variable (!)

Intervalle manuel : vous déterminez les classes

FICHE 14 DISCRÉTISATION & COMPARAISON

La comparaison est un objectif spécifique de la cartographie

Que l'on souhaite comparer des positions relatives ou absolues, il s'agit de

- comparer des distributions de deux phénomènes sur un même espace
- comparer des distributions d'un même phénomène sur un même espace sans le temps
- comparer des distribution d'un même phénomène sur deux espaces différents.

FICHE 15a REPRÉSENTATIONS GRAPHIQUES

3 règles de lisibilité : encombrement du dessin (densité), rapport entre les deux/trois dimensions du plan (lisibilité angulaire), mise en valeur de la variable (lisibilité rétinienne).

- *La densité graphique* est relative à la charge d'un graphique en information textuelle et/ou Graphique (densité forte, dessin illisible, densité trop faible, dessin inutile).
- *La lisibilité angulaire* est relative au fait qu'un graphique peut nuire à sa lisibilité Par exemple Un diagramme par exemple doit être compris dans un rapport 1X2 : hauteur au maximum de 2 largeurs, et la largeur doit correspondre à 2 hauteurs.
- *La lisibilité rétinienne* est relative au contraste des informations.

Indispensables :

- 1) Un titre général
- 2) Des titres sur chaque axe
- 3) Les sources (dont la date)
- 4) Légende

LE DIAGRAMME A BANDES

La base est constante et la hauteur proportionnelle à l'effectif ou la fréquence.

LE DIAGRAMME A SECTEURS

A chaque modalité est affecté un secteur de cercle dont la surface est proportionnelle aux effectifs.

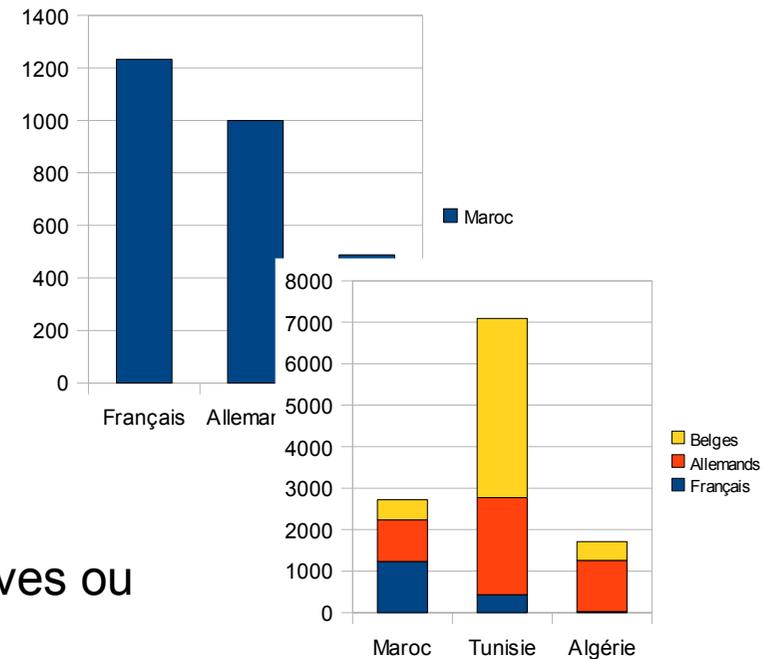
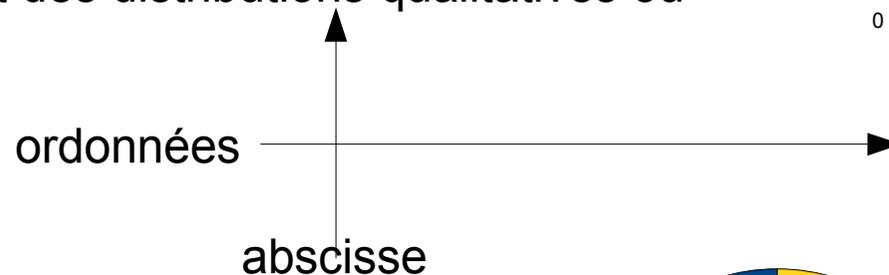
FICHE 15b REPRÉSENTATIONS GRAPHIQUES : LE DIAGRAMME

Le diagramme n'est que la représentation de différentes modalités.

A) Diagramme « Bâtons »

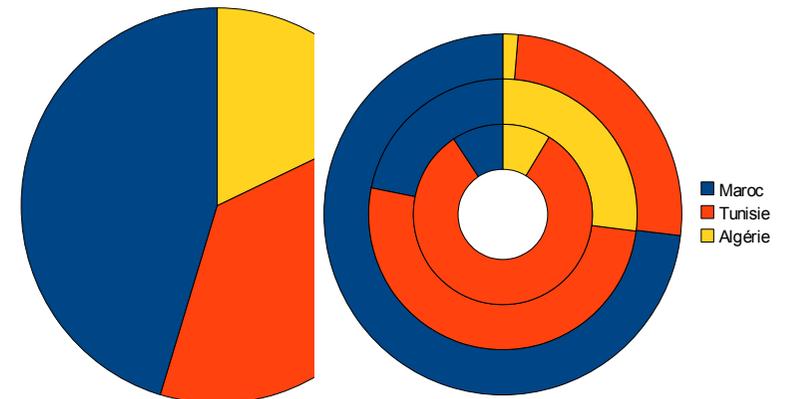
Le diagramme a une base constante et la hauteur est proportionnelle à l'effectif ou la fréquence contrairement à l'histogramme.

Il sert à représenter graphiquement des distributions qualitatives ou quantitatives discrètes sans regroupement en classes.



B) Le diagramme « secteurs ».

Le diagramme peut être présenté sous la forme de secteurs dénommés usuellement « camemberts ». L'angle de chaque secteur représente une proportion de l'effectif.



FICHE 15c REPRÉSENTATIONS GRAPHIQUES : L'HISTOGRAMME

L'histogramme est un mode de représentation des distributions à caractère quantitatif continu dont les valeurs sont regroupées en classes.

A) L'histogramme

Ce graphique est un ensemble de rectangles contigus où chaque rectangle est associé à une classe (axe des) et dont la surface est proportionnelle à l'effectif ou à la fréquence de cette classe (axe des).

B) Les courbes de fréquences

Les courbes peuvent permettre de croiser les fréquences croissantes et décroissantes et déterminer ainsi des classes.

FICHE 15d REPRESENTATION GRAPHIQUE COMPAREE : LA PYRAMIDE DES AGES

La pyramide des ages n'est ni plus ni moins que la représentation graphique de deux sous-populations qui partagent des classes similaires.

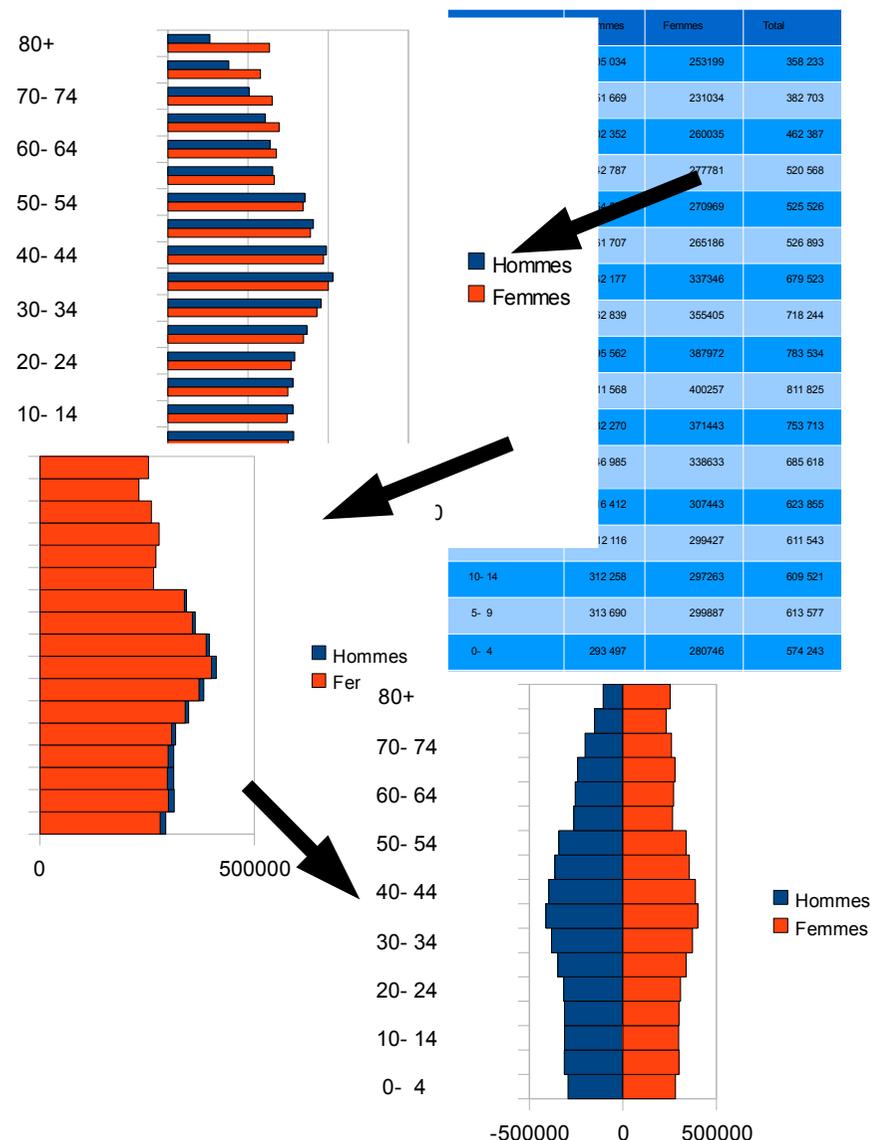
La procédure est la suivante:

1) construire un histogramme avec les deux valeurs

2) réduire le chevauchement des barres à 100 % l'espacement à 0%

3) Dans une des colonnes les données doivent être transformées en valeurs négatives.

Un travail est ensuite nécessaire pour positionner l'axe au centre et éliminer cosmétiquement les valeurs négatives.



FICHE 16 LISTE DES FORMULES OPEN OFFICE V3

SOUS.TOTAL SOMME	Renvoie SOMME, MOYENNE, ECARTYPE, etc. résultats de données filtrées. Somme le contenu des cellules. SOMME(nombre_1;nombre_2;...;nombre_30)
FREQUENCE FREQUENCE (Données ; Classes)	Renvoie une matrice catégorisant les valeurs d'un ensemble de données dans des intervalles donnés.
MOYENNE MOYENNEA NB NBVAL MEDIANE MODE MOYENNE.REDUITE	Renvoie la moyenne des arguments, en ignorant le texte. Renvoie la moyenne des arguments, incluant du texte (évalué comme 0). Compte les nombres dans une liste d'arguments, ignorant le texte. Compte les valeurs non vides dans la liste des arguments. Renvoie la médiane d'un ensemble de nombres. Renvoie la valeur commune dans un ensemble de nombres. Renvoie la moyenne d'un ensemble de nombres, ignorant une proportion de valeurs hautes et basses.
SOMME.CARRES.ECARTS ECARTYPE ECARTYPEA ECARTYPEP ECARTYPEPA VAR VARA VAR.P VAR.PA	Renvoie la somme des carrés de la déviation de la moyenne. Renvoie l'écart type de l'échantillon. Renvoie l'écart type de l'échantillon (autorisant du texte et des valeurs logiques). Renvoie l'écart type de la population. Renvoie l'écart type de la population (autorisant du texte et des valeurs logiques). Renvoie la variance de l'échantillon. Renvoie la variance de l'échantillon (autorisant du texte et des valeurs logiques). Renvoie la variance de la population. Renvoie la variance de la population (autorisant du texte et des valeurs logiques).
MAX MIN CENTILE RANG.POURCENTAGE QUARTILE RANG PETITE.VALEUR	Renvoie la valeur maximum dans une liste d'arguments. Renvoie la valeur minimum dans une liste d'arguments. Renvoie un centile spécifié dans une liste de nombres. Renvoie le rang de pourcentage d'un nombre dans une liste de nombres. Renvoie un quartile spécifié dans une liste de nombres. Renvoie le rang d'un nombre dans une liste de nombres. Renvoie la nème plus petite valeur dans une liste de nombres.