UE 202
Conception cartographique

Séance 6 : Initiation à Magrit
- Cartes en symboles prop.
- Les regroupements en classes, ou discrétisations

TDs 6 et 7
Laurent Jégou
jegou@univ-tlse2.fr

Objectifs de la séance : pratique de la cartographie

- Découvrir l'utilisation de Magrit : fonds, données.
- Pratiquer la représentation des données quantitatives discrètes en symboles proportionnels.
- Comprendre l'intérêt de réduire l'information.
- Connaître les différentes méthodes, leurs domaines d'application et leurs conséquences sur la carte produite.

Préparation : rassembler les fichiers utiles

Nous allons utiliser (sur IRIS) :

- un fond de carte des communes de l'Ariège au format GeoJSON, qui provient du produit « AdminExpress » de l'IGN (contours administratifs, libres et gratuits)

→ **ariege_coms.geojson**

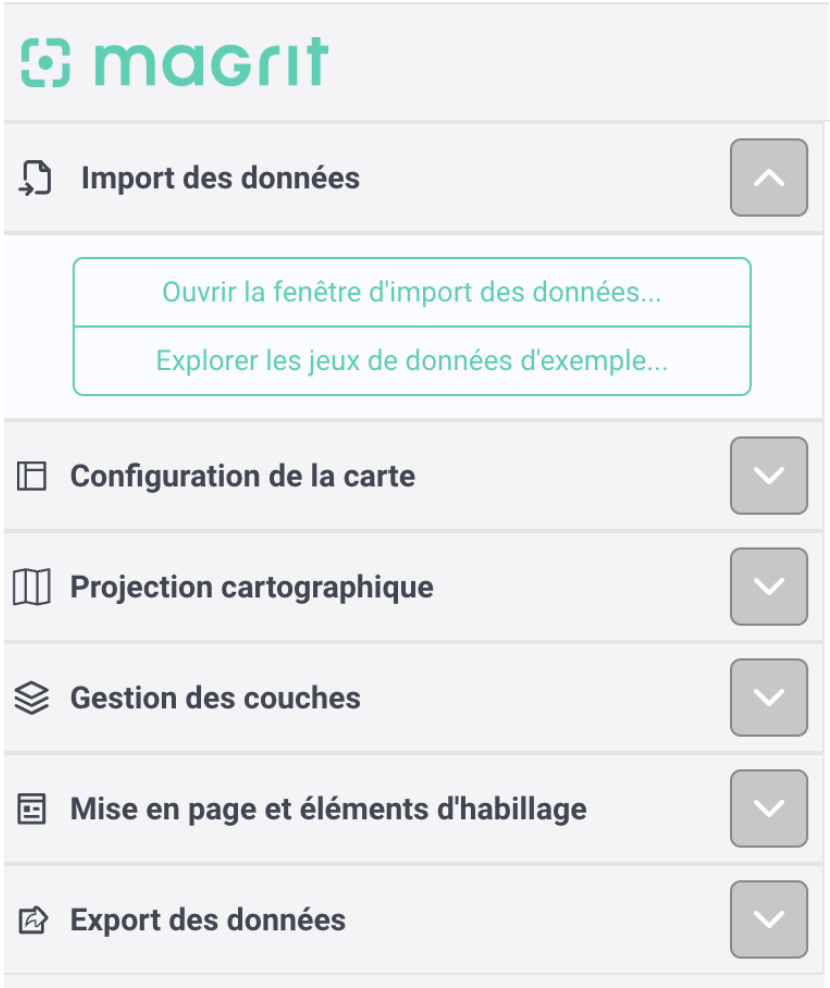
- les données extraites du site de l'INSEE dans la première partie sur les équipements touristiques.

→ **equip-tour-transp-com-2021-ariege.xlsx**

Première étape : ouvrir Magrit dans un navigateur

http://magrit.cnrs.fr/

Fonctionnement de Magrit



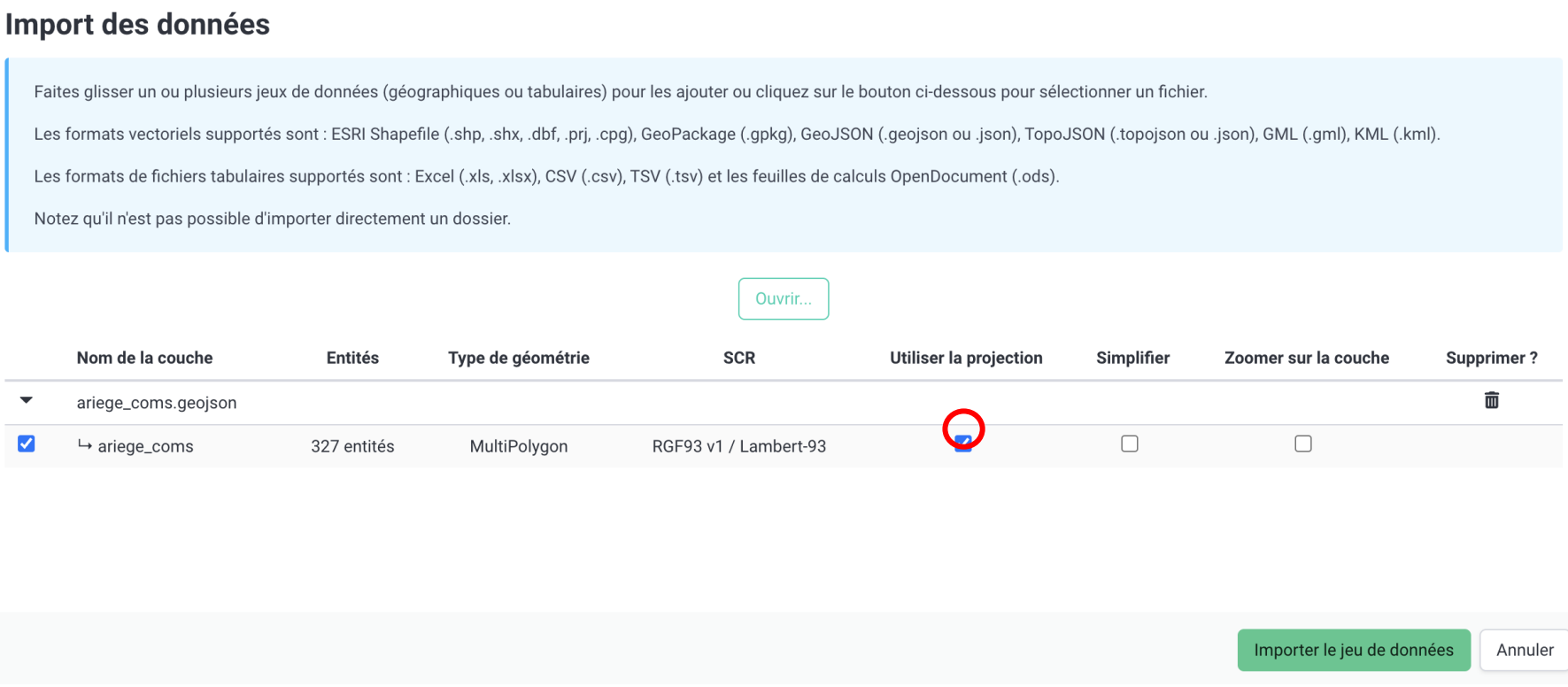
L'utilisation de Magrit est structurée par les étapes figurées dans les cases de la colonne de gauche, de haut en bas.

- Import des données : fonds et statistiques
- Configuration de la carte
- Projection (au besoin)
- Gestion des couches superposées (fond, cartes)
- Mise en page et habillage (titre, échelle, sources…)
- Export de la carte et des données

Ouverture du fond de carte

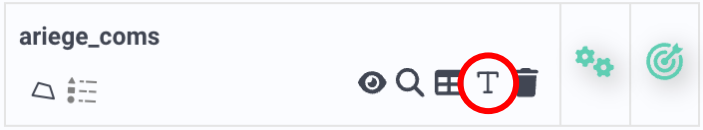
Tout d'abord, on va ajouter le fond ***ariege_coms.geojson*** (sur IRIS)
Il s'agit du fond de carte principal, qui va supporter les cartes thématiques.

On peut aussi utiliser des fonds secondaires, dits d'habillage, qui vont permettre de mieux comprendre, par exemple des routes, des rivières...
Cochez la case « Utiliser la projection »

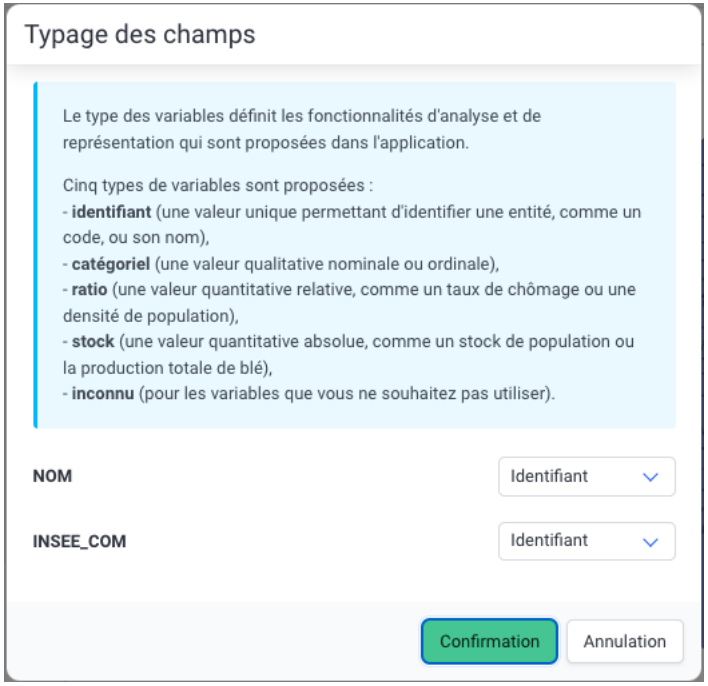


Ouverture du fond de carte

Le fichier fond contient deux *champs* (colonnes) d'information en plus de la géométrie des communes : leurs codes INSEE et leurs noms.



On obtient la fenêtre suivante, qui permet de vérifier / corriger les types :



Ouverture du fond de carte

Le fichier fond contient deux *champs* (colonnes) d'information en plus de la géométrie des communes : leurs codes INSEE et leurs noms.

Magrit possède un vocabulaire un peu différent pour les types de données :

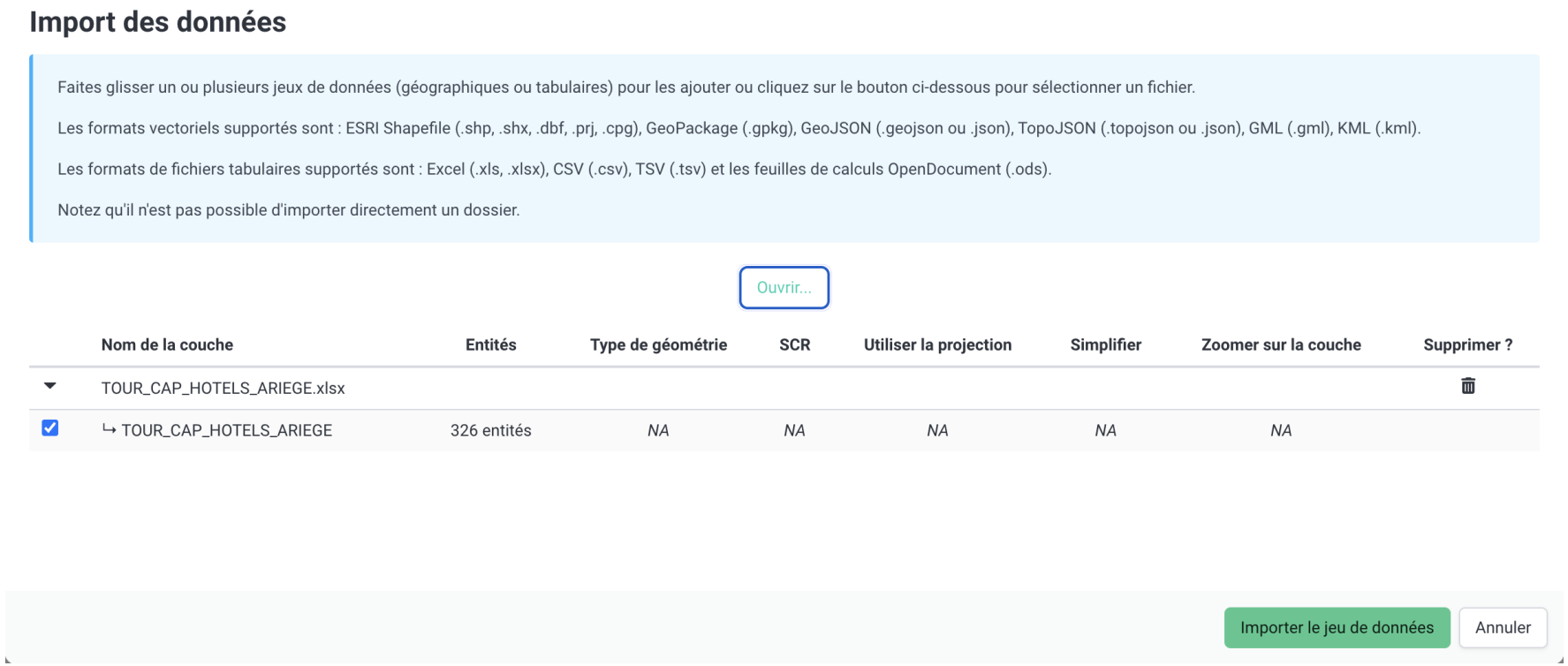
Nom statistique	**Nom dans Magrit**
Identifiant	Identifiant
Quantitative discrète	Stock
Quantitative continue, relative	Ratio
Qualitative ordonnée	Catégoriel
Qualitative nominale	Catégoriel

Les deux colonnes du fond de carte, Nom et INSEE_COM sont bien des identifiants, donc rien à corriger à cette étape.

Ouverture du fichier de données xls

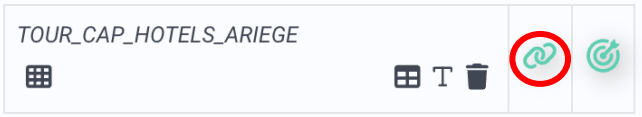
Ensuite, il faut charger les données à représenter sur le fond de carte des communes.

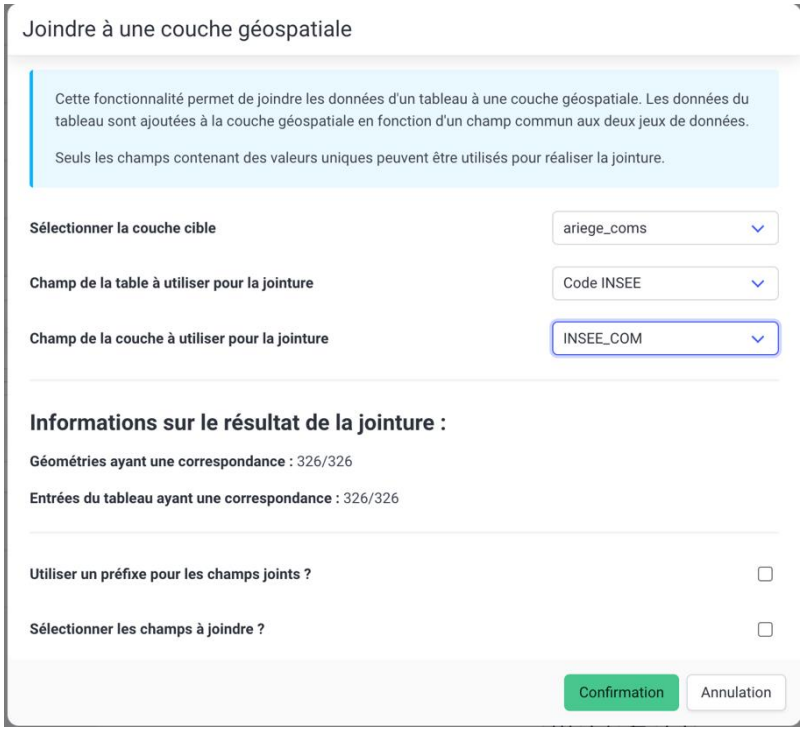
Via l'ajout d'un jeu de données, chargez le fichier :
TOUR_CAP_HOTELS_ARIEGE.xlsx



Ouverture du fichier de données xls

- « Jointure » fond - données



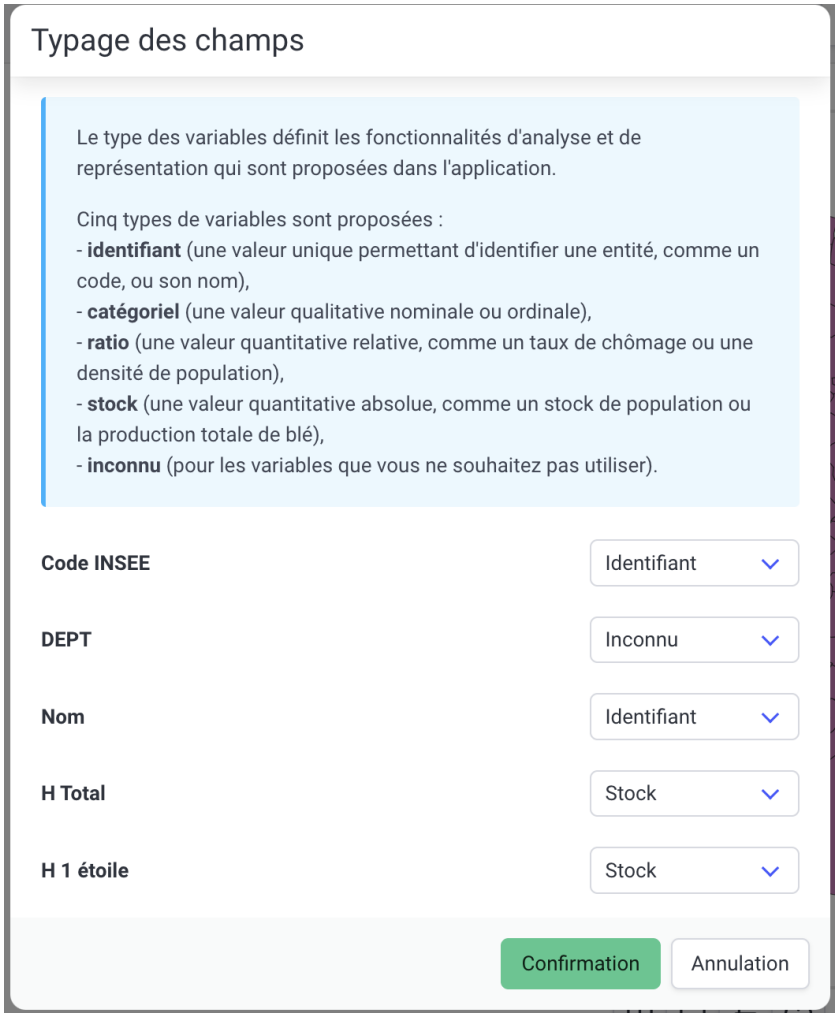


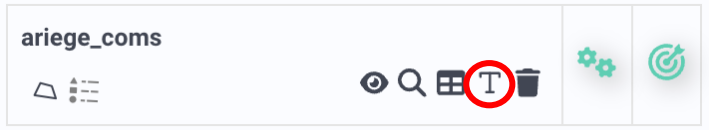
Pour que Magrit sache quelle ligne du fichier de **données** correspond à quel **élément géométrique** du fond de carte, il faut réaliser l'opération de *jointure* (ou mise en relation).

Cela consiste simplement à indiquer quels sont les champs qui contiennent *l'information commune* aux deux fichiers, c'est-à-dire le code géographique INSEE.

Ici, c'est le champ « **INSEE_COM** » du fond de carte qui correspond au champ « **Code INSEE** » du fichier de données Excel.

Typage des données



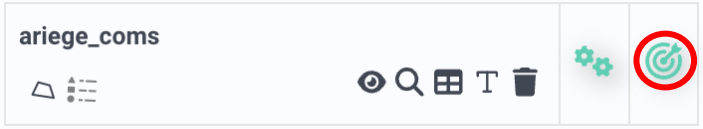


Pour terminer la préparation des données, il faut typer les nouvelles colonnes jointes au fond de carte.

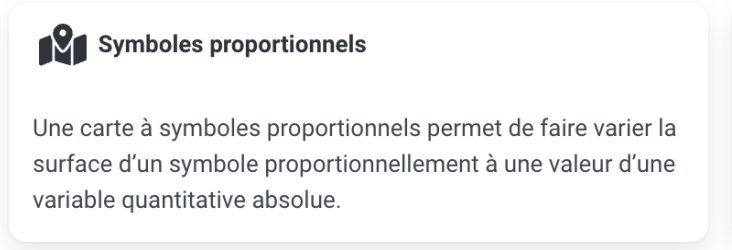
Hôtel Total et 1 étoile, qui sont des nombres d'équipements, des comptages absolus, bruts, des *quantitatives discrètes* sont donc à typer en « **Stock** ».

Représentation en symboles

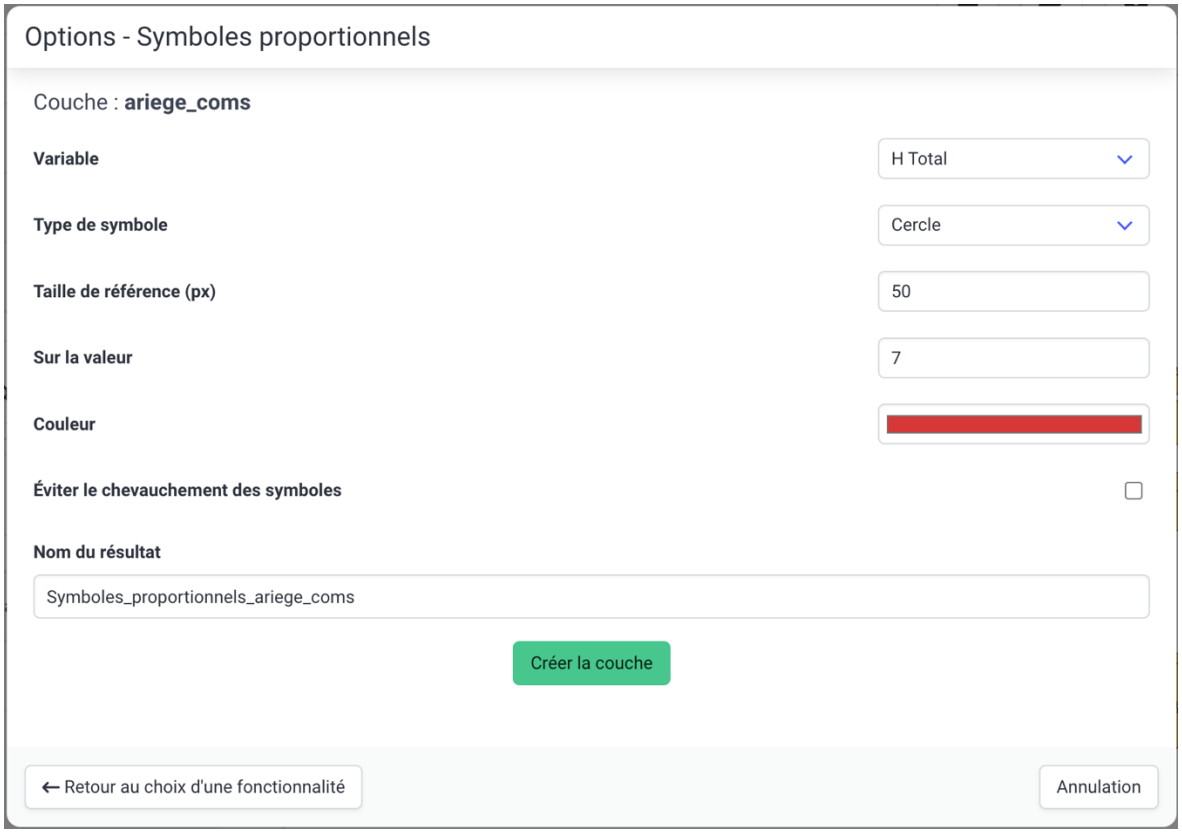
Pour commencer une carte, il faut cliquer sur **l'icône en forme de cible** de la couche du fond de carte :



Ensuite, on choisit le type de carte en symboles proportionnels :



Représentation en symboles

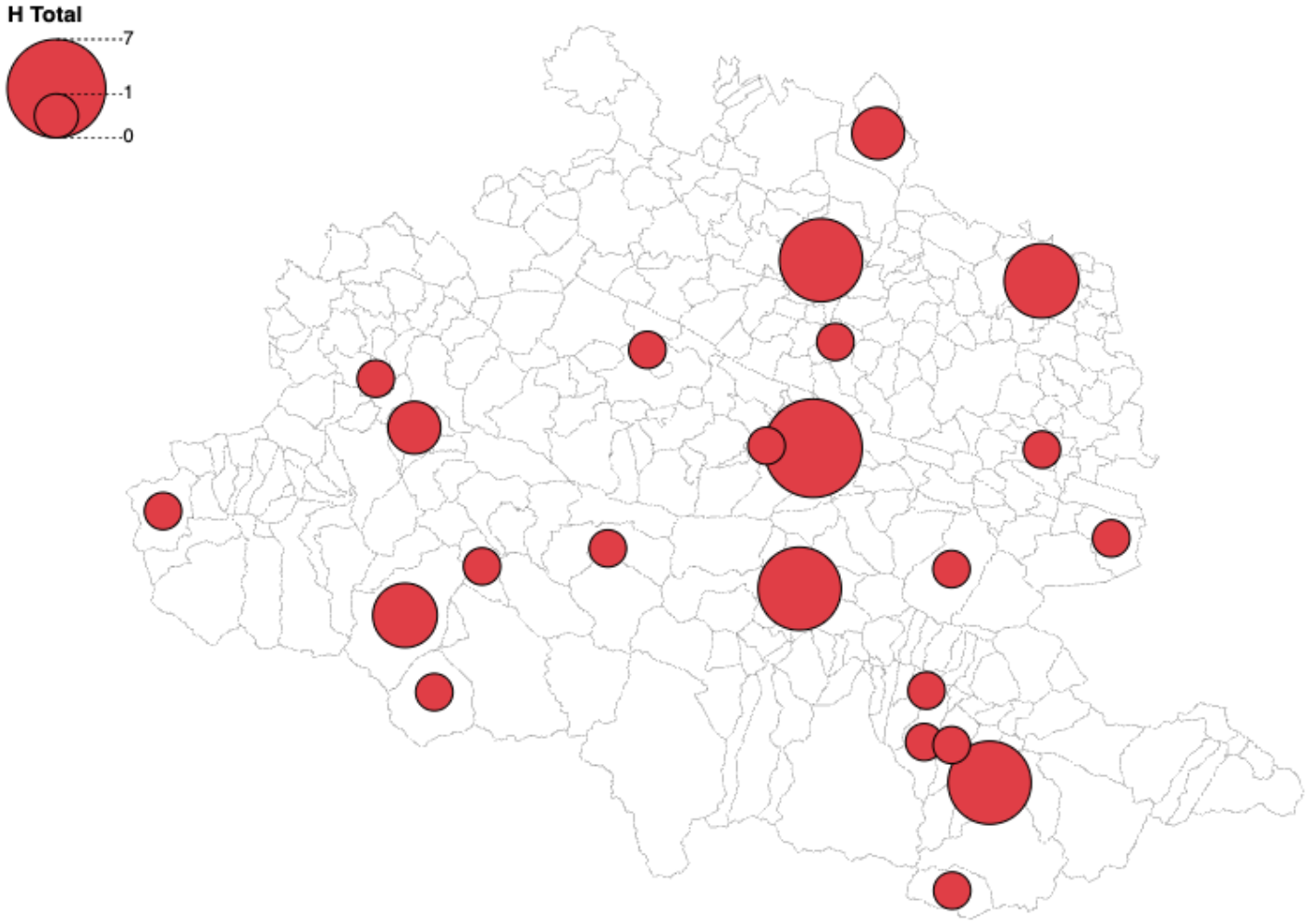


- choisir un **champ**,
- définir la **taille** du symbole à dessiner pour une valeur de la variable.

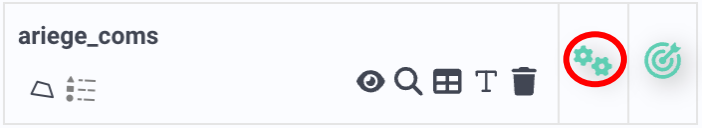
Tous les autres symboles seront proportionnels à celui-ci.

Généralement on définit ce ratio taille-valeur sur la valeur maximale (choix par défaut dans Magrit), le plus gros symbole étant le plus délicat à dessiner.

La carte en cours de réalisation



Réglages visuels



Pour que notre carte soit d'une présentation lisible, on peut choisir un fond blanc et un contour des polygones en gris moyen.

Une règle de graphisme importante, la **hiérarchie visuelle** : l'information importante doit être bien visible, les informations secondaires doivent passer au second plan visuel.

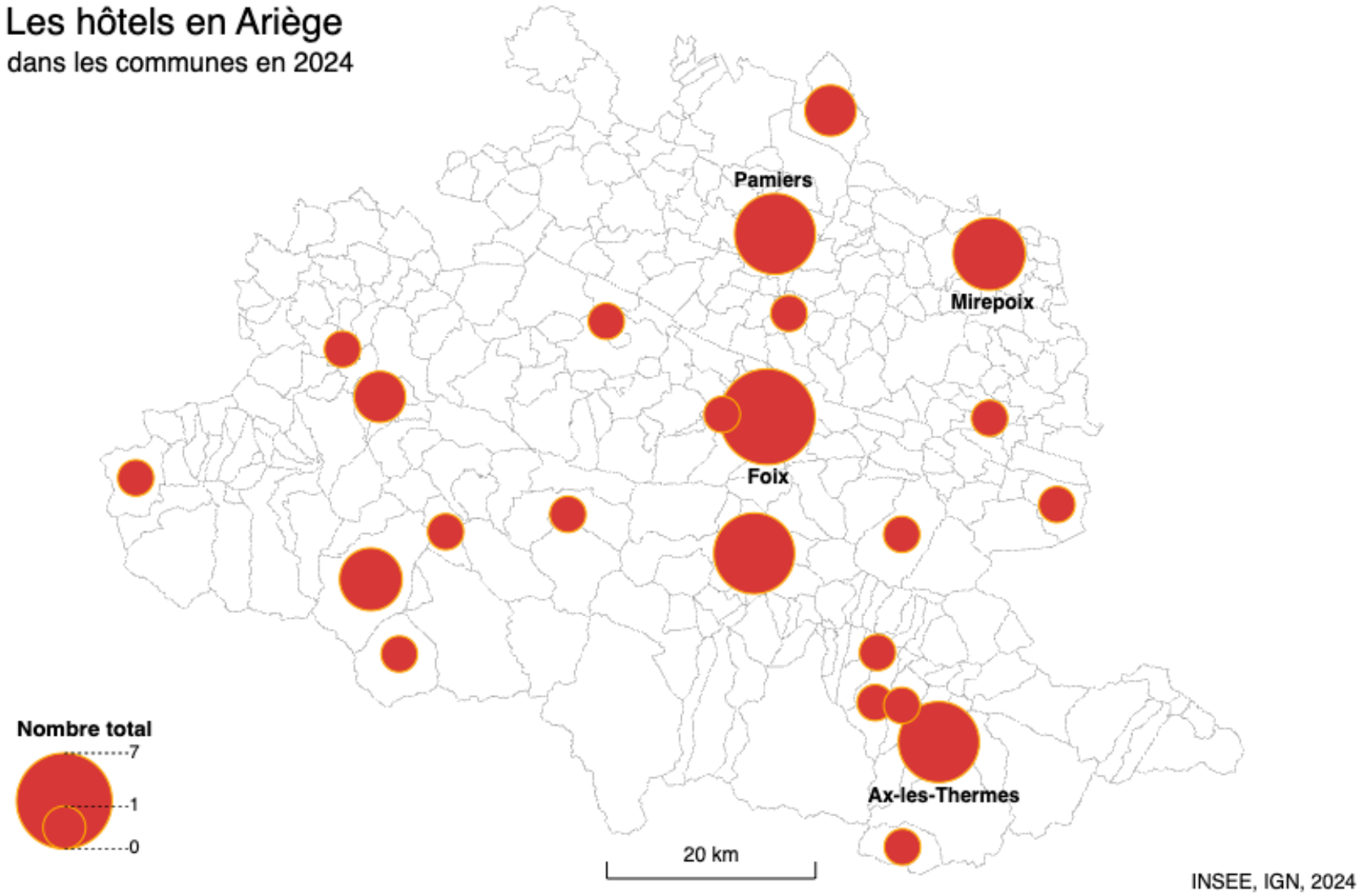
Habillage de la carte

Cette carte n'est pas terminée, car il lui manque des éléments importants pour être bien comprise :

- Un titre *factuel*, qui répond aux questions : *quoi*, *où* et *quand*
- Un titre de légende qui n'est pas redondant avec le titre principal et aide à lire la carte en indiquant *les unités* utilisées
- Une échelle graphique (une barre graduée)
- Des sources (INSEE pour les stats. et IGN pour le fond)
- Un.e auteur.e (et éventuellement une date de réalisation)
- L'orientation vers le Nord (optionnelle si le nord est vers le haut)

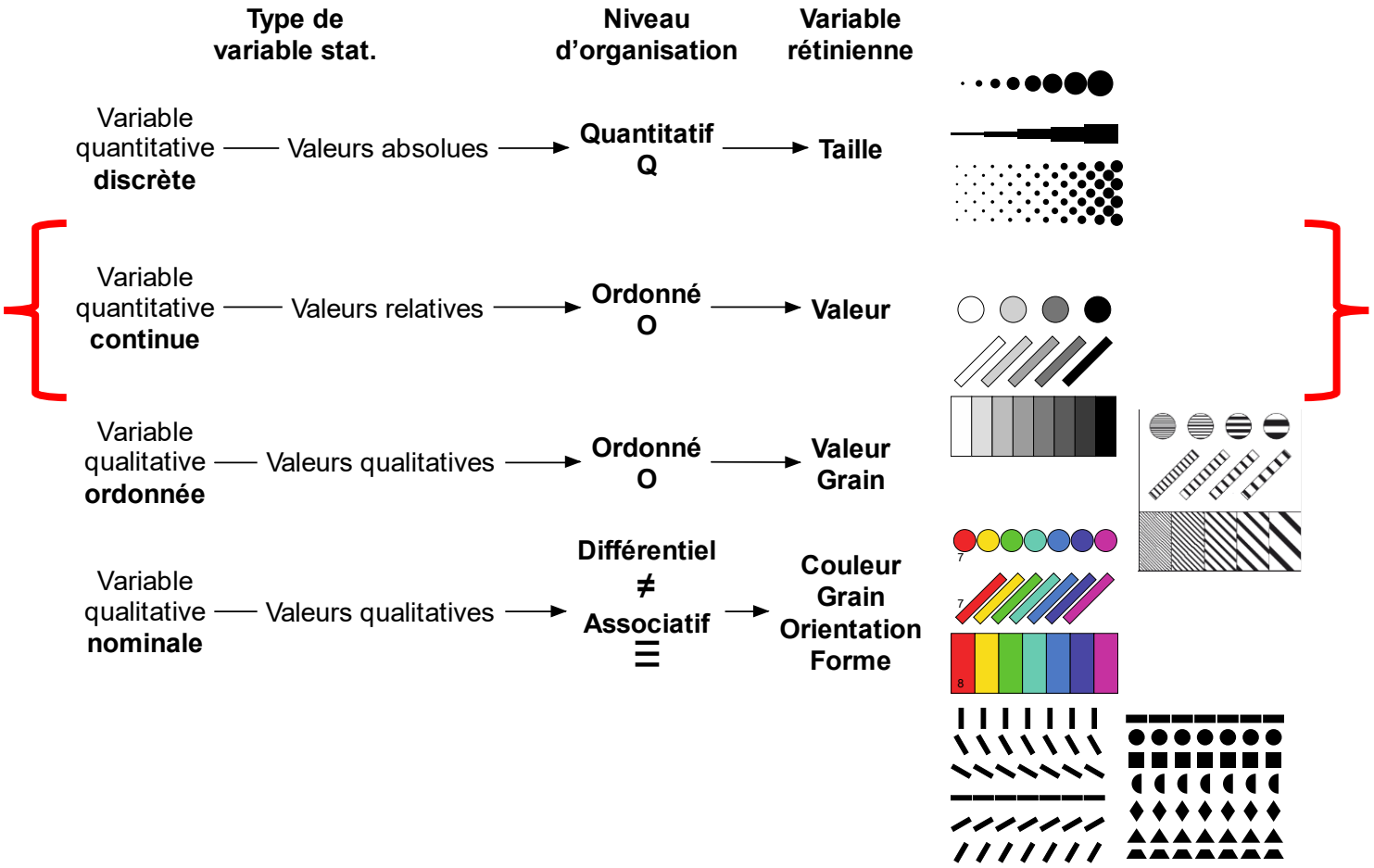
➜ Magrit permet d'ajouter ces éléments dans la case suivante, « Mise en page et éléments d'habillage».

Avec la possibilité d'ajouter des libellés de texte, on peut nommer les principales villes possédant des hôtels sur la carte.

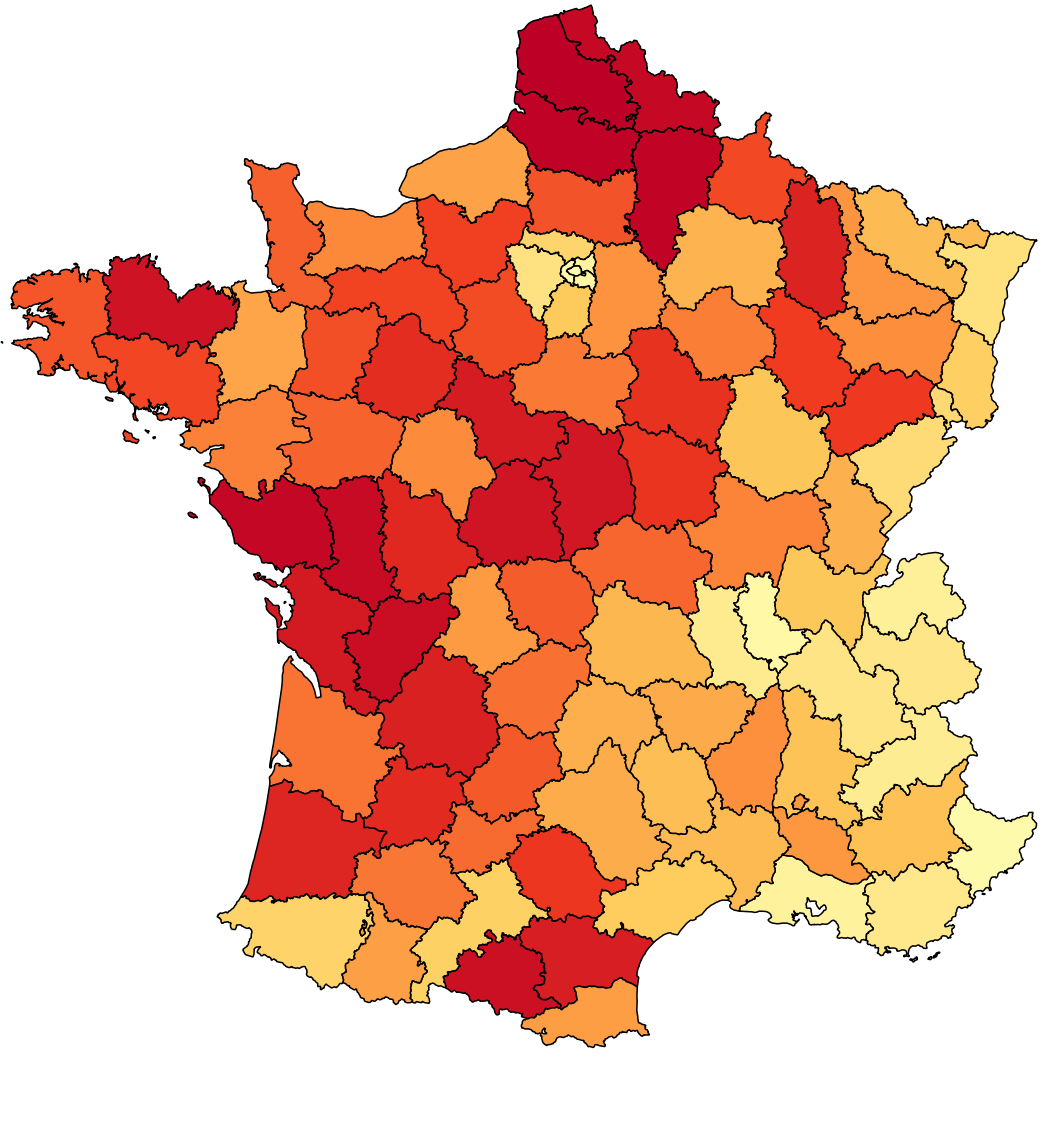


Vers la représentation des variables relatives, quanti. continues, en couleurs

Pour ce type de variable, c'est l'ordre qu'il est important de représenter comme niveau d'organisation entre les individus, avant les quantités.



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



En pourcentages

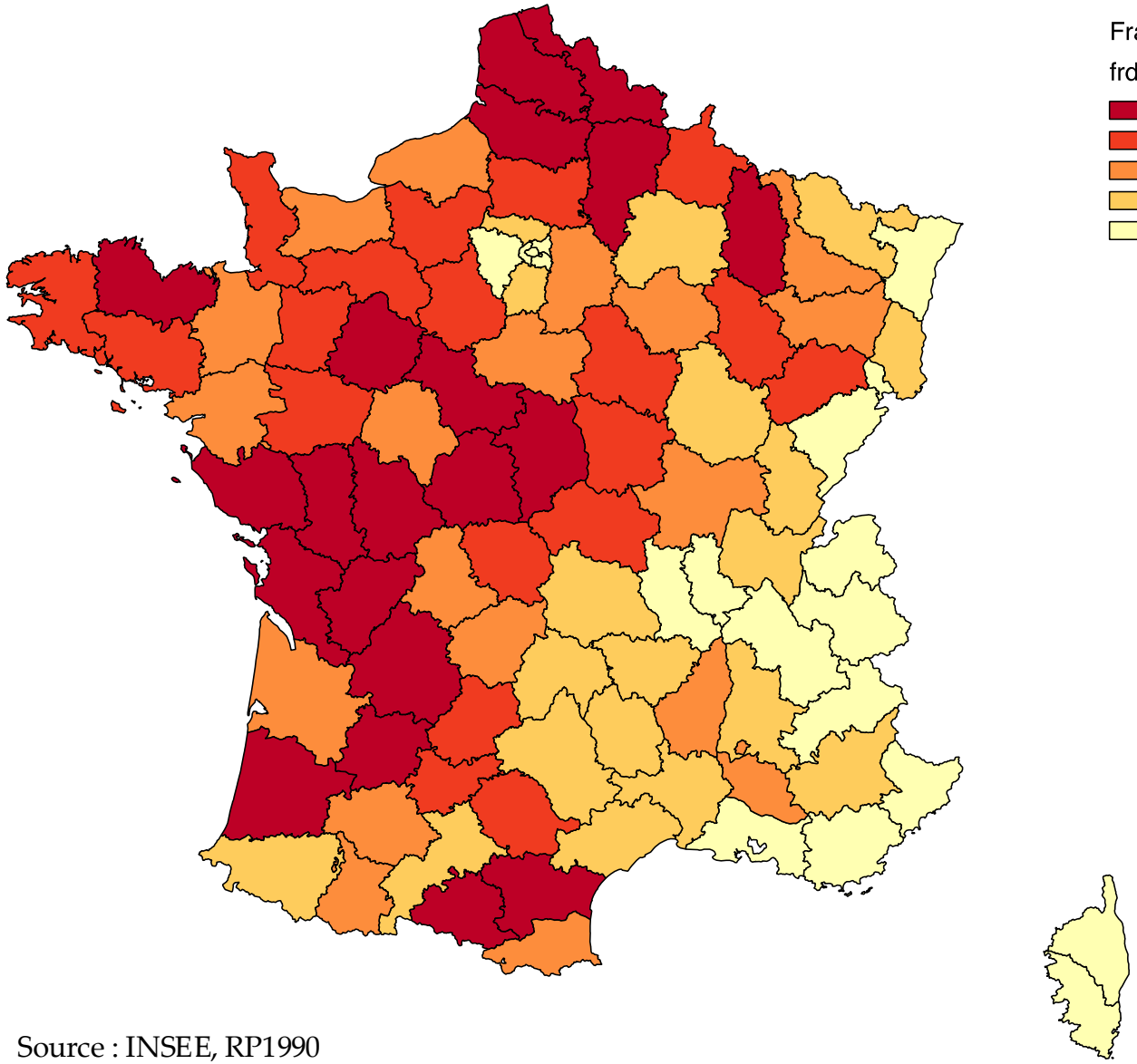
France data départements

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	

Source : INSEE, RP1990

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



Source : INSEE, RP1990

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.

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

9 classes s'il oppose deux teintes :

Les logiciels vous proposeront des palettes de couleurs automatiques, mais vous pouvez aussi définir vos propres paliers de gradients colorés, par exemple avec cet outil : http://www.geotests.net/couleurs/gradients.html

Rappel des principes

- Objectif : simplifier la carte pour rendre l'information plus lisible.
- Préserver les éléments importants de la variable : les seuils
  (zéro pour +/-, indice 100, etc.), le sens de variation.
- Respecter la forme de la **distribution** des valeurs
  (celle du *graphique de fréquences*)
- Créer des classes qui correspondent à des
  « groupes pseudo-naturels » de valeurs (sous-ensembles qui se distinguent du reste, dans la distribution des valeurs).

Concrètement :

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

L'image colorée que représente la carte doit donner une idée *correcte et pertinente* 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 manques entre deux classes successives).
- Les classes ne doivent pas se chevaucher (un individu ne peut appartenir qu'à une classe à la fois).
- Les couleurs choisies doivent évoquer le phénomène.

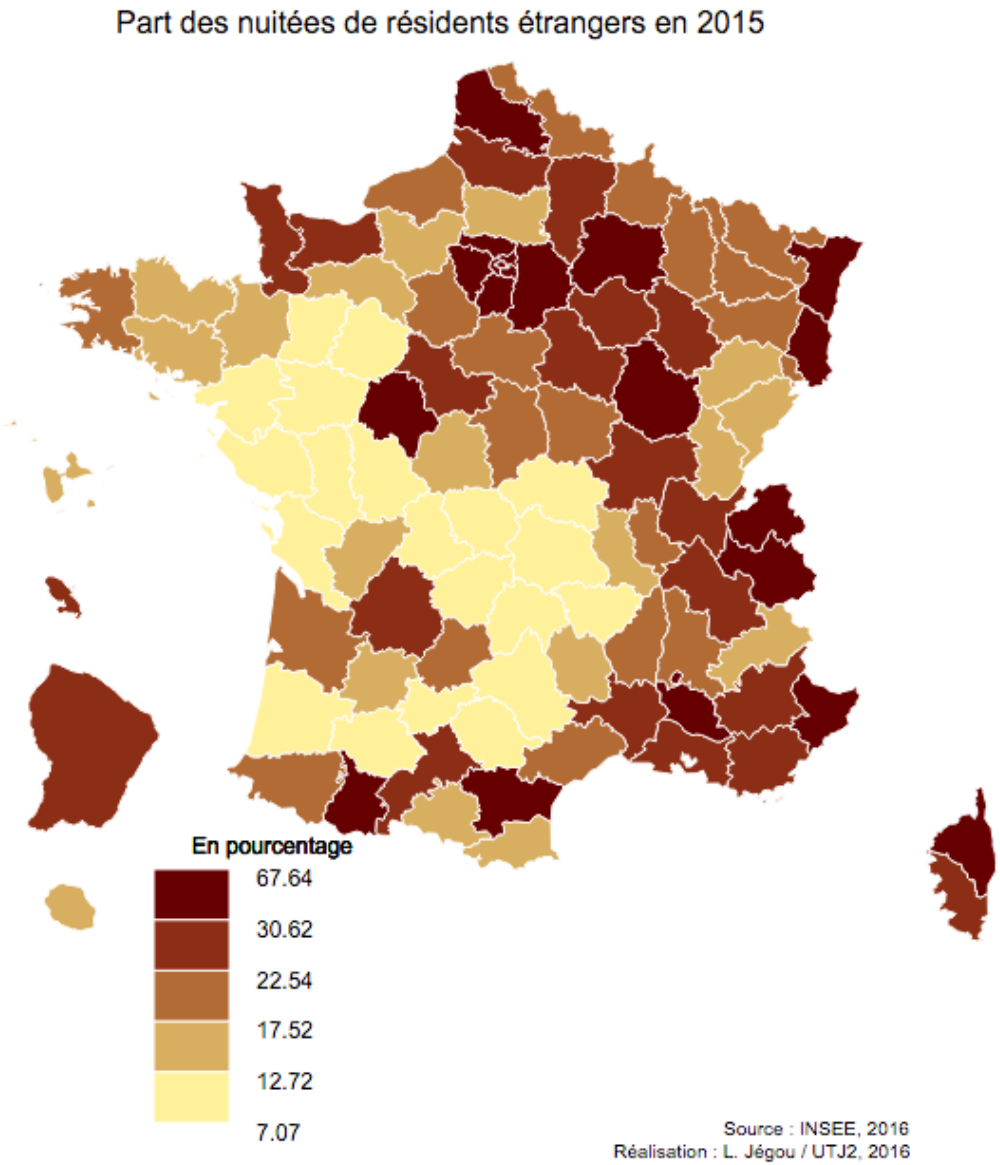
Les différentes méthodes de discrétisation automatique

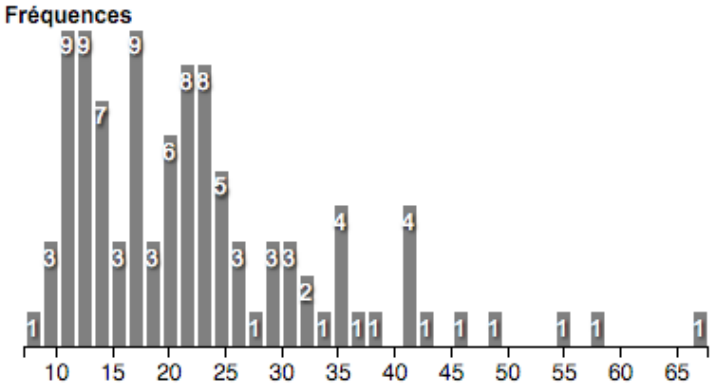
- Selon le logiciel de conception cartographique utilisé, différentes méthodes de regroupement en classes seront disponibles, parfois sous des noms différents.

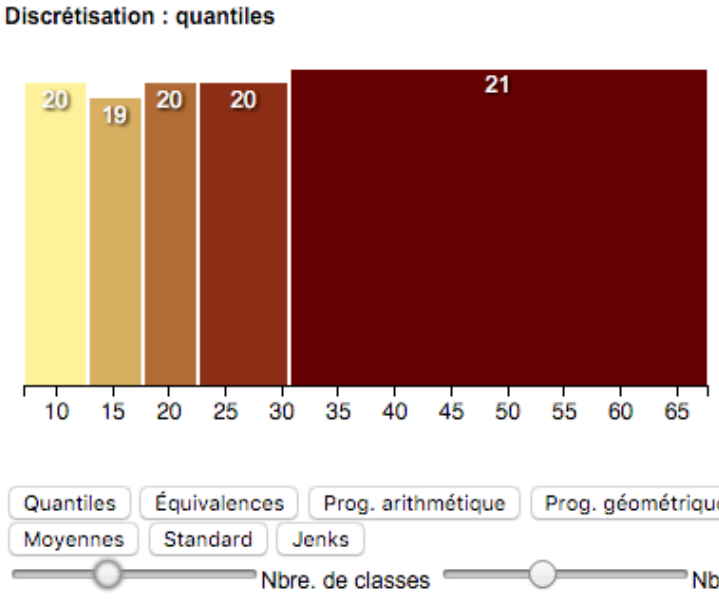
- Les illustrations sont issues de cet exemple interactif en ligne : http://www.geotests.net/test/discretisations/
- Les données sont disponibles au format Excel ici : http://www.geotests.net/test/discretisations/hotels2015etr.xls

La méthode des quantiles
(ou équifréquences)

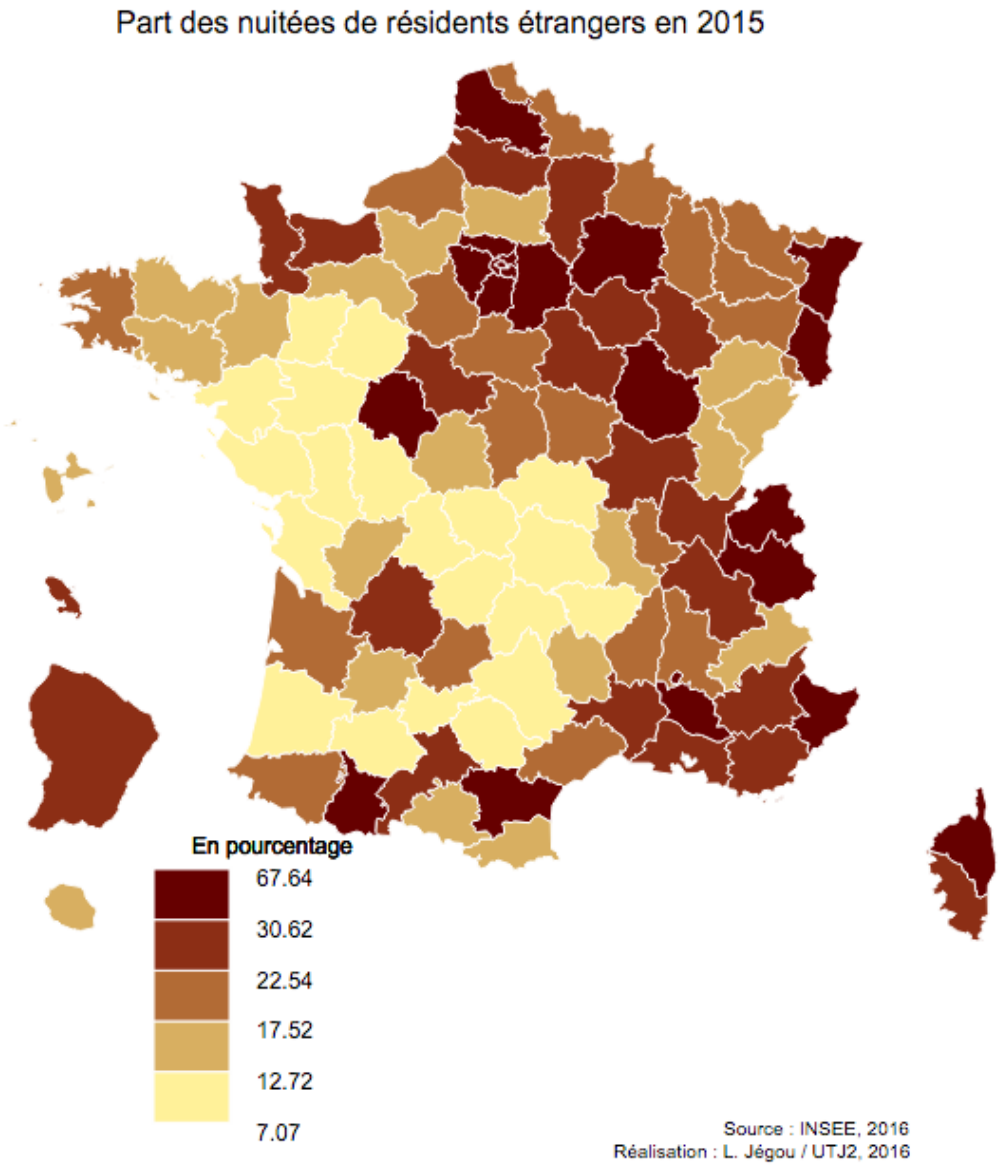
- **Méthode :** On compte le même nombre d'individus par classe
- **Calcul** : On divise la population par le nombre de classes désiré, on obtient le nombre d'individus par classe. Ensuite, on détermine les bornes des classes en triant la variable dans l'ordre croissant et en cherchant les bornes correspondant 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 de 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.*

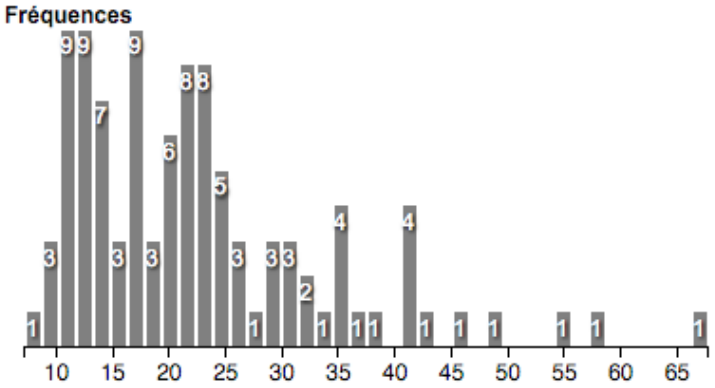


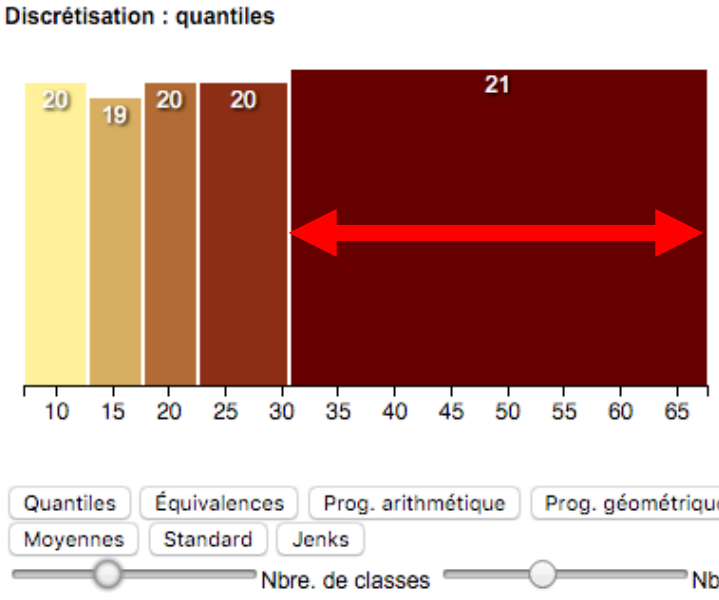




La méthode des quantiles : exemple



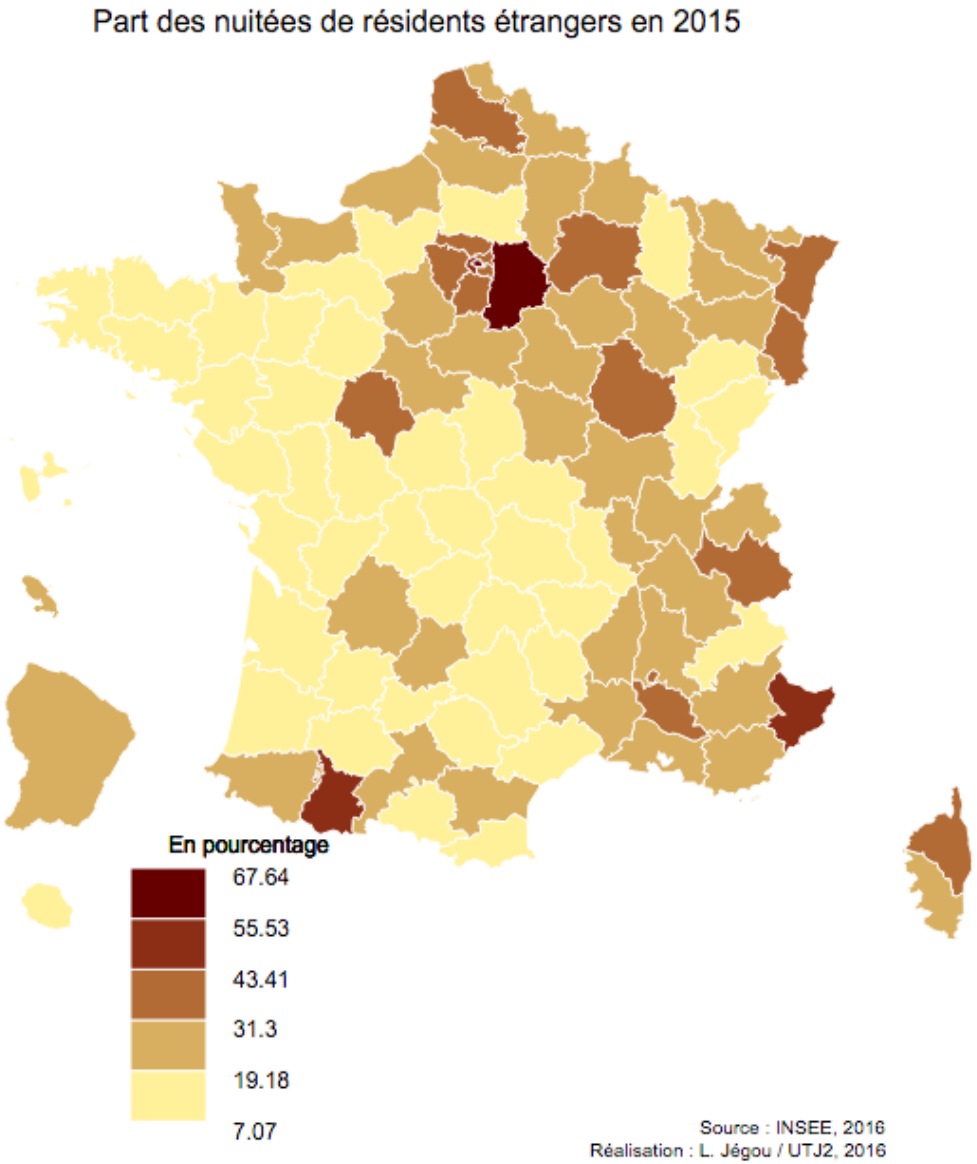


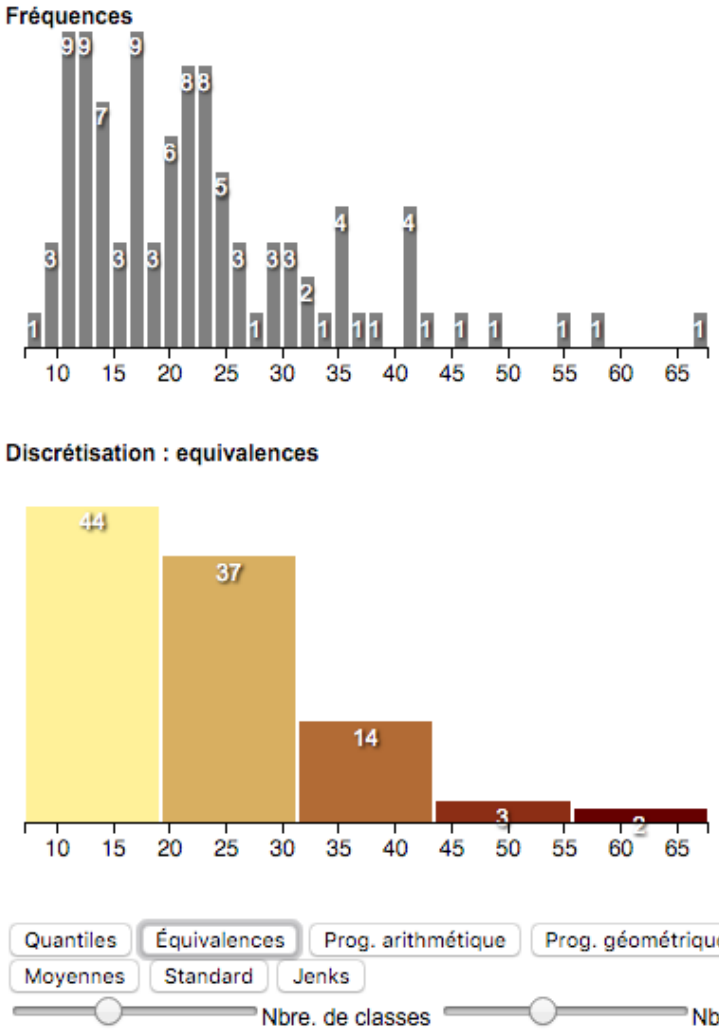


La méthode des quantiles : exemple

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 à tendance à produire des classes de poids très inégal, et donc des cartes avec une répartition des couleurs relativement hétérogène.





La méthode des équivalences : exemple

La méthode de la progression arithmétique

- La méthode est conçue pour les *distributions asymétriques* à gauche, c'est-à-dire 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.

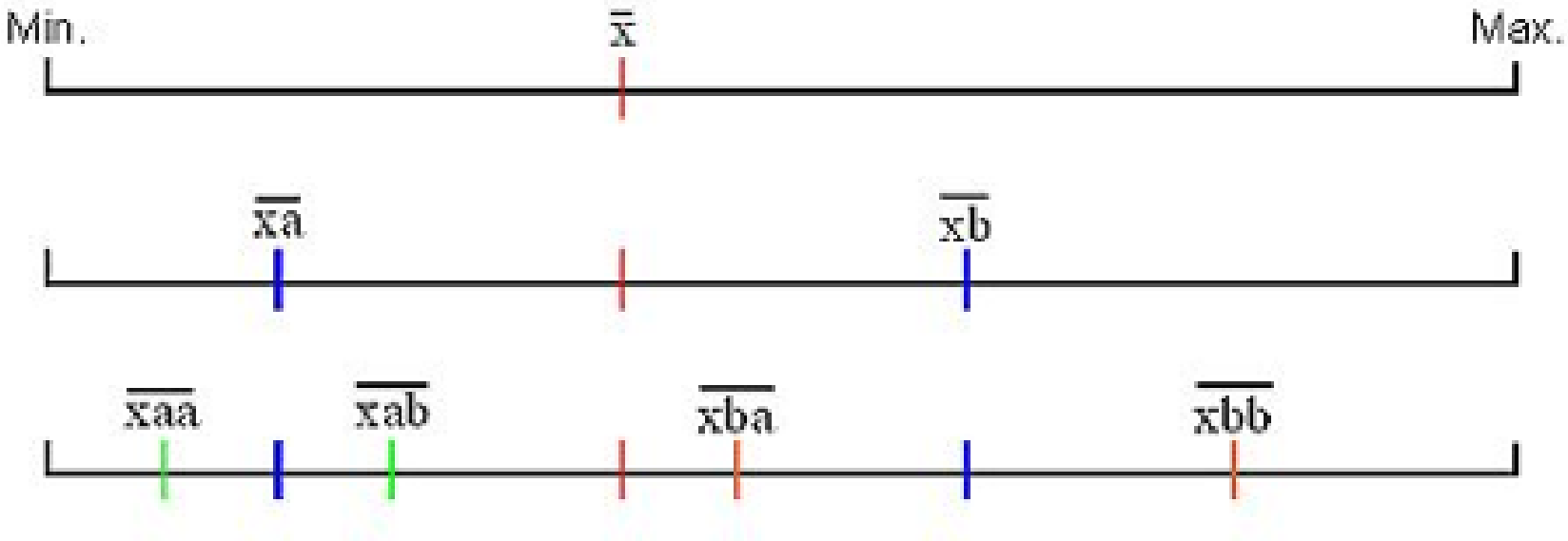
La méthode des moyennes emboîtées

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 MA de la partie A de la variable. MA découpe la partie A en deux classes : AA et AB.
3. Calculer la moyenne MB de la partie B de la variable. MB découpe la partie B en deux classes : BA et BB.

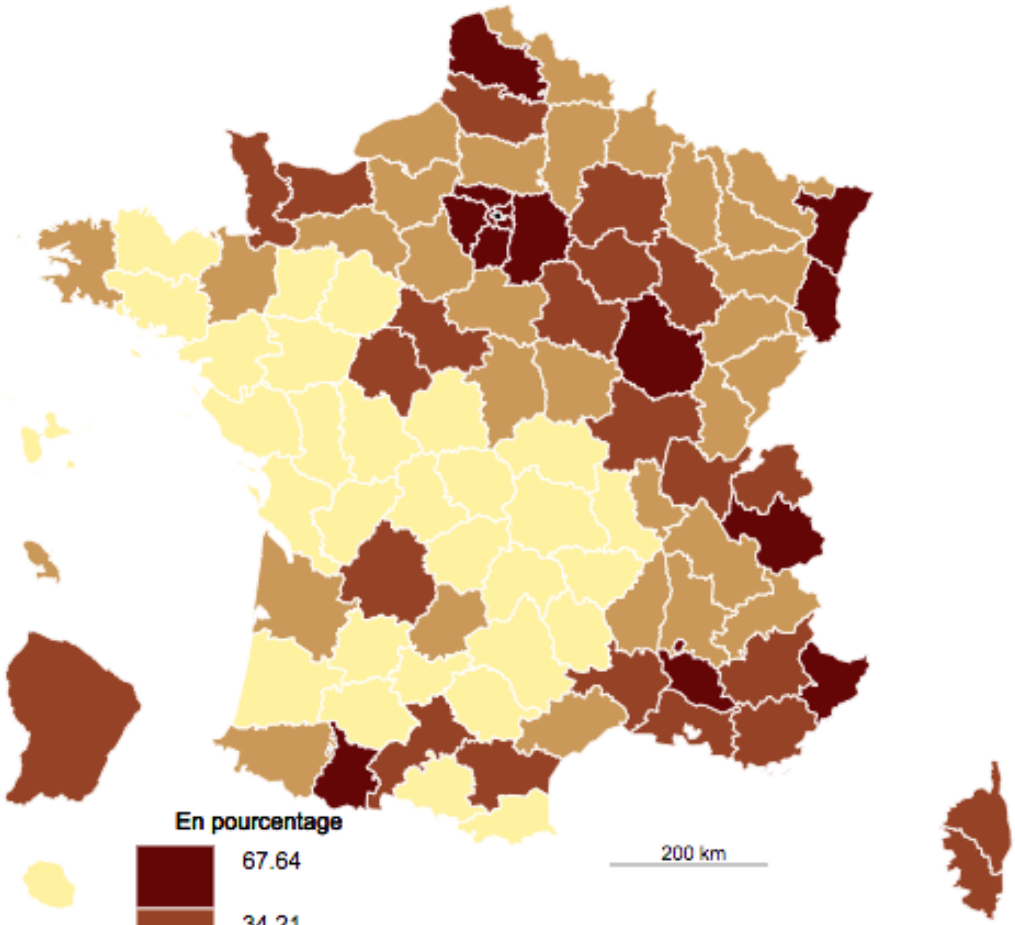
On obtient donc quatre classes, AA, AB, BA et BB, en ayant calculé trois moyennes emboîtées. Si l'on poursuit le processus encore une fois, on obtient 8 classes.



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.

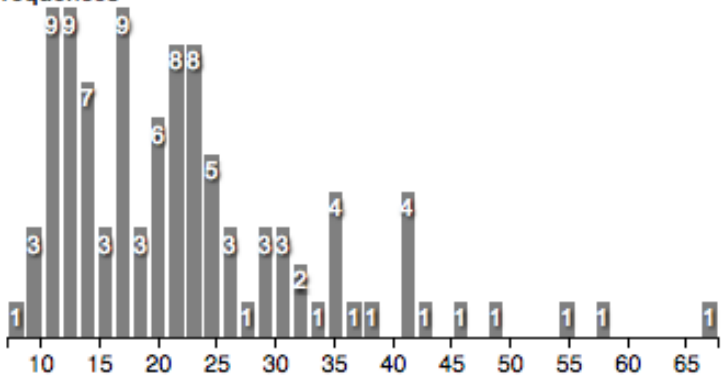
Inconvénients : De par sa méthode de calcul, cette discrétisation peut produire des classes vides ou très hétérogènes, dans le cas de distributions très dissymétriques ou plurimodales. Enfin, le nombre de classes produit, fixe, peut ne pas être adapté au nombre d'individus de la variable. *Il faut donc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

Part des nuitées de résidents étrangers en 2015

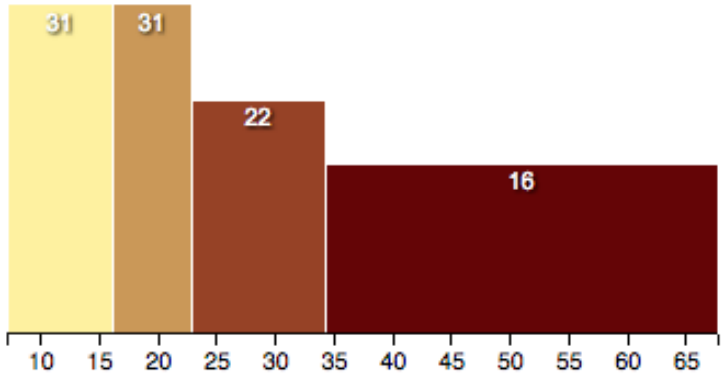


Sources : INSEE, 2016, IGN
Réalisation : M@ppemonde - 12/2016

Fréquences



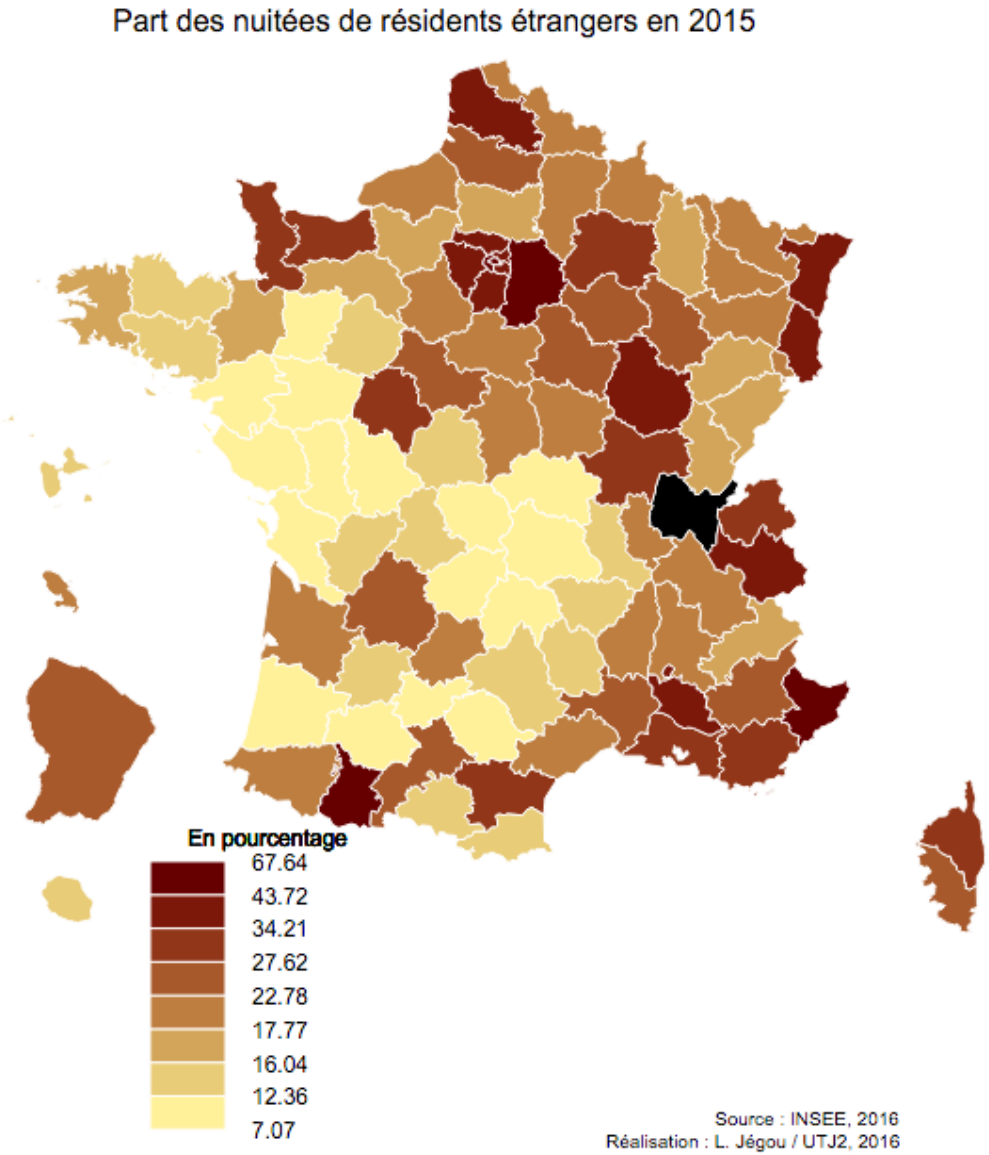
Discrétisation : moyennes

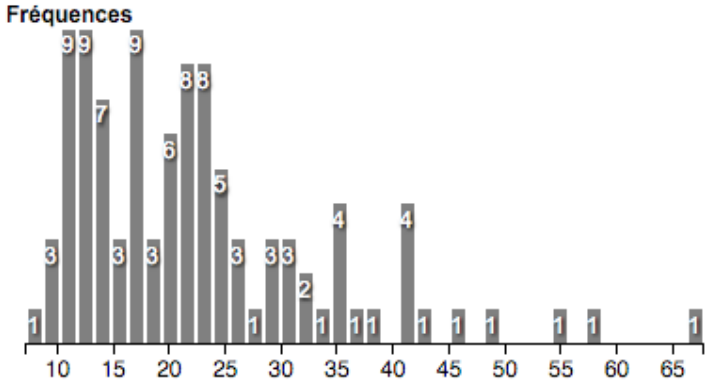


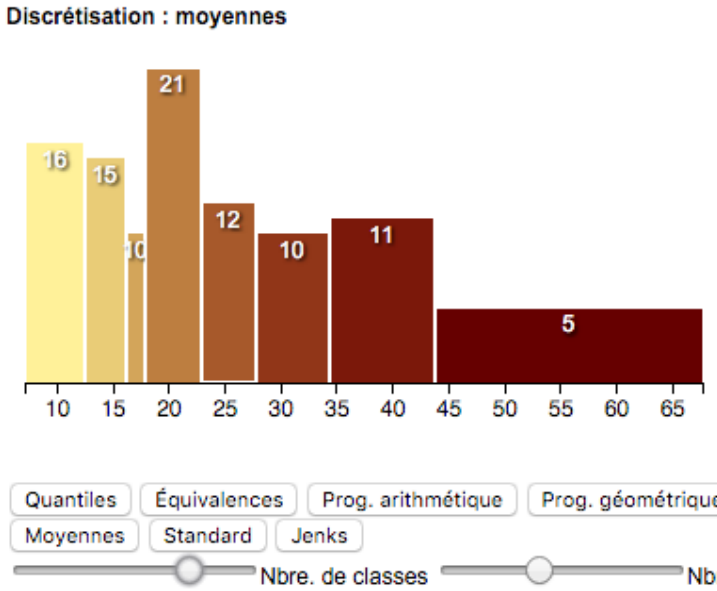
Quantiles | Équivalences | Prog. arithmétique | Prog. géométrique
Moyennes | Standard | Jenks

Nbre. de classes

Nbre. de barres de fréquences







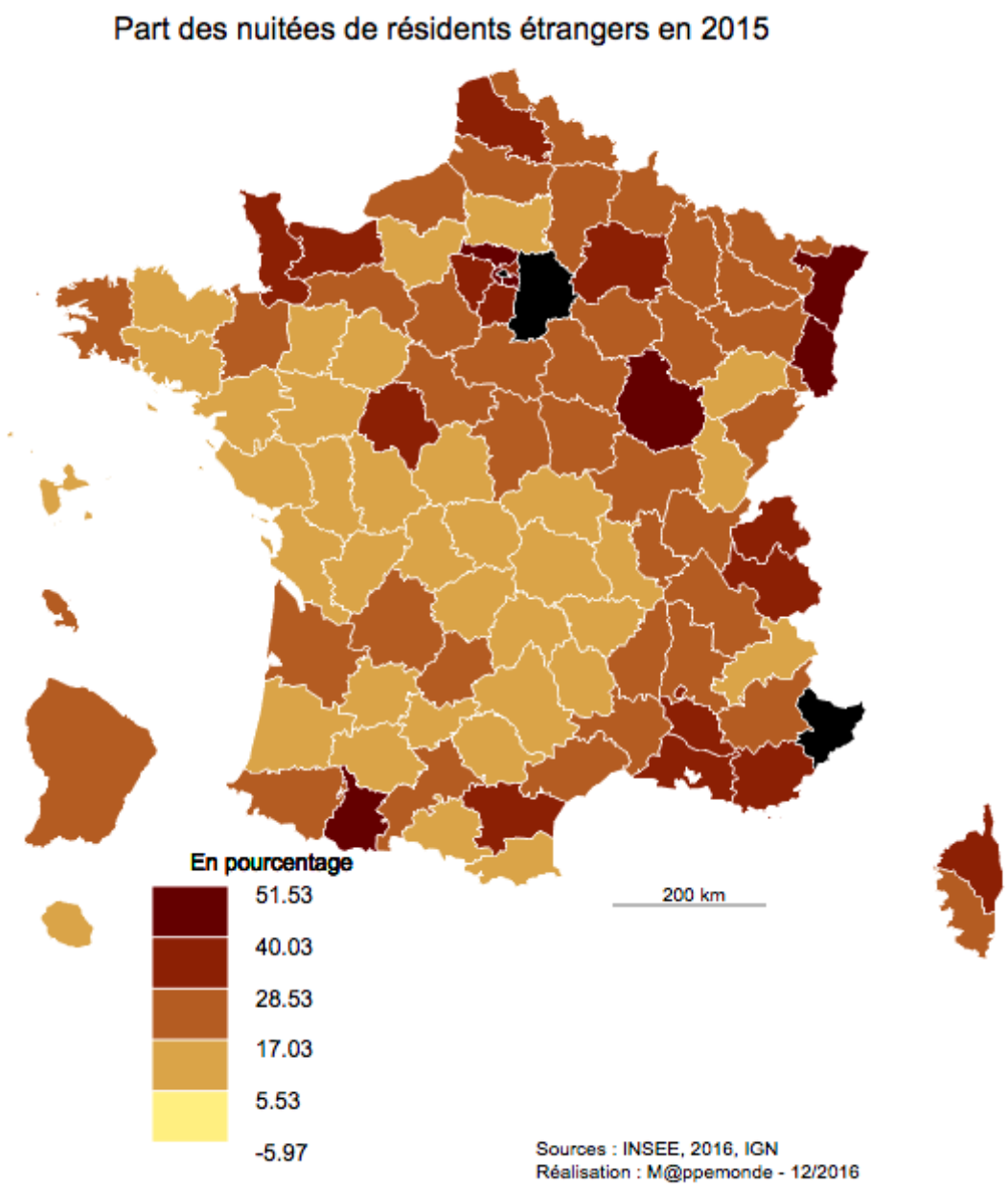
La méthode des moyennes emboîtées : exemple

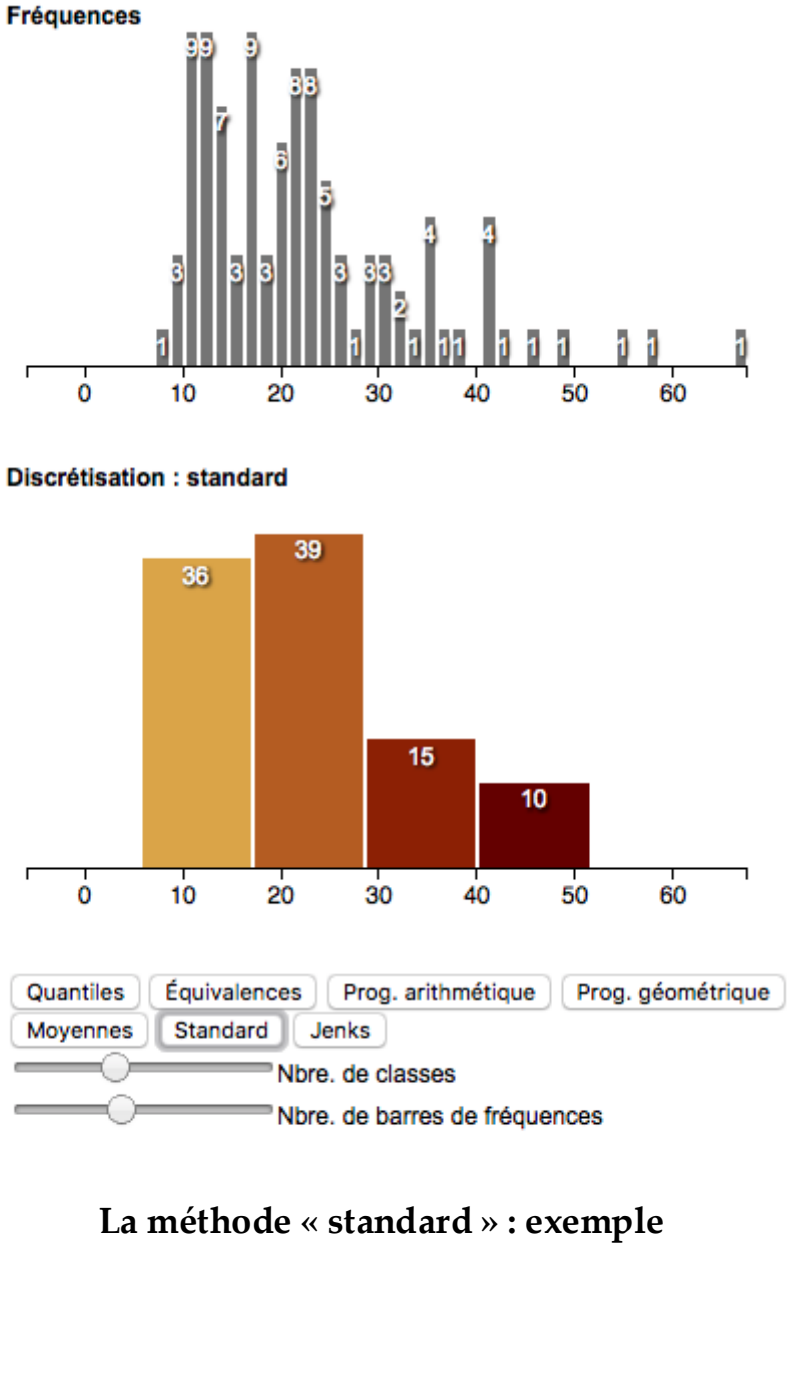
La méthode « standard »
(basée sur l'écart-type)

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.

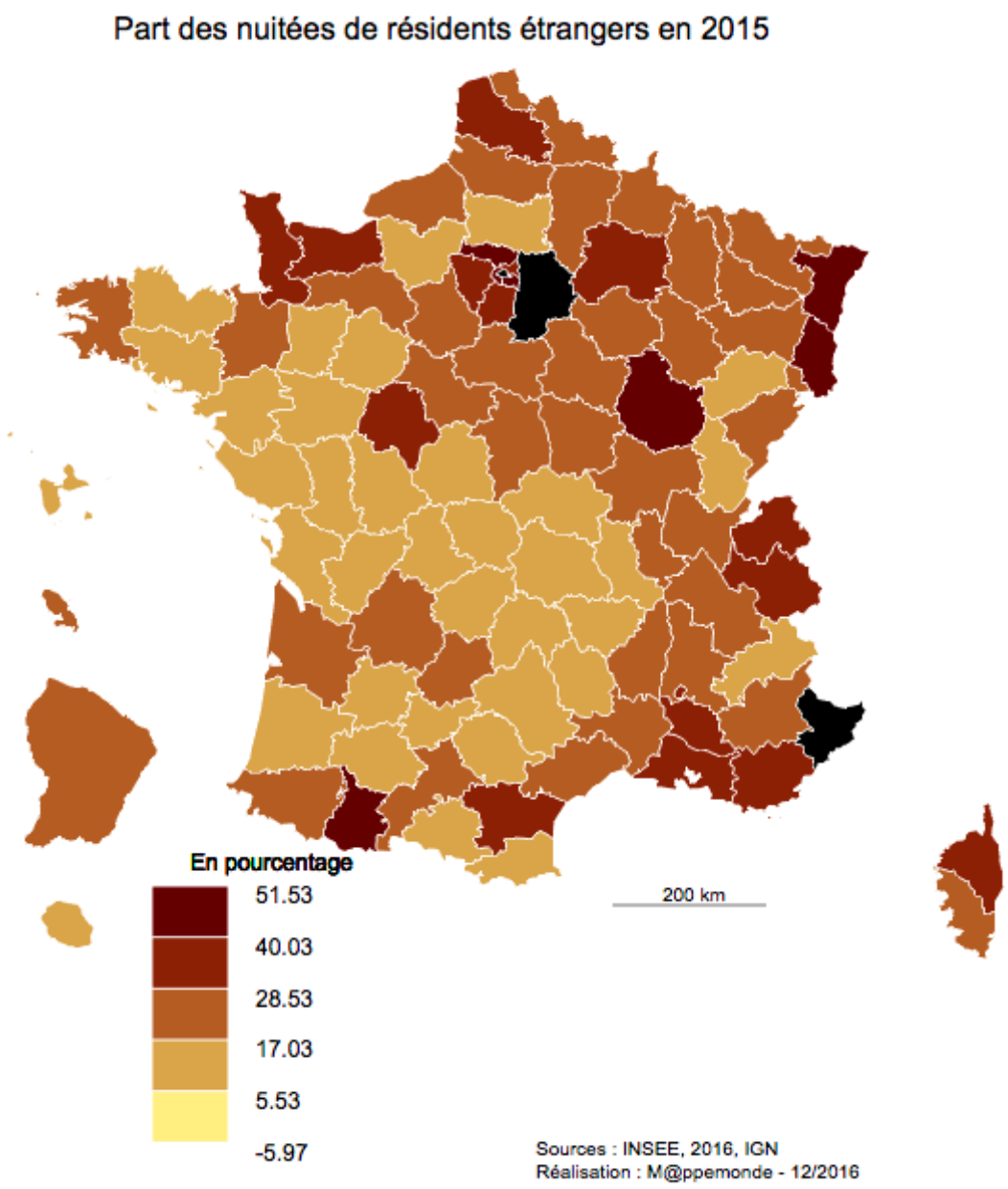
Calcul : On calcule d'abord la moyenne M de la variable, puis son écart-type E. Ensuite, suivant le nombre de classes voulu, on applique la procédure suivante :

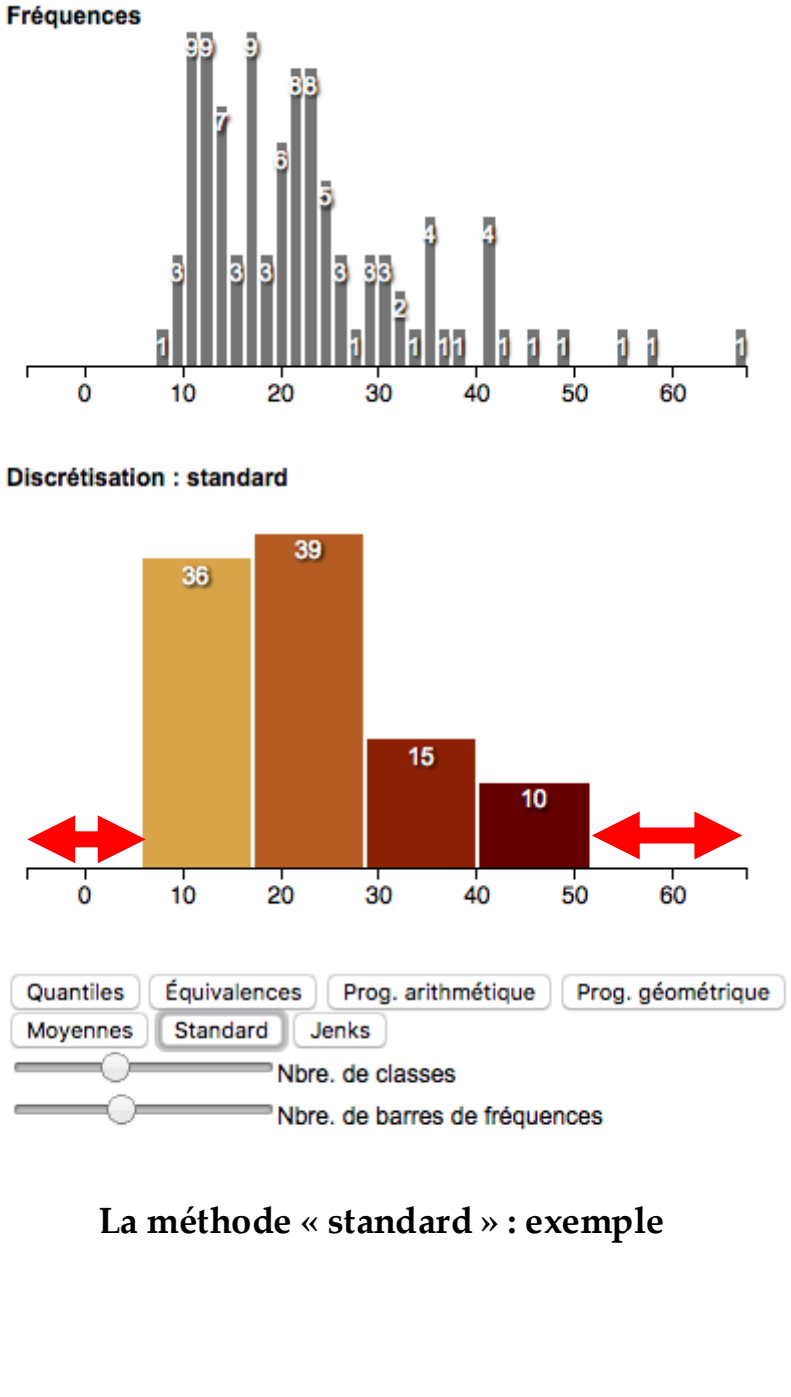
- 3 classes:
  Min, m-0.5E, m+0.5E, Max
- 5 classes:
  Min, m-1.5E, m-0.5E, m+0.5E, m+1.5E, Max
- 7 classes:
  Min, m-2.5E, m-1.5E, m-0.5E, m+0.5E, m+1.5E, m+2.5E, Max





La méthode « standard » : exemple





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. (On peut les commenter facilement).

- **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.

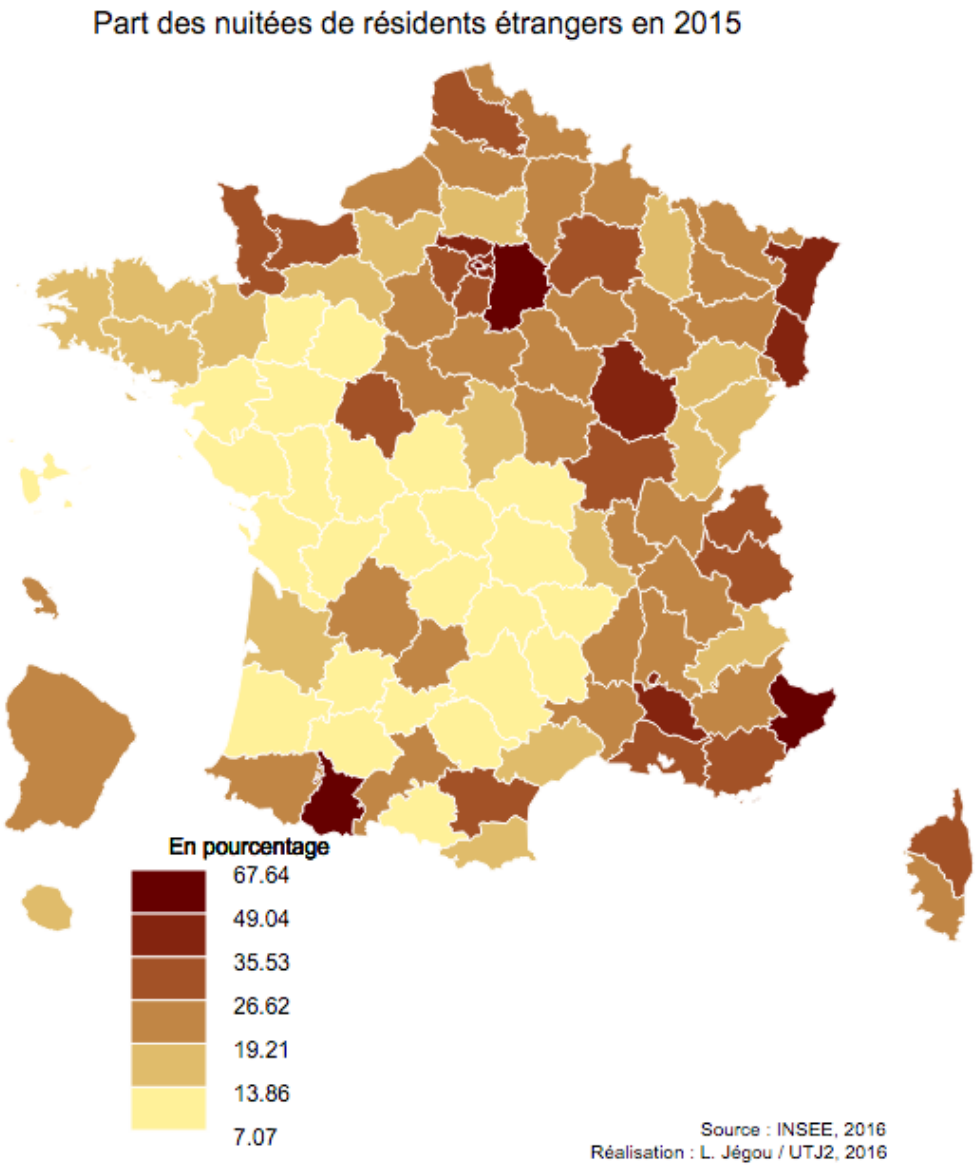
La méthode de Jenks
(ou nuages de points, voire « seuils naturels »)

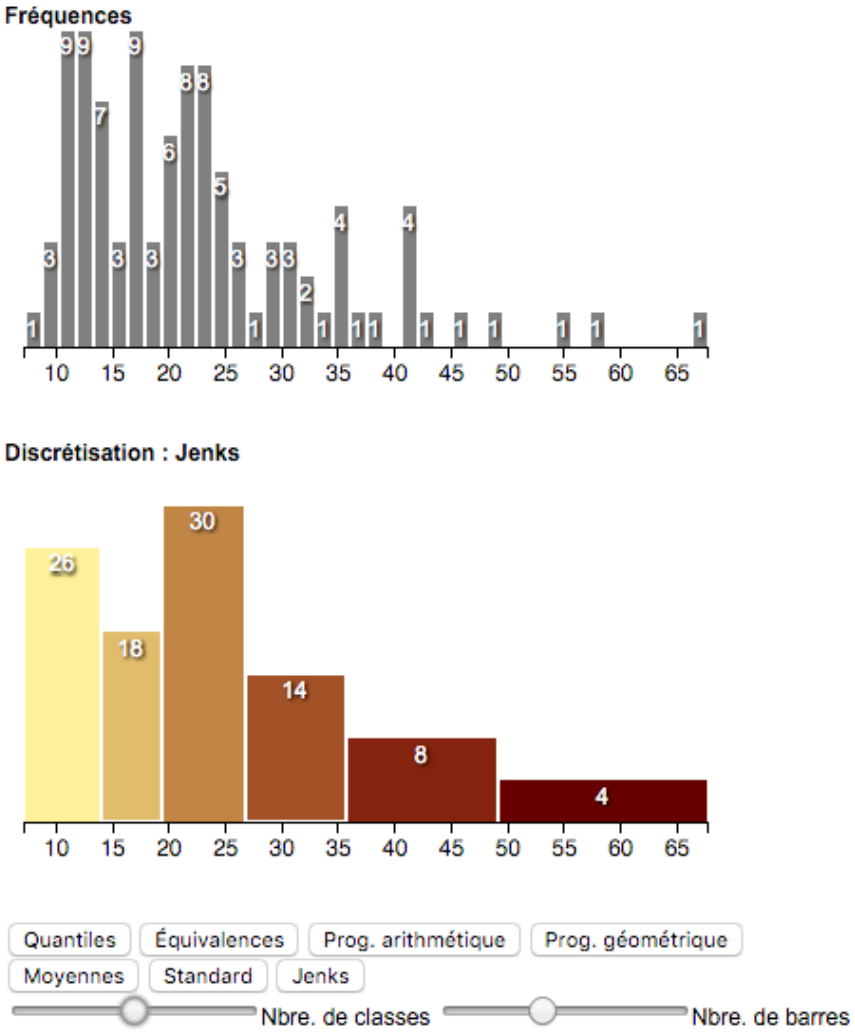
Méthode : La méthode de Jenks est l'une des plus intéressantes car elle permet d'obtenir des classes très proches de la distribution de la variable, en créant des classes contenant des valeurs homogènes mais des classes les plus hétérogènes entre elles. Elle a été mise au point par G. F. Jenks, un cartographe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

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 intra-classe 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 cartes associées ne produira pas des cartes facilement comparables entre elles (évolution dans le temps, comparaison de composantes d'une population ou de thèmes).





La méthode de Jenks : exemple