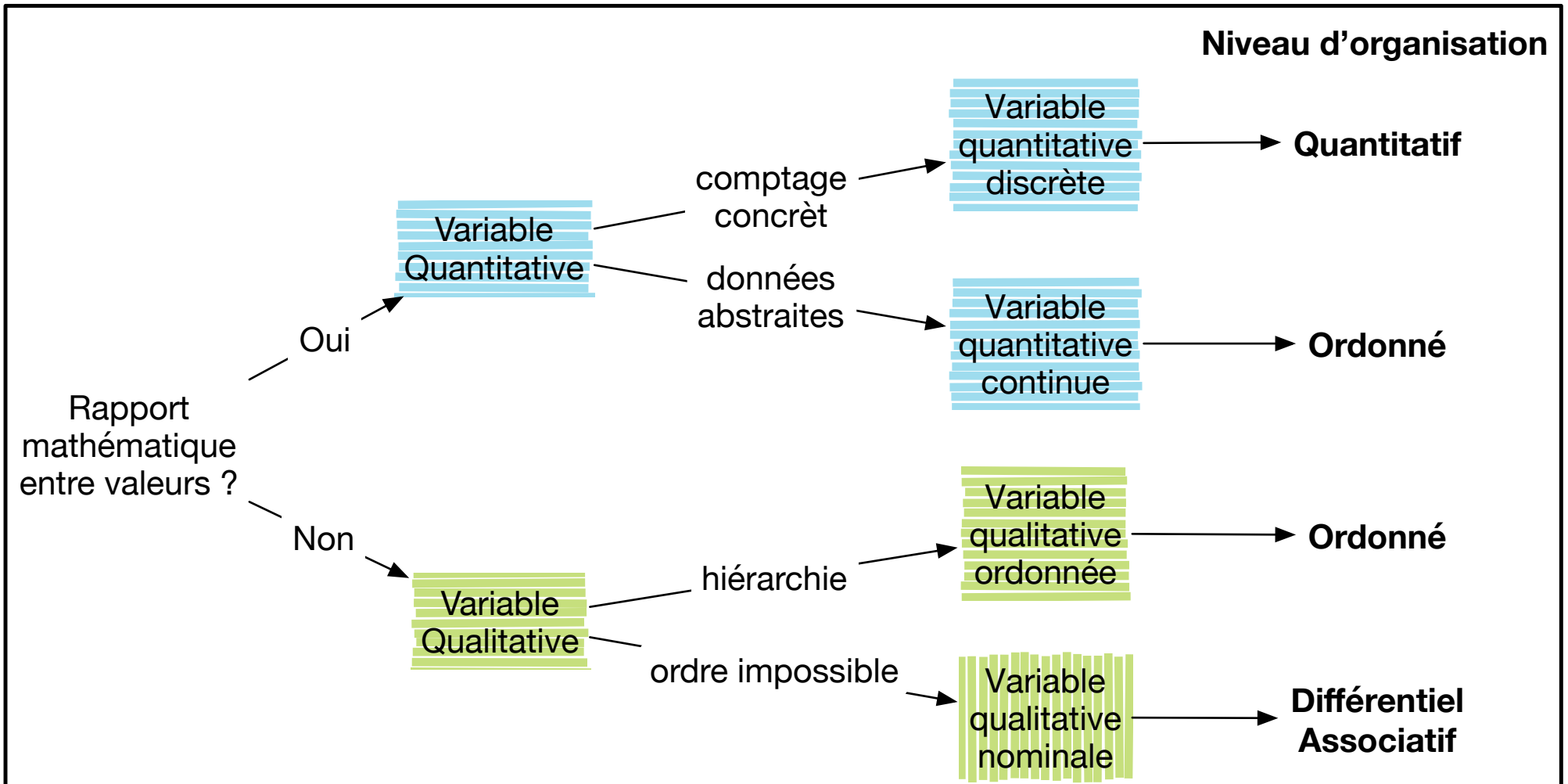

CONCEPTION CARTOGRAPHIQUE : SÉMIOLOGIE ET DISCRÉTISATIONS

L3 APTER / L. JÉGOU



OBJECTIFS DE LA SÉANCE

- Rappels de méthodologie : sémiologie graphique
- Comprendre l'intérêt de réduire l'information
- Discrétisation : connaître les différentes méthodes, leurs domaines d'application et leurs conséquences sur la carte produite



NIVEAU DES VARIABLES RÉTINIENNES

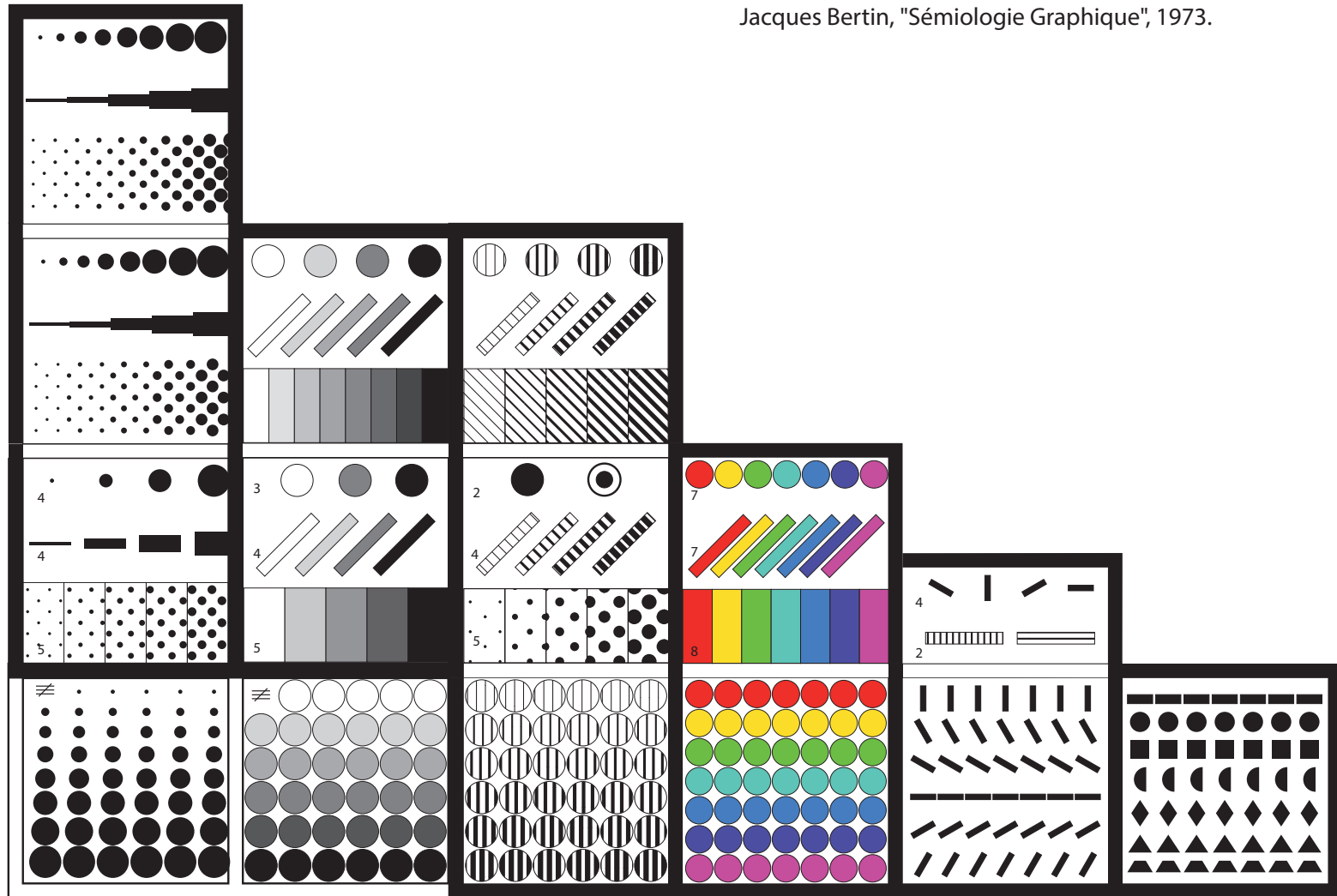
Jacques Bertin, "Sémiologie Graphique", 1973.

QUANTITÉ
Q
Tous les signaux sont perçus
PROPORTIONNELS
entre eux

ORDRE
O
Tous les signaux
sont perçus comme
ORDONNÉS

SÉLECTION
≠
Tous les signaux
sont perçus comme
DIFFÉRENTS
et forment des FAMILLES

ASSOCIATION
≡
Tous les signaux
peuvent être perçus comme
SEMBLABLES



TAILLE

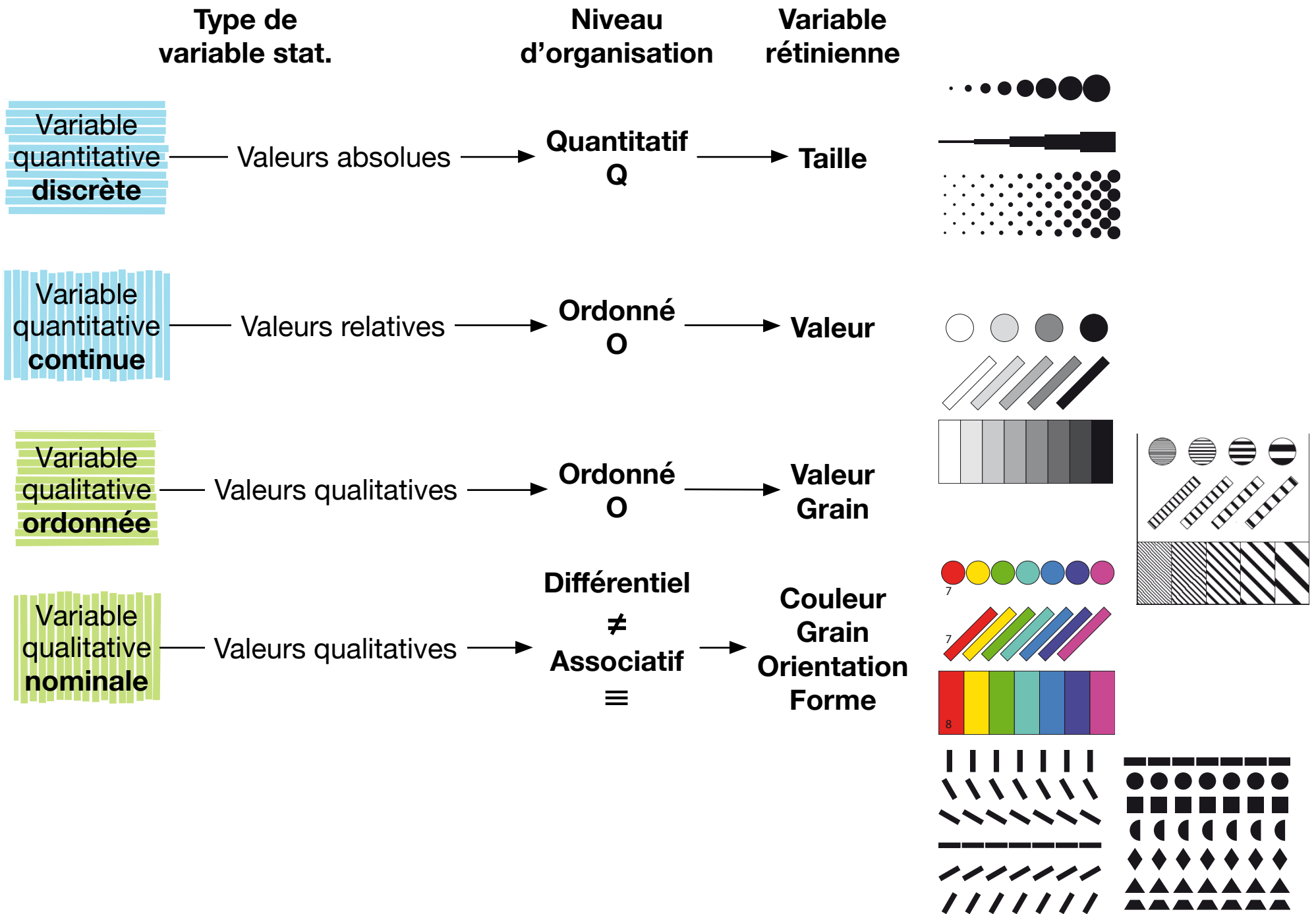
VALEUR

GRAIN

COULEUR

ORIENTATION

FORME

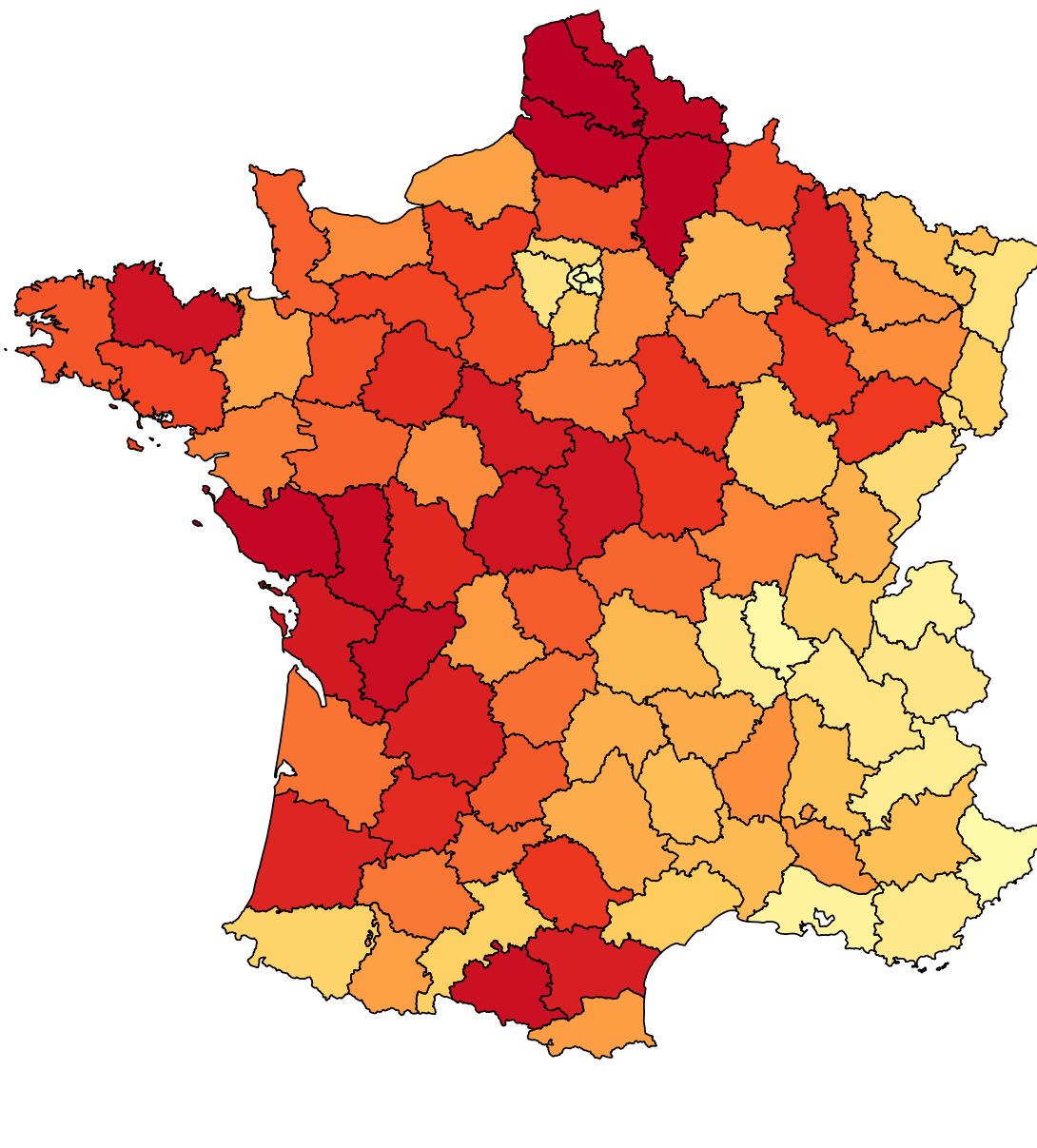




REGROUPER EN CLASSES POUR SIMPLIFIER : LA DISCRÉTISATION

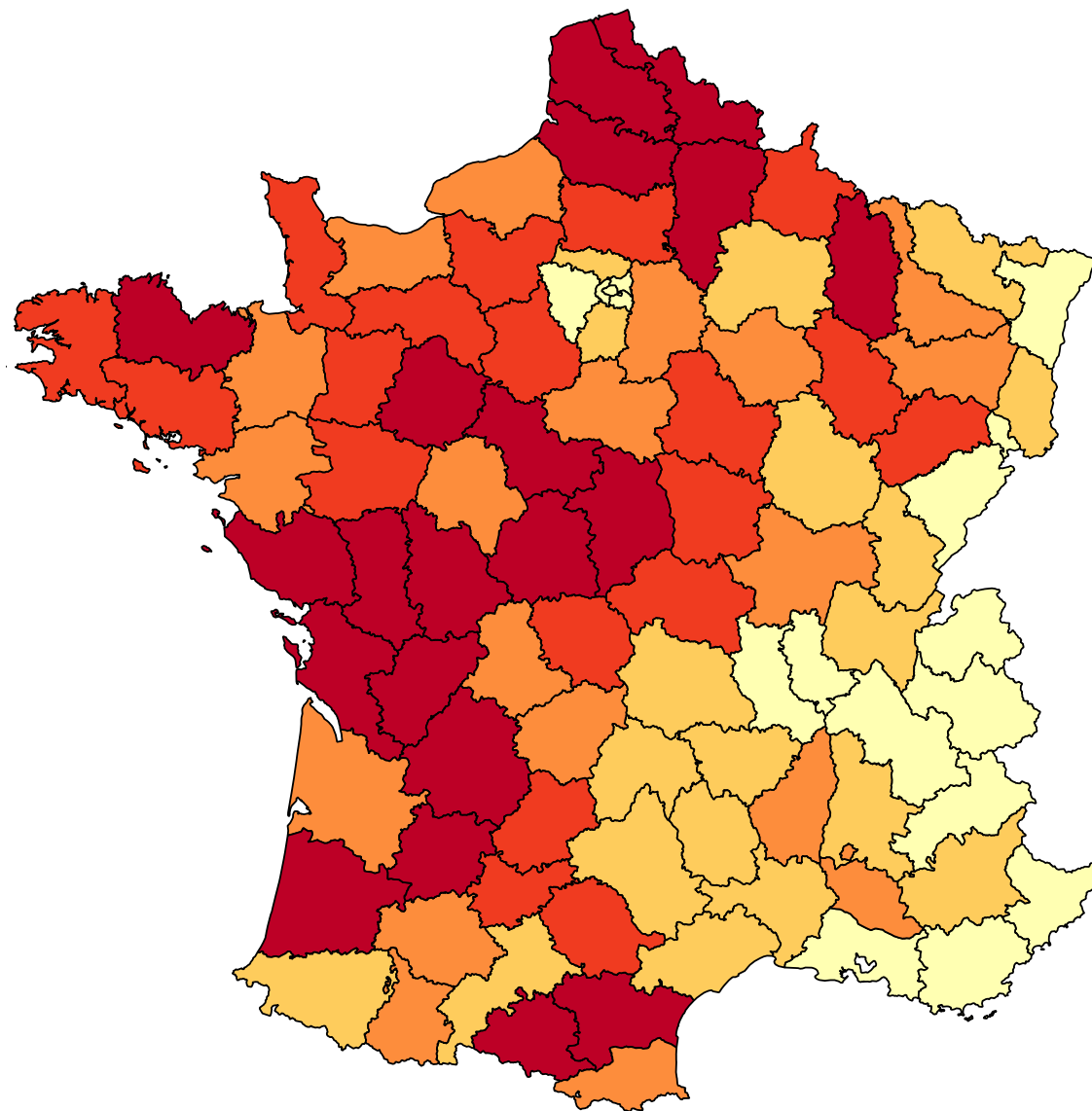
Part des maisons individuelles dans le total des logements, en 1990

En pourcentages



France data départements		frdepts	
69.31 - 73.09	56.75 - 56.92	45.95 - 45.99	
68.57 - 69.31	56.52 - 56.75	45.77 - 45.95	
67.77 - 68.57	56.13 - 56.52	45.37 - 45.77	
67.53 - 67.77	55.65 - 56.13	44.92 - 45.37	
67.44 - 67.53	55.51 - 55.65	44.51 - 44.92	
66.55 - 67.44	55.26 - 55.51	44.07 - 44.51	
66.09 - 66.55	54.80 - 55.26	43.06 - 44.07	
65.96 - 66.09	54.71 - 54.80	41.58 - 43.06	
64.90 - 65.96	54.62 - 54.71	40.98 - 41.58	
63.94 - 64.90	53.42 - 54.62	40.88 - 40.98	
63.66 - 63.94	53.17 - 53.42	40.44 - 40.88	
63.35 - 63.66	53.08 - 53.17	39.84 - 40.44	
62.83 - 63.35	52.99 - 53.08	38.52 - 39.84	
62.17 - 62.83	52.89 - 52.99	38.38 - 38.52	
61.84 - 62.17	52.13 - 52.89	38.07 - 38.38	
61.63 - 61.84	51.50 - 52.13	37.07 - 38.07	
61.61 - 61.63	51.27 - 51.50	36.53 - 37.07	
61.55 - 61.61	50.98 - 51.27	36.26 - 36.53	
61.12 - 61.55	50.40 - 50.98	35.38 - 36.26	
61.05 - 61.12	49.69 - 50.40	34.81 - 35.38	
61.03 - 61.05	48.83 - 49.69	33.14 - 34.81	
60.98 - 61.03	48.15 - 48.83	31.75 - 33.14	
60.72 - 60.98	48.12 - 48.15	31.22 - 31.75	
60.39 - 60.72	48.03 - 48.12	30.42 - 31.22	
60.24 - 60.39	47.82 - 48.03	29.47 - 30.42	
59.94 - 60.24	47.62 - 47.82	28.69 - 29.47	
59.79 - 59.94	47.50 - 47.62	27.80 - 28.69	
58.45 - 59.79	47.36 - 47.50	26.53 - 27.80	
57.35 - 58.45	47.14 - 47.36	24.11 - 26.53	
57.09 - 57.35	47.08 - 47.14	22.17 - 24.11	
56.92 - 57.09	46.41 - 47.08	21.01 - 22.17	
	46.27 - 46.41	16.40 - 21.01	
	46.16 - 46.27	12.01 - 16.40	
	46.06 - 46.16	0.79 - 12.01	
	45.99 - 46.06		

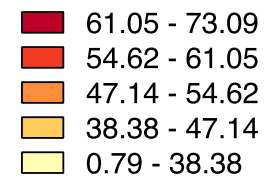
Part des maisons individuelles dans le total des logements, en 1990



En pourcentages

France data départements

frdepts



LE CHOIX DU NOMBRE DE CLASSES

Le nombre *maximal* de classes que l'on peut réaliser en préservant la lisibilité de la carte en couleurs peut varier entre 7 et 9.

Attention à ce que la qualité d'impression soit adaptée.

7 classes si le dégradé de couleurs se réalise sur une seule teinte :



9 classes s'il fait jouer deux teintes :



RAPPEL DES PRINCIPES

- ❑ Uniquement pour les variables **de type quantitatif continu** (ou relatif), généralement issues de calculs de ratios.
- ❑ Préserver les éléments importants de la variable : seuils (zéro pour +/-, indice 100, etc.), sens de variation.
- ❑ Respecter la forme de la distribution des valeurs (celle du *graphique de fréquences*)
- ❑ Créer des classes qui correspondent à des « groupes naturels » de valeurs (sous-ensembles).

CONCRÈTEMENT :

Respect de l'information d'origine : la répartition des couleurs sur la carte doit être fortement liée à la répartition des individus dans la variable (la distribution).

L'image colorée que représente la carte doit donner une idée **juste** de l'information géographique contenue dans la variable.

- ❑ Il ne doit pas y avoir de classe vide.
- ❑ Toutes les classes doivent être jointives (pas de valeurs entre deux classes successives).
- ❑ Les classes ne doivent pas se chevaucher (un individu ne peut appartenir qu'à une classe à la fois).

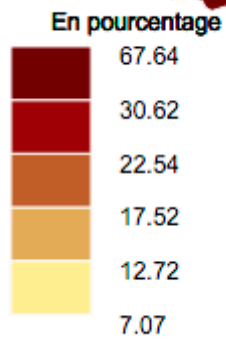
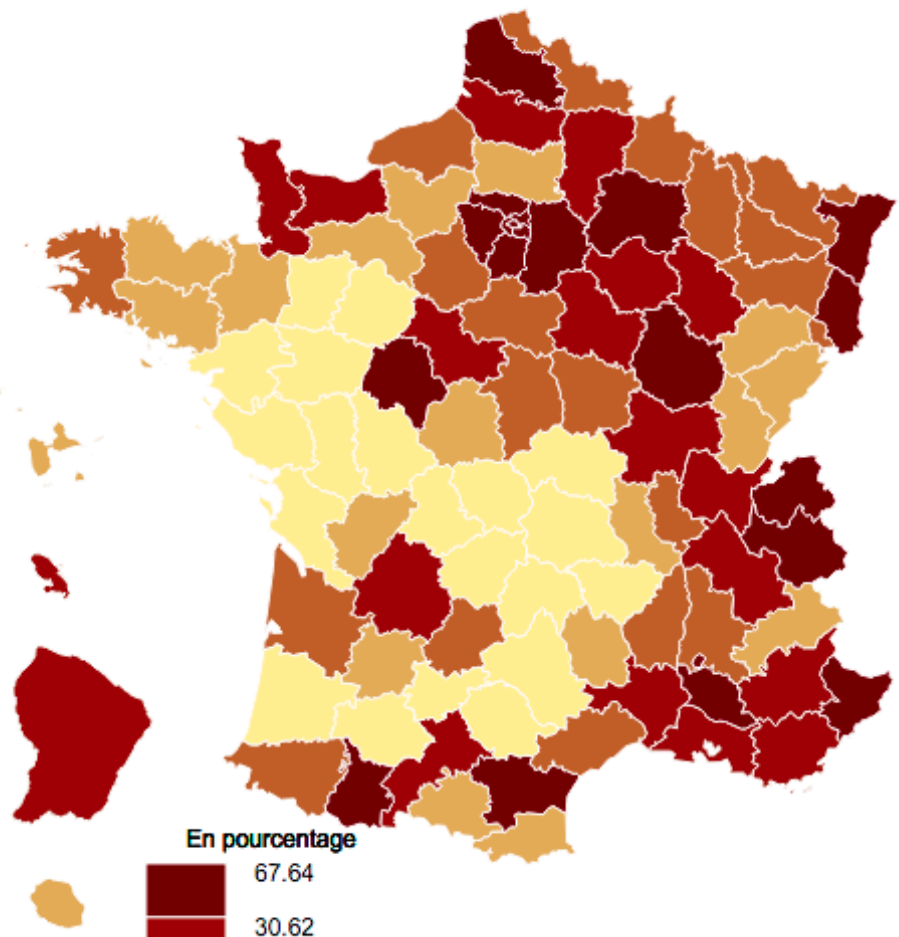
LES DIFFÉRENTES MÉTHODES DE DISCRÉTISATION AUTOMATIQUE

- Selon le logiciel de conception cartographique utilisé, différentes méthodes de regroupement en classes seront disponibles. Nous allons ici passer en revue les principales.
- Les illustrations sont issues de cet exemple interactif en ligne : <http://mappemonde.mgm.fr/geovisu/119/>
- Les données sont disponibles au format Excel ici : <http://www.geotests.net/test/discretisations/hotels2015etr.xls>

LA MÉTHODE DES QUANTILES (OU ÉQUIFRÉQUENCES)

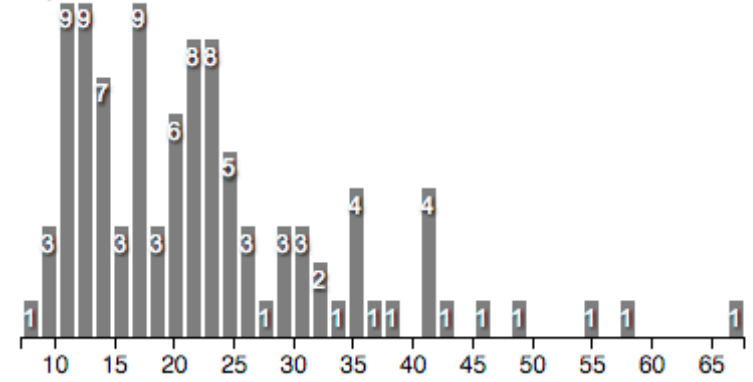
- **Méthode** : On compte le même nombre d'individus par classe
- **Calcul** : On divise la population par le nombre de classes désiré, on obtient le nombre d'individus par classe. Ensuite, on détermine les bornes des classes en triant la variable dans l'ordre croissant et en cherchant les bornes correspondant au nombre d'individus voulu.
- **Avantages** : La carte produite va présenter une répartition des couleurs assez équilibrée, si les unités qui composent le fond de carte sont sensiblement de même surface. Comme la taille de chaque classe est identique, on peut les comparer entre elles.
- **Inconvénients** : La méthode ne tient pas compte des valeurs des individus pour les regrouper, les classes vont donc pouvoir contenir des individus de valeurs assez différentes, notamment dans les classes extrêmes.

Part des nuitées de résidents étrangers en 2015

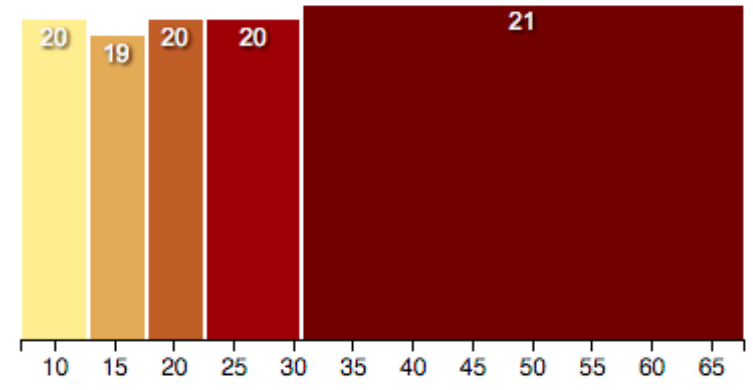


Source : INSEE, 2016
Réalisation : L. Jégou / UTJ2, 2016

Fréquences



Discretisation : quantiles



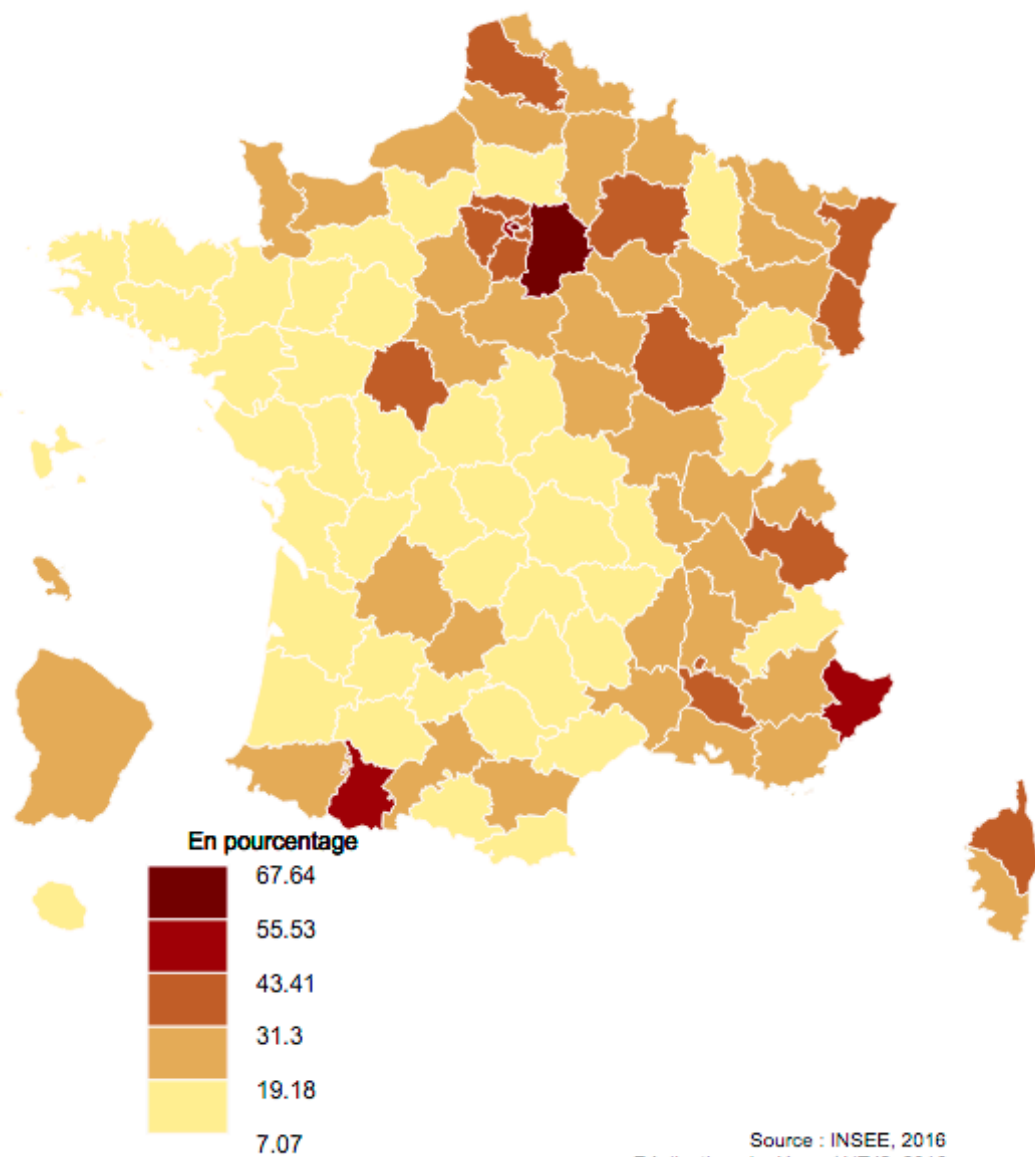
Nbre. de classes Nbre. de barres

La méthode des quantiles : exemple

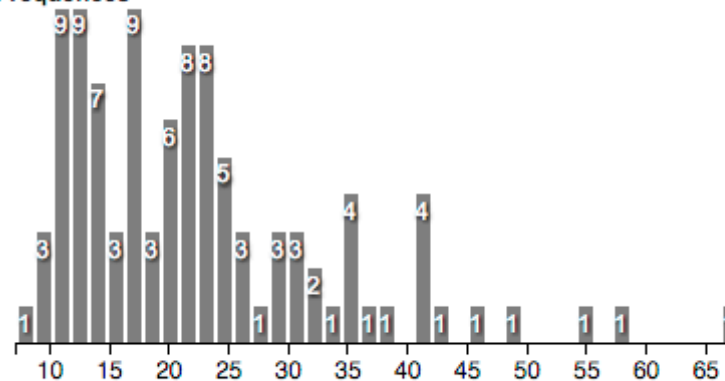
LA MÉTHODE DES ÉQUIVALENCES (OU ÉQUIDISTANCES)

- **Méthode** : même amplitude, écart de valeurs, par classe. Ainsi, chaque classe couvrira une portion identique de l'étendue totale de la variable.
- **Calcul** : On divise l'étendue totale de la variable (max-min) par le nombre de classes et on obtient l'étendue par classe. Ensuite, on détermine les bornes des classes en triant la variable dans l'ordre croissant et l'on cherche les bornes correspondant aux étendues trouvées.
- **Avantages** : Cette méthode de discrétisation n'est pas liée à la variable, donc on peut l'utiliser pour comparer des variables distinctes, des phénomènes différents. On peut aussi y voir un moyen de mesure de la distribution par rapport à une distribution uniforme.
- **Inconvénients** : Cette méthode est peu utilisée à cause de limites importantes. La plus grave est qu'il arrive souvent qu'une classe ainsi définie ne comprenne aucune valeur, ce qui fausse la discrétisation. Cette discrétisation a tendance à produire des classes de poids très inégal, et donc des cartes avec une répartition des couleurs relativement hétérogène.

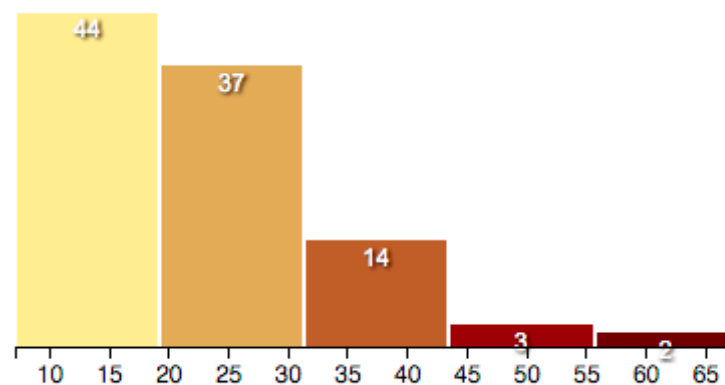
Part des nuitées de résidents étrangers en 2015



Fréquences



Discretisation : equivalences



Quantiles Équivalences Prog. arithmétique Prog. géométrique

Moyennes Standard Jenks

Nbre. de classes Nbre. de barres

La méthode des équivalences : exemple

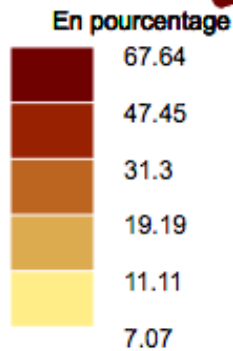
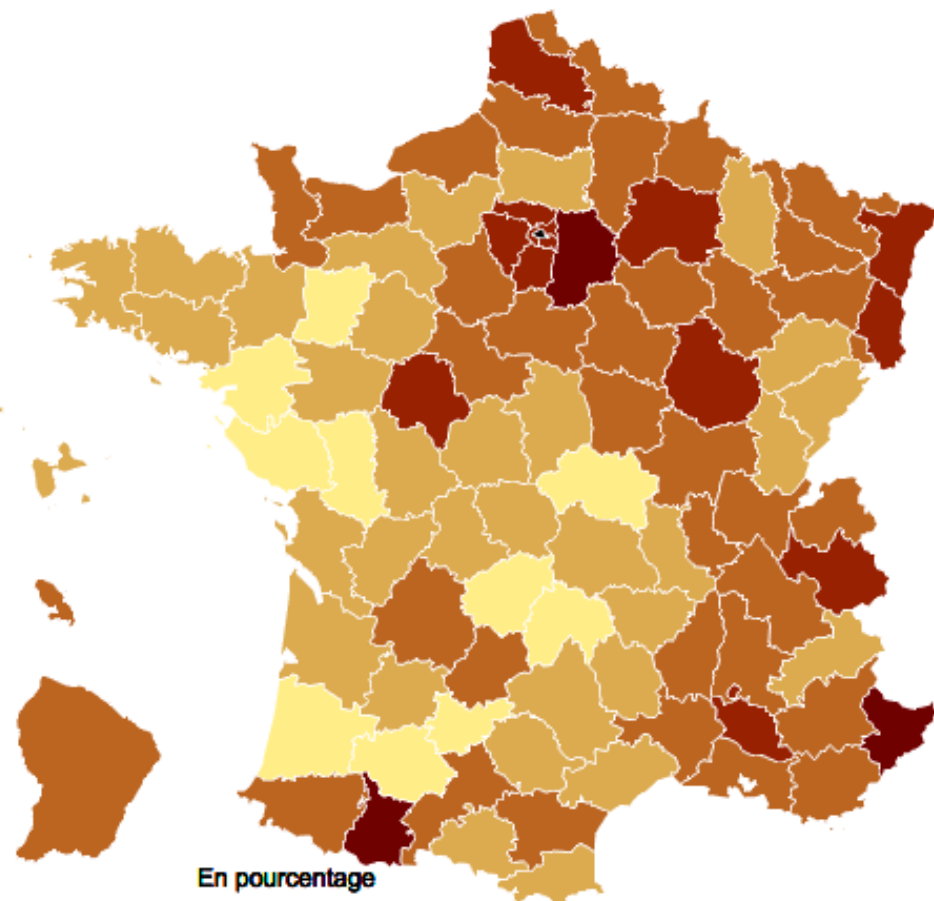
LA MÉTHODE DE LA PROGRESSION ARITHMÉTIQUE

Méthode : Cette méthode consiste à découper la variable avec des classes dont l'étendue va suivre une progression arithmétique (une suite).

Calcul :

1. Pour commencer on calcule la raison R de la suite :
 $R = \text{étendue} / (n!)$
Exemple pour 5 classes : $R = \text{étendue} / (1+2+3+4+5)$
2. Ensuite, on calcule l'amplitude des classes en multipliant la raison R par le numéro de la classe :
C1 [min, min+R]
C2 [min+R, min+2.R]
C3 [min+2.R, min+3.R]
etc.

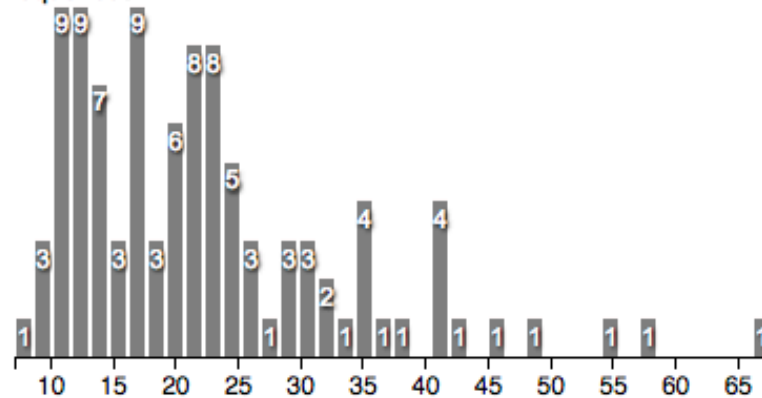
Part des nuitées de résidents étrangers en 2015



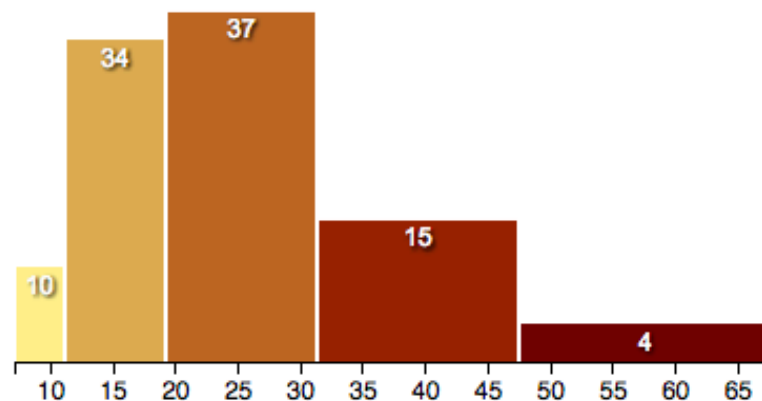
200 km

Sources : INSEE, 2016, IGN
Réalisation : M@ppemonde - 12/2016

Fréquences



Discretisation : progarithm



Nbre. de classes

Nbre. de barres de fréquences

La méthode de la progression arithmétique : exemple

LA MÉTHODE DE LA PROGRESSION ARITHMÉTIQUE

- La méthode est conçue pour les distributions *asymétriques* avec beaucoup de valeurs faibles et peu de valeurs fortes.
- Si, au contraire, la distribution s'apparente à la loi normale (courbe en cloche), la méthode n'est pas adaptée et produira des classes très hétérogènes.

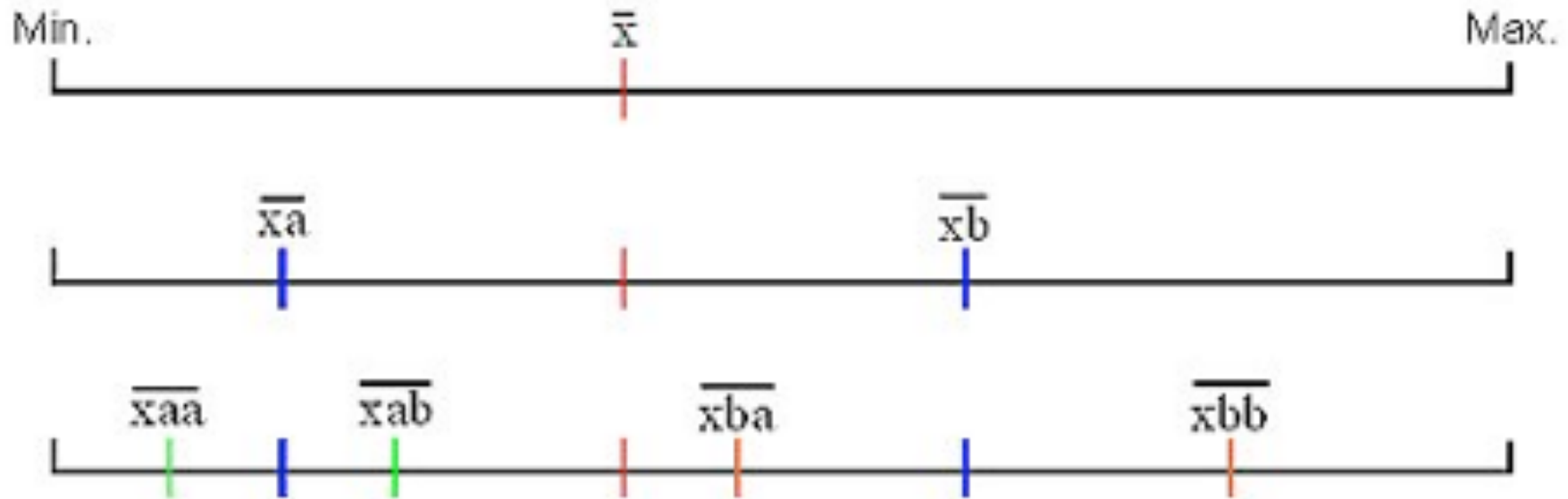
LA MÉTHODE DES MOYENNES EMBOÎTÉES

Méthode : Cette méthode consiste à découper la variable en utilisant des moyennes successives comme limites de classes.

Calcul : Cette méthode ne peut produire que 4 ou 8 classes exploitables. Voilà pourquoi, étape par étape :

1. Calculer la moyenne de la variable, M . M découpe la variable en deux parties, A et B .
2. Calculer la moyenne M_A de la partie A de la variable. M_A découpe la partie A en deux classes : AA et AB .
3. Calculer la moyenne M_B de la partie B de la variable. M_B découpe la partie B en deux classes : BA et BB .

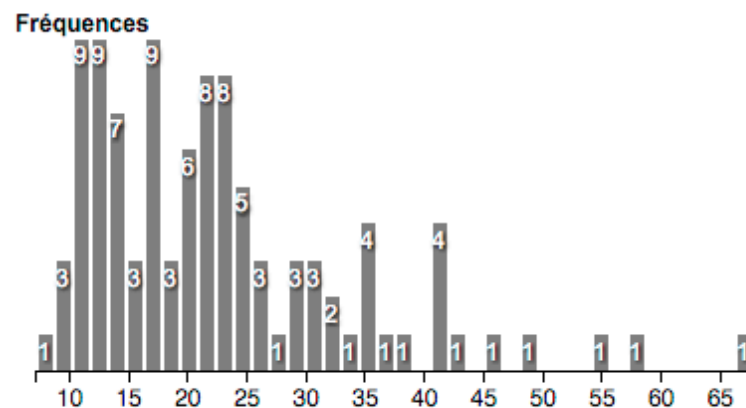
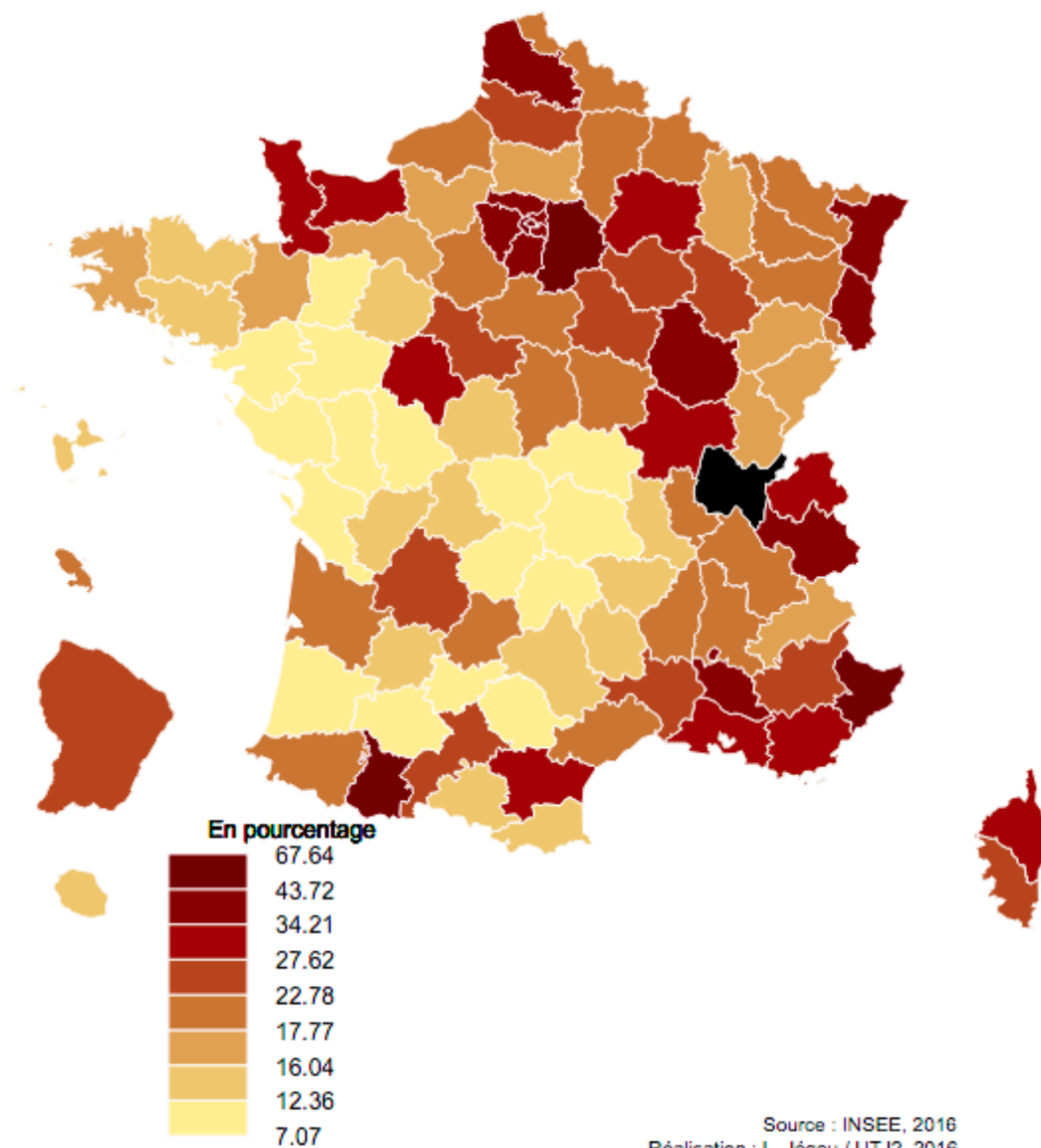
On obtient donc quatre classes, AA , AB , BA et BB , en ayant calculé trois moyennes emboîtées. Si l'on poursuit le processus encore une fois, on obtient 8 classes.



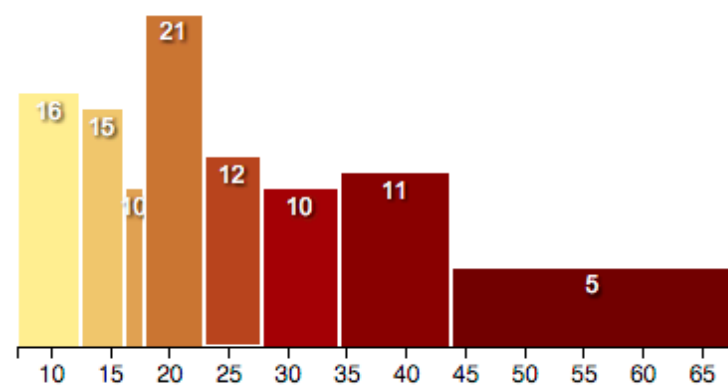
Avantages : Cette méthode a l'avantage d'être fortement liée à la distribution de la variable. Si vous réalisez une carte unique et que vous voulez « coller » à la distribution, cette méthode est généralement assez bien adaptée. Autre avantage : comme les limites des classes sont des *moyennes*, les cartes produites sont aisément interprétables.

Inconvénients : De par sa méthode de calcul, cette discrétisation peut produire des classes vides ou très hétérogènes, dans le cas de distributions très dissymétriques ou plurimodales. Enfin, le nombre de classes produit, fixe, peut ne pas être adapté au nombre d'individus de la variable.

Part des nuitées de résidents étrangers en 2015



Discrétisation : moyennes



La méthode des moyennes emboîtées : exemple

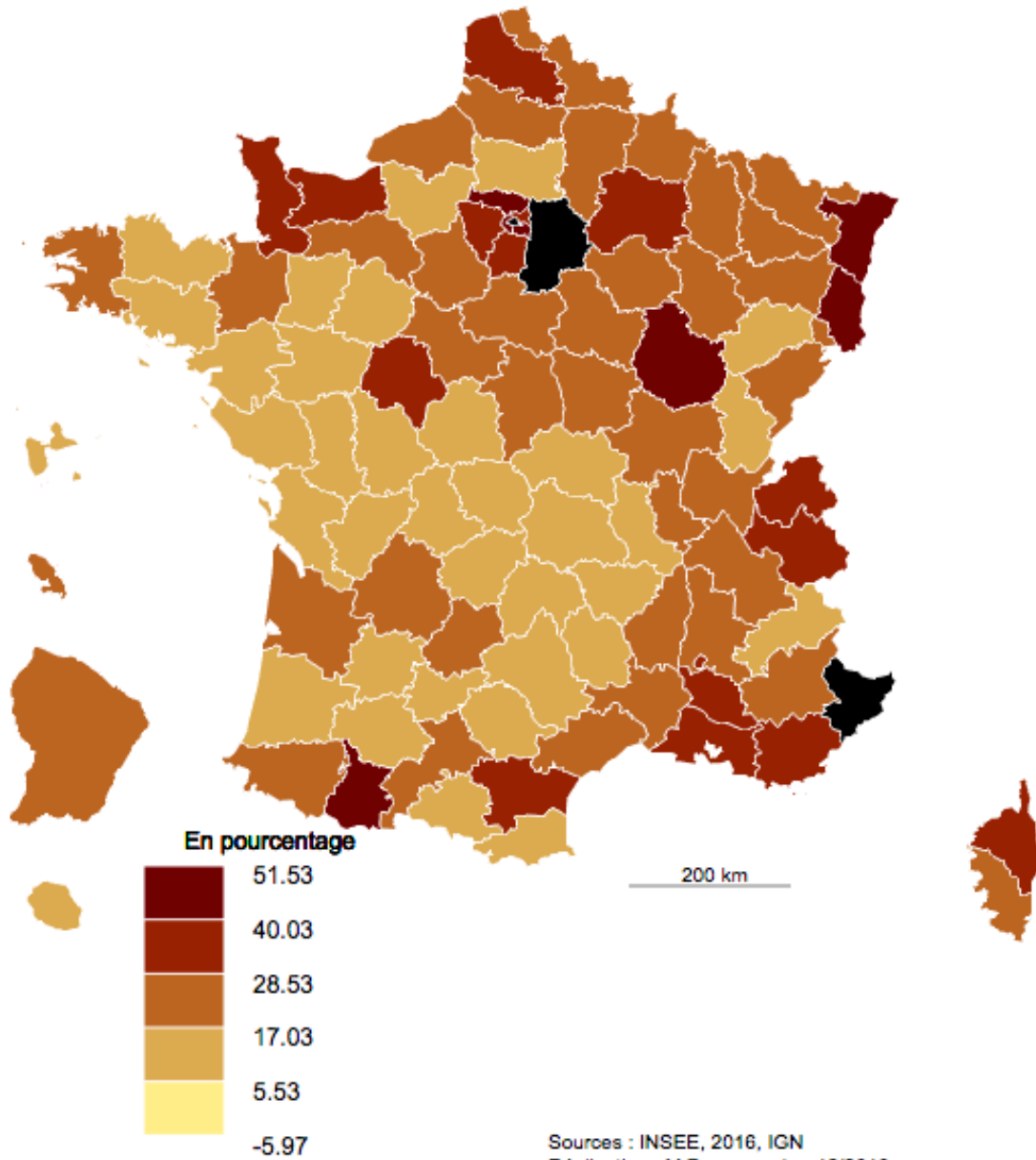
LA MÉTHODE « STANDARD » (BASÉE SUR L'ÉCART-TYPE)

Méthode : « *Standard deviation* » signifie écart-type en anglais. Cette discrétisation est basée sur la moyenne, à laquelle on enlève ou on ajoute des fractions ou des multiples d'écart-type.

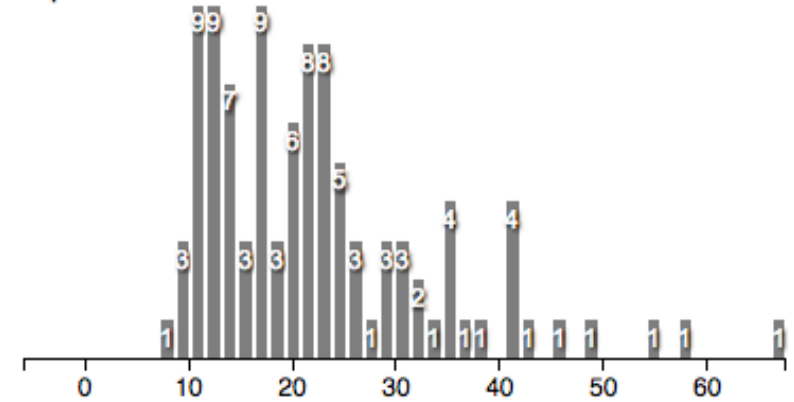
Calcul : On calcule d'abord la moyenne M de la variable, puis son écart-type E . Ensuite, suivant le nombre de classes voulu, on applique la procédure suivante :

- 3 classes:
Min, $m-0.5E$, $m+0.5E$, Max
- 5 classes:
Min, $m-1.5E$, $m-0.5E$, $m+0.5E$, $m+1.5E$, Max
- 7 classes:
Min, $m-2.5E$, $m-1.5E$, $m-0.5E$, $m+0.5E$, $m+1.5E$, $m+2.5E$, Max

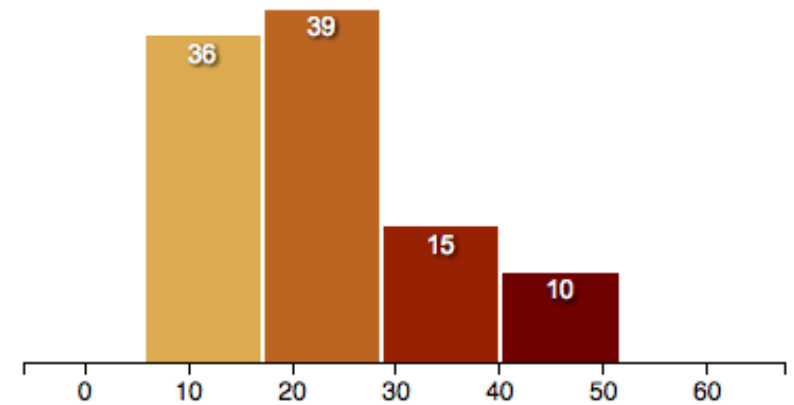
Part des nuitées de résidents étrangers en 2015



Fréquences



Discrétisation : standard



Nbre. de classes

Nbre. de barres de fréquences

La méthode « standard » : exemple

LA MÉTHODE « STANDARD » (BASÉE SUR L'ÉCART-TYPE)

- **Avantages** : Cette méthode est basée sur des indicateurs statistiques de *centralité* et de *dispersion*. Elle permet d'obtenir des séries de cartes comparables entre elles, de plus avec un découpage en classes sensible à la *dispersion* de la variable.
- **Inconvénients** : C'est encore une méthode très liée à la variable et à sa dispersion. Donc à *ne pas utiliser sur des dispersions trop grandes*, des dissymétries trop importantes, sous peine de classe vide. De plus, comme pour les moyennes emboîtées, cette méthode s'organise autour d'une moyenne centrale. Elle ne fonctionnera donc pas très bien avec des distributions *plurimodales* où la moyenne est moins significative.

LA MÉTHODE DE JENKS (OU NUAGES DE POINTS, VOIRE « SEUILS NATURELS »)

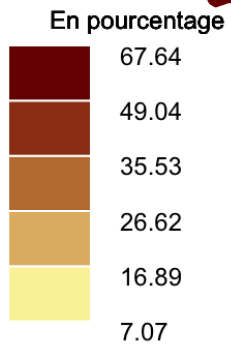
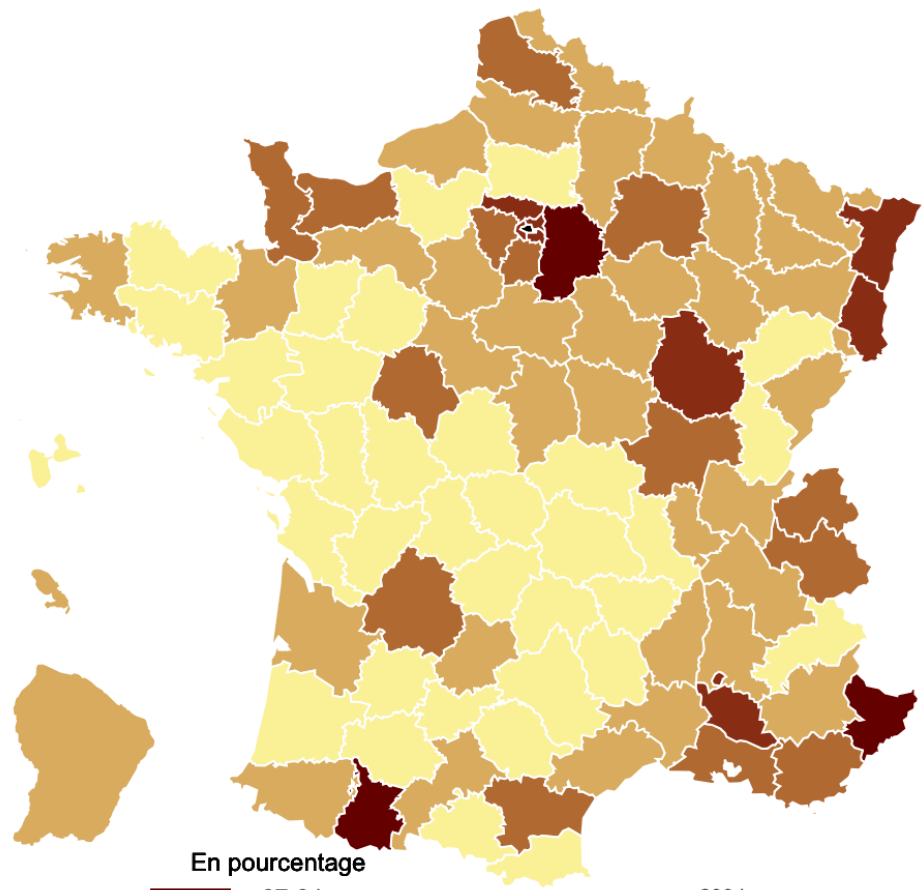
Méthode : La méthode de Jenks est l'une des plus intéressantes car elle permet d'obtenir des classes très proches de la distribution de la variable, en créant des classes contenant des valeurs homogènes mais des classes les plus hétérogènes entre elles. Elle a été mise au point par G. F. Jenks, un cartographe américain, dans les années 1960.

Calcul : Elle n'est calculable que par informatique, car elle utilise des traitements qui peuvent être longs. Le calcul va chercher à réduire la variance intraclasse et à maximiser la différence entre les moyennes des classes, en faisant varier finement les bornes des classes par étapes.

Avantages : Cette méthode produira une carte assez objective, *si le nombre de classes choisi est assez grand*. De par son calcul, elle est mieux adaptée aux variables multimodales et possédant un nombre d'individus assez important pour pouvoir faire ressortir des seuils.

Inconvénients : Le regroupement produit est très dépendant de la distribution de la variable, donc l'utilisation sur une série de différentes variables ne produira pas des cartes facilement comparables entre elles.

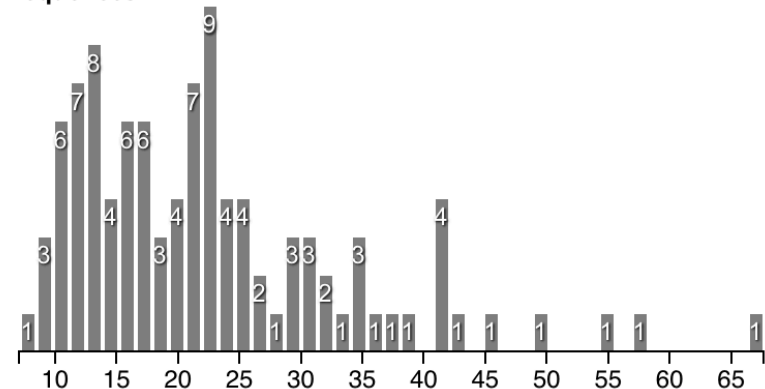
Part des nuitées de résidents étrangers en 2015



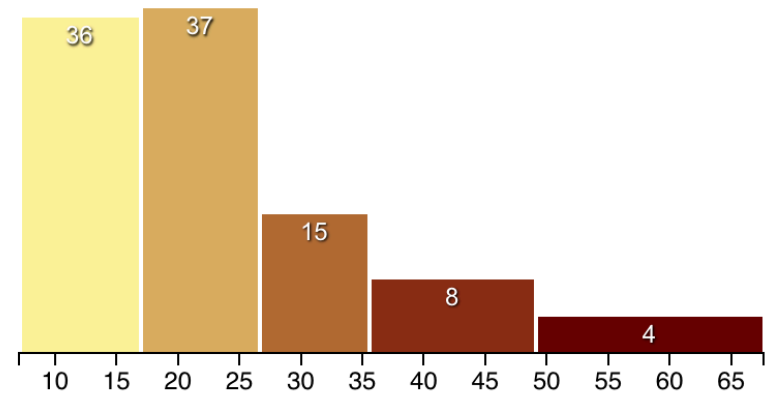
200 km

Sources : INSEE, 2016, IGN
Réalisation : M@ppemonde - 12/2016

Fréquences



Discretisation : Jenks

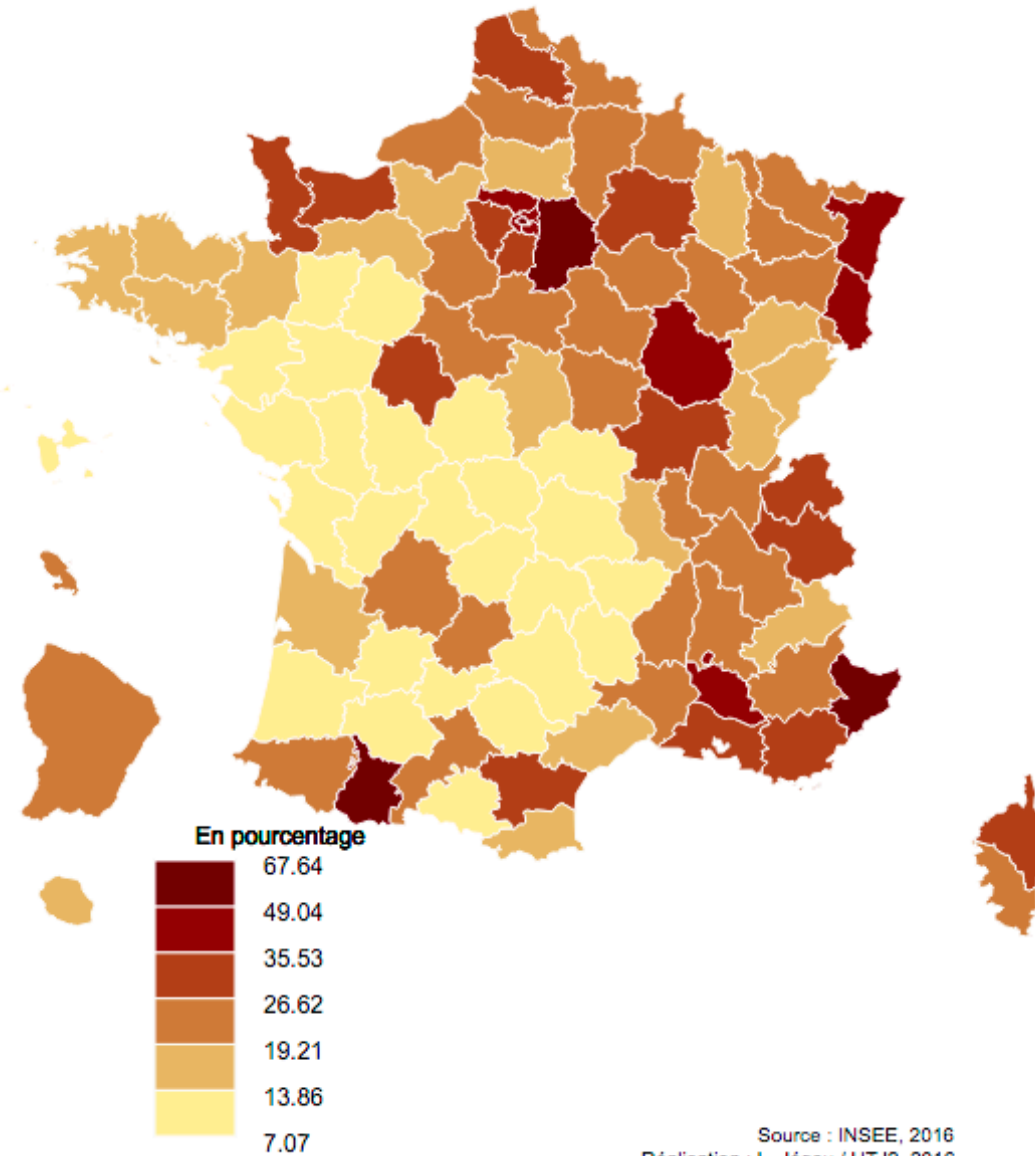


Nbre. de classes

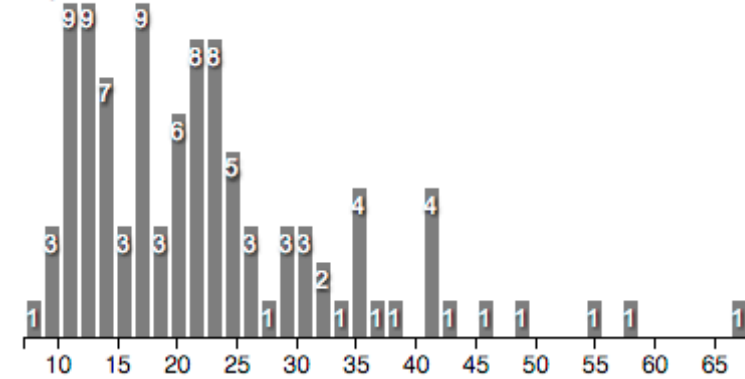
Nbre. de barres de fréquences

**La méthode de Jenks :
exemple avec 5 classes**

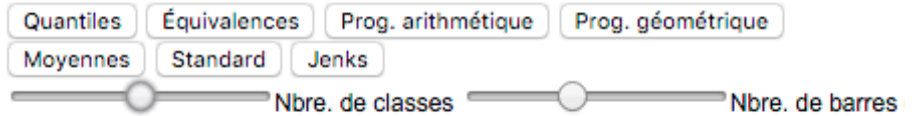
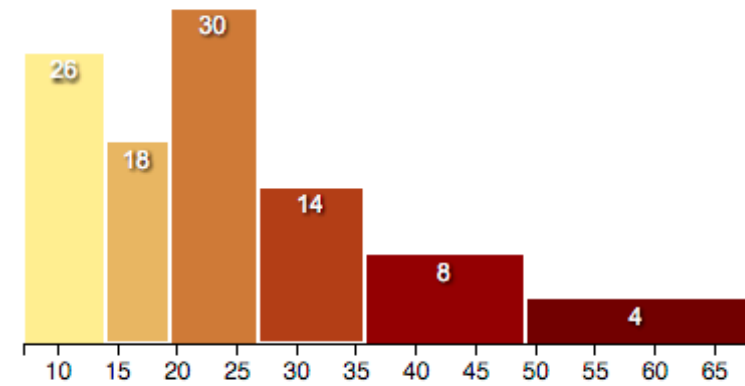
Part des nuitées de résidents étrangers en 2015



Fréquences



Discretisation : Jenks



**La méthode de Jenks :
exemple avec 6 classes**