

ANALYSE DES DONNÉES LICENCE PRO 2025-2026

LA DISCRÉTISATION EN CARTOGRAPHIE

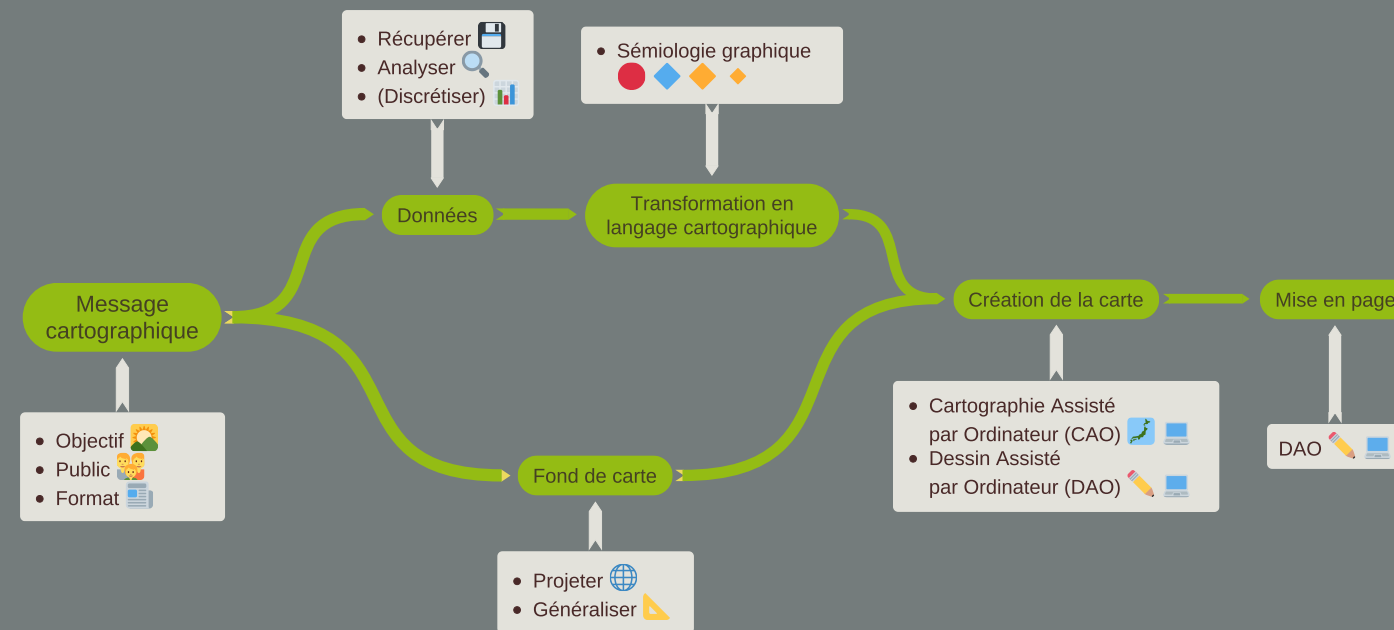
   Florian Bayer

ENSG 

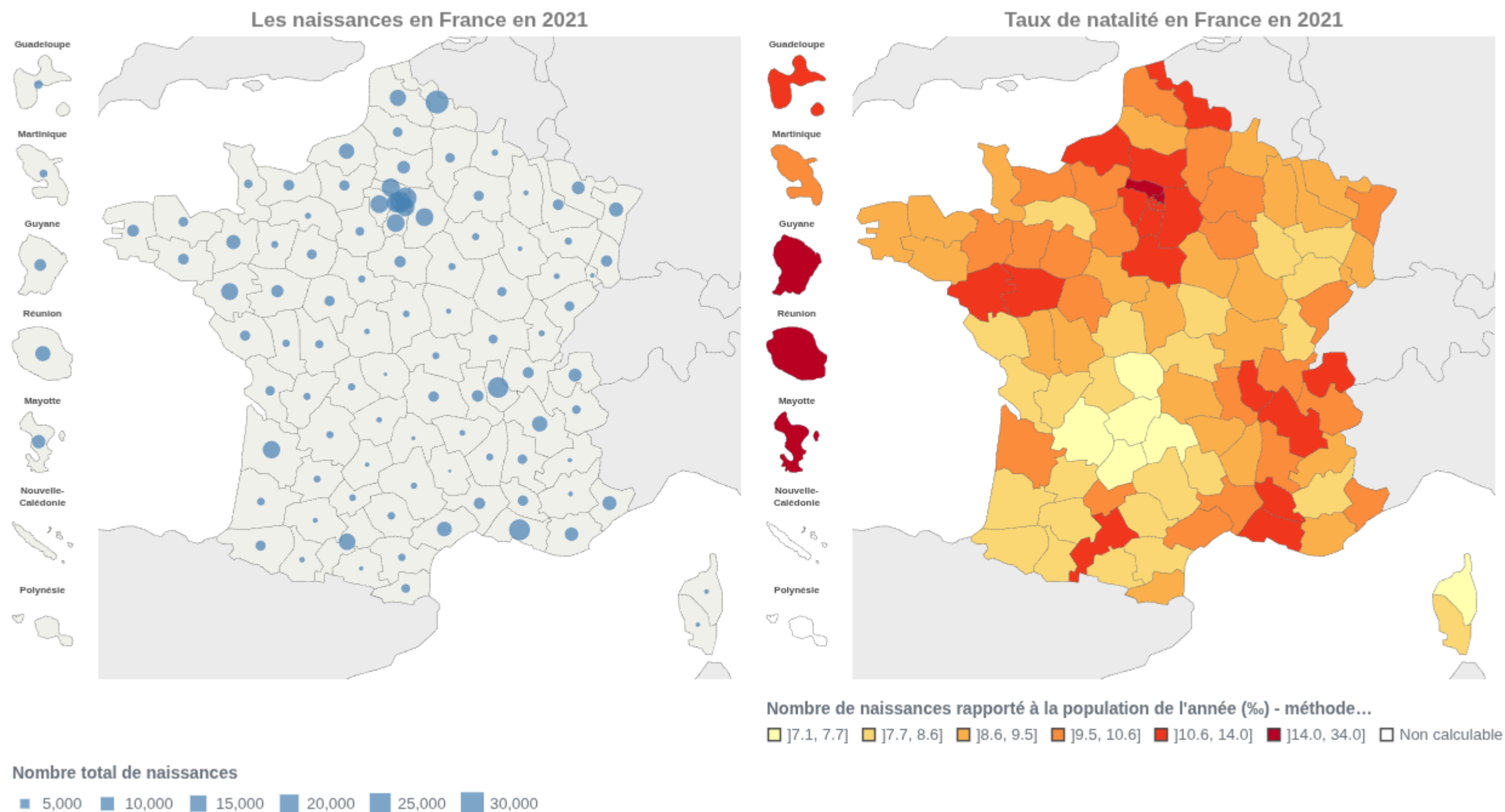
CONCEPTS ET OUTILS UTILISÉS DANS CET ENSEIGNEMENT



SCHÉMA DE PRODUCTION CARTOGRAPHIQUE



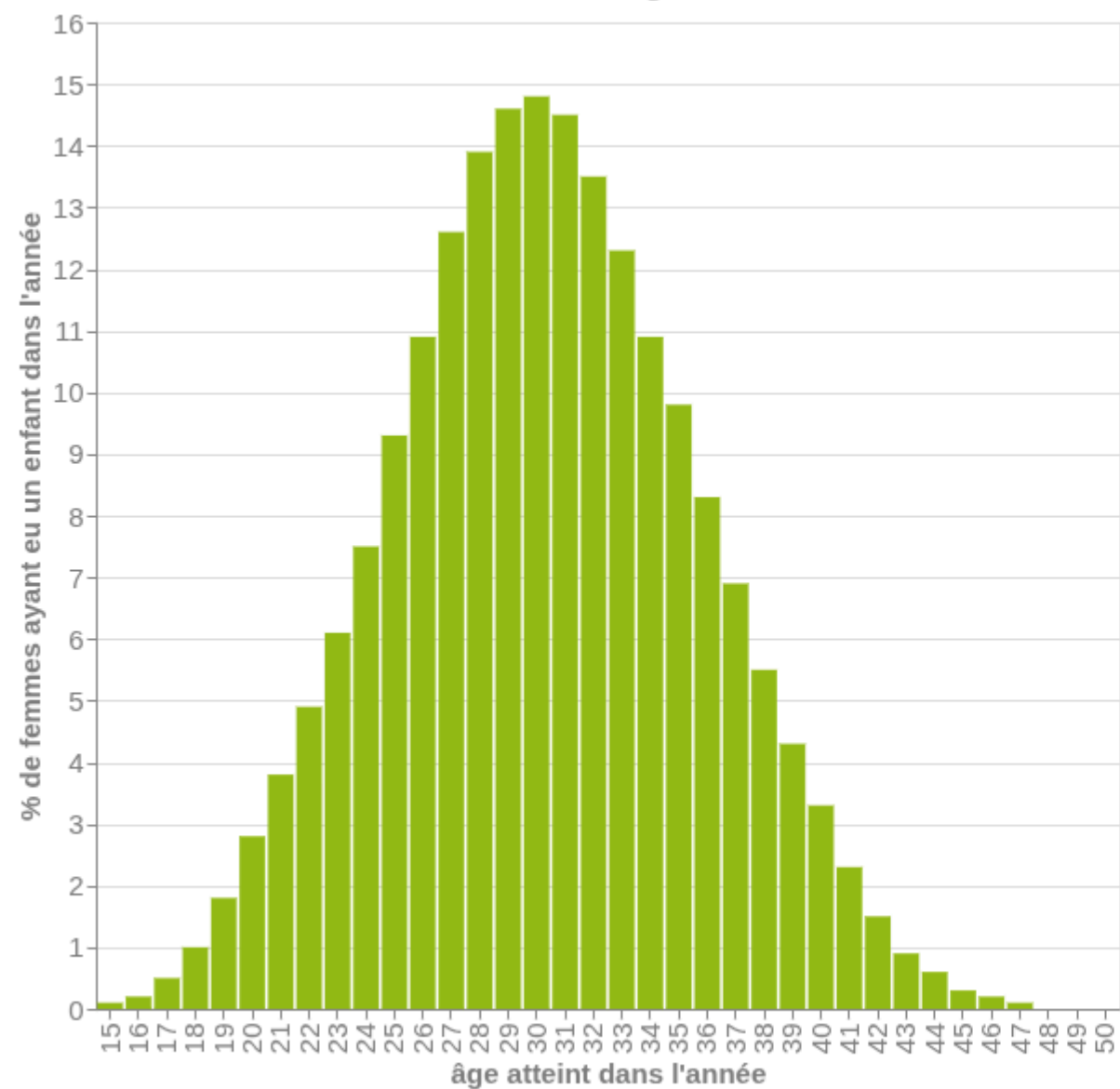
CARTOGRAPHIER POUR ÉTUDIER LA VARIABILITÉ SPATIALE



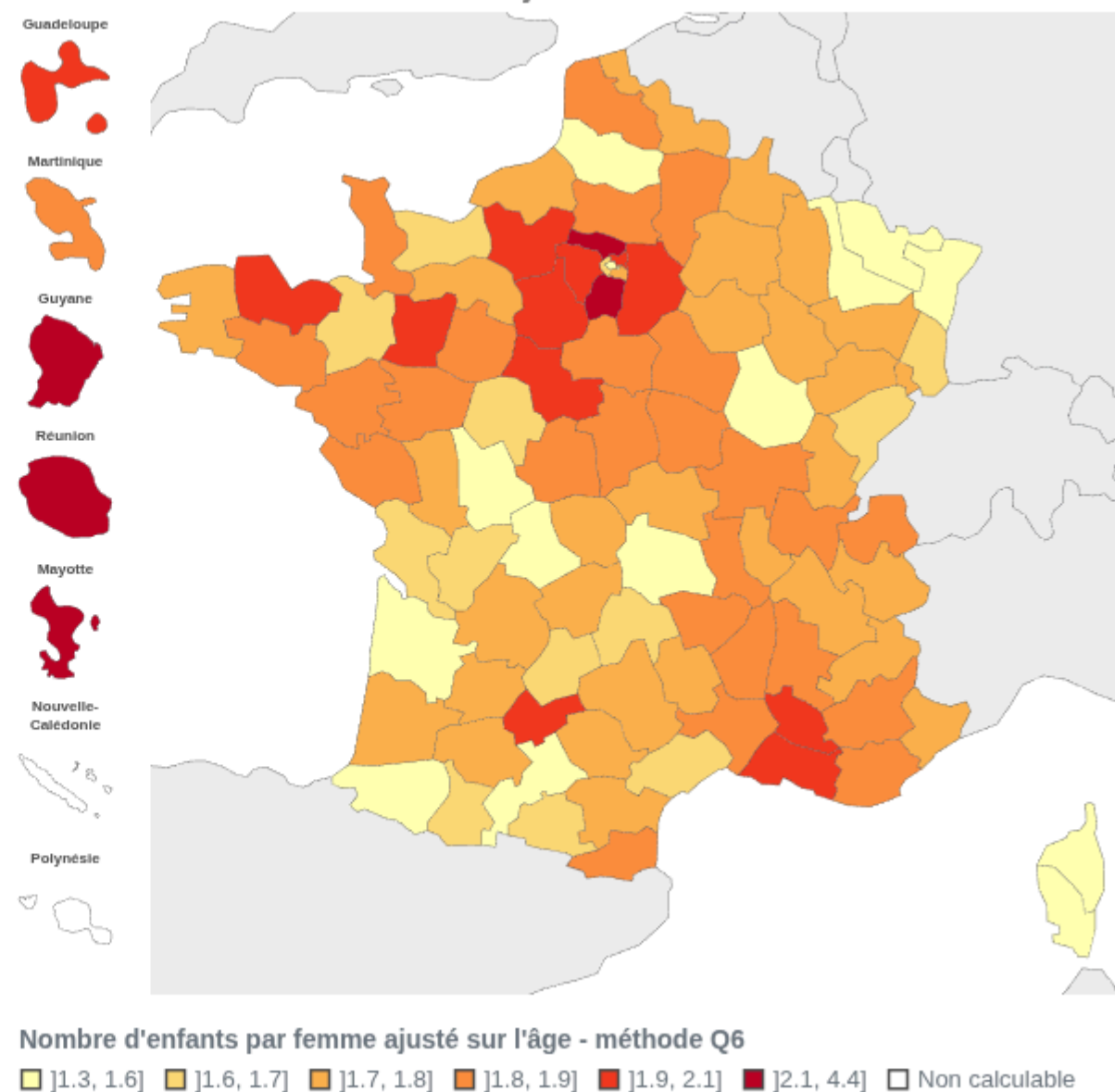


EN ÉTANT ATTENTIF AUX "SENS" DE VOS INDICATEURS

Taux de fécondité selon l'âge de la mère en 2014



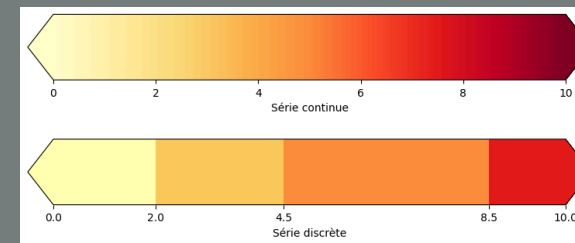
Indicateur conjoncturel de fécondité en 2021



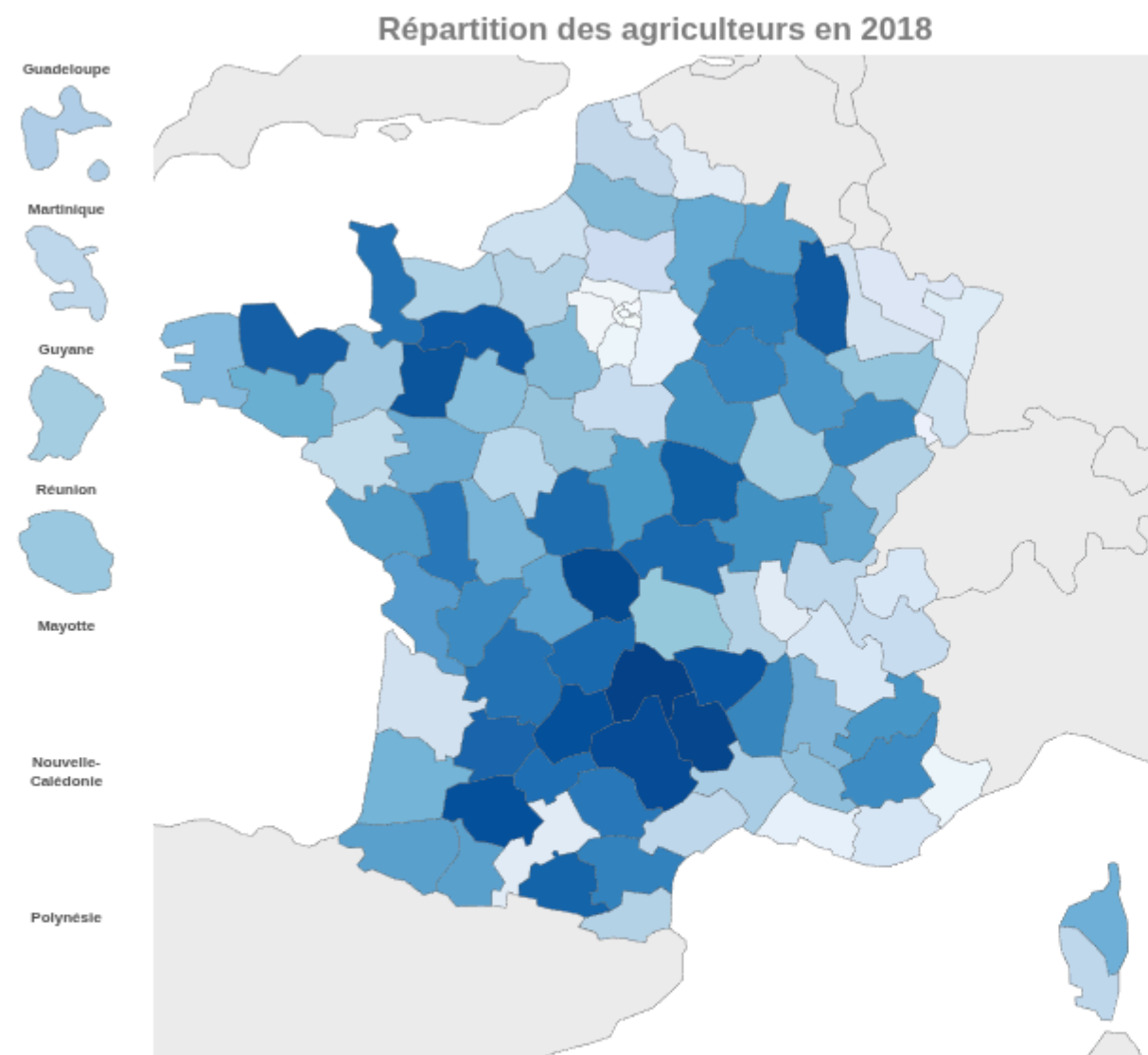
LA DISCRÉTISATION

En cartographie et en statistique, il est parfois nécessaire de simplifier l'information à transmettre.

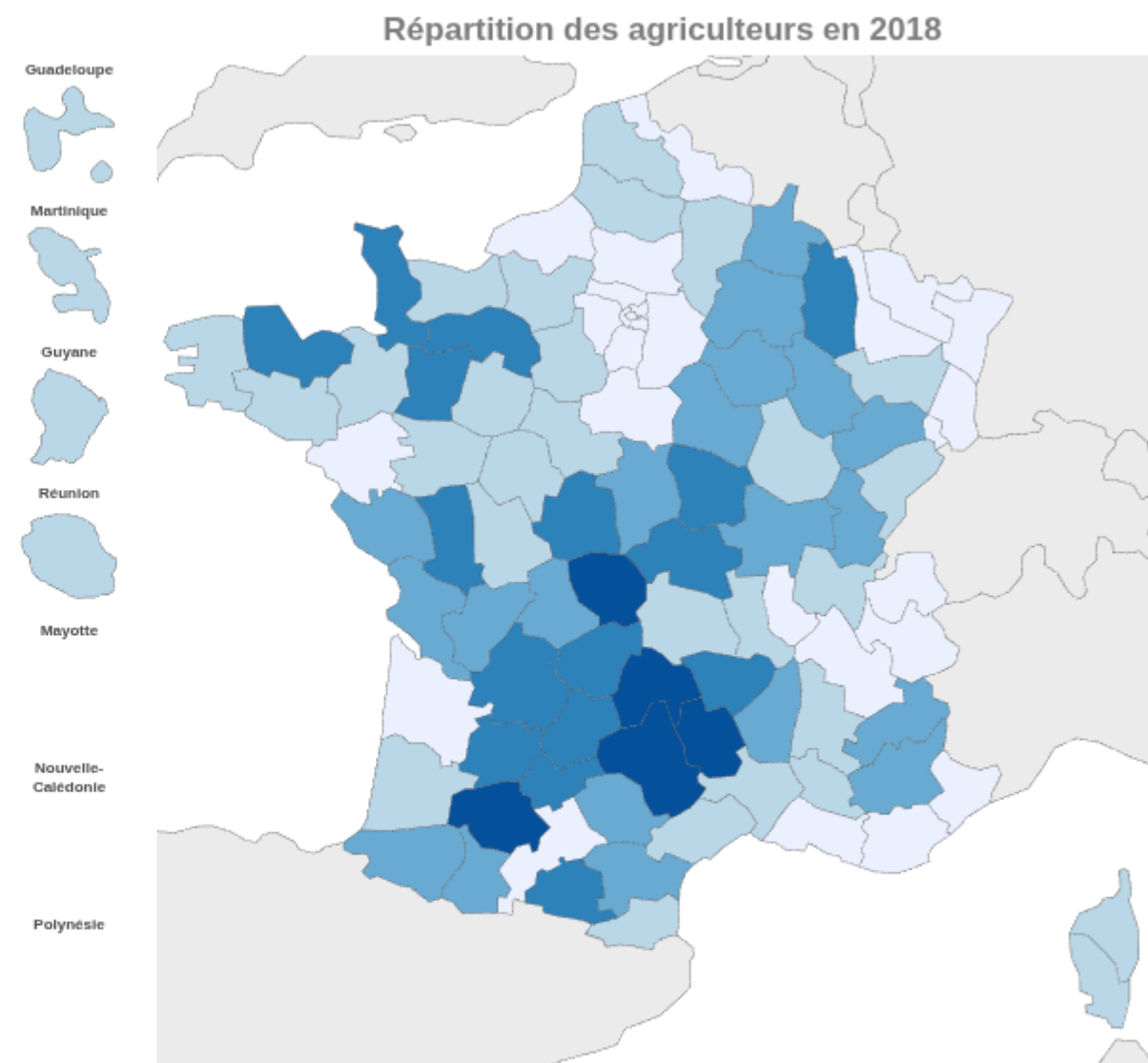
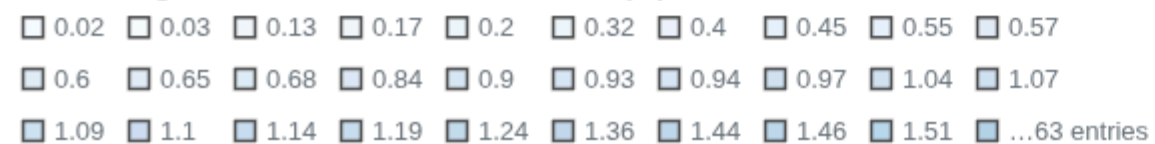
- Notamment lorsque la quantité d'information à représenter est très importante.
- La réduction de l'information au sein de classes est appelée la discrétisation.



? POURQUOI DISCRÉTISER ?



Part des agriculteurs sur l'ensemble des CSP(%)



Part des agriculteurs sur l'ensemble des CSP(%)



L'œil humain n'est pas en mesure d'associer un chiffre précis à la variable visuelle valeur (à l'inverse de la taille+forme). Seule la notion d'ordre est "innée" avec le rapport de noir et blanc sur une surface donnée.

Pour avoir associée à un niveau de gris un chiffre, il faut donc discrétiser

MINIMISER LA VARIANCE INTRA-CLASSE, MAXIMISER LA VARIANCE INTER-CLASSE (1)



MINIMISER LA VARIANCE INTRA-CLASSE, MAXIMISER LA VARIANCE INTER-CLASSE (2)

5.1	6	7	8	8.8	10
-----	---	---	---	-----	----

5.1	6	7	8	8.8	10
-----	---	---	---	-----	----

5.1	6	7	8	8.8	10
-----	---	---	---	-----	----

LE MESSAGE CARTOGRAPHIQUE



CONCEPTS-CLÉS DE LA CARTOGRAPHIE

- La carte communique une information par l'image
- Elle utilise un langage conceptualisé par Jacques Bertin, la sémiologie graphique : alphabet, vocabulaire et syntaxe
- Des biais cognitifs interviendront dans la conception de la carte (vision du cartographe sur ce qu'il observe).

Pour réduire ces biais et rendre votre message efficace, il faut :

- Utiliser les règles de la conception cartographique
- Penser la carte pour son public et non pour soi
- Adapter le message cartographique (public, support, objectifs).

AVANT DE FAIRE UNE CARTE

Identifier :

l'objectif de votre carte

- Dois-je faire une carte pour y répondre ?
- Dans quel contexte ?
(Explorer ? Communiquer ?)
- Quel est le message à faire passer ?

le public de votre carte

- Des experts sur le sujet ?
- Des novices ?

le support de la carte

- Papier ? Informatique ?
- Couleur ? Noir et blanc ?

LE MESSAGE CARTOGRAPHIQUE

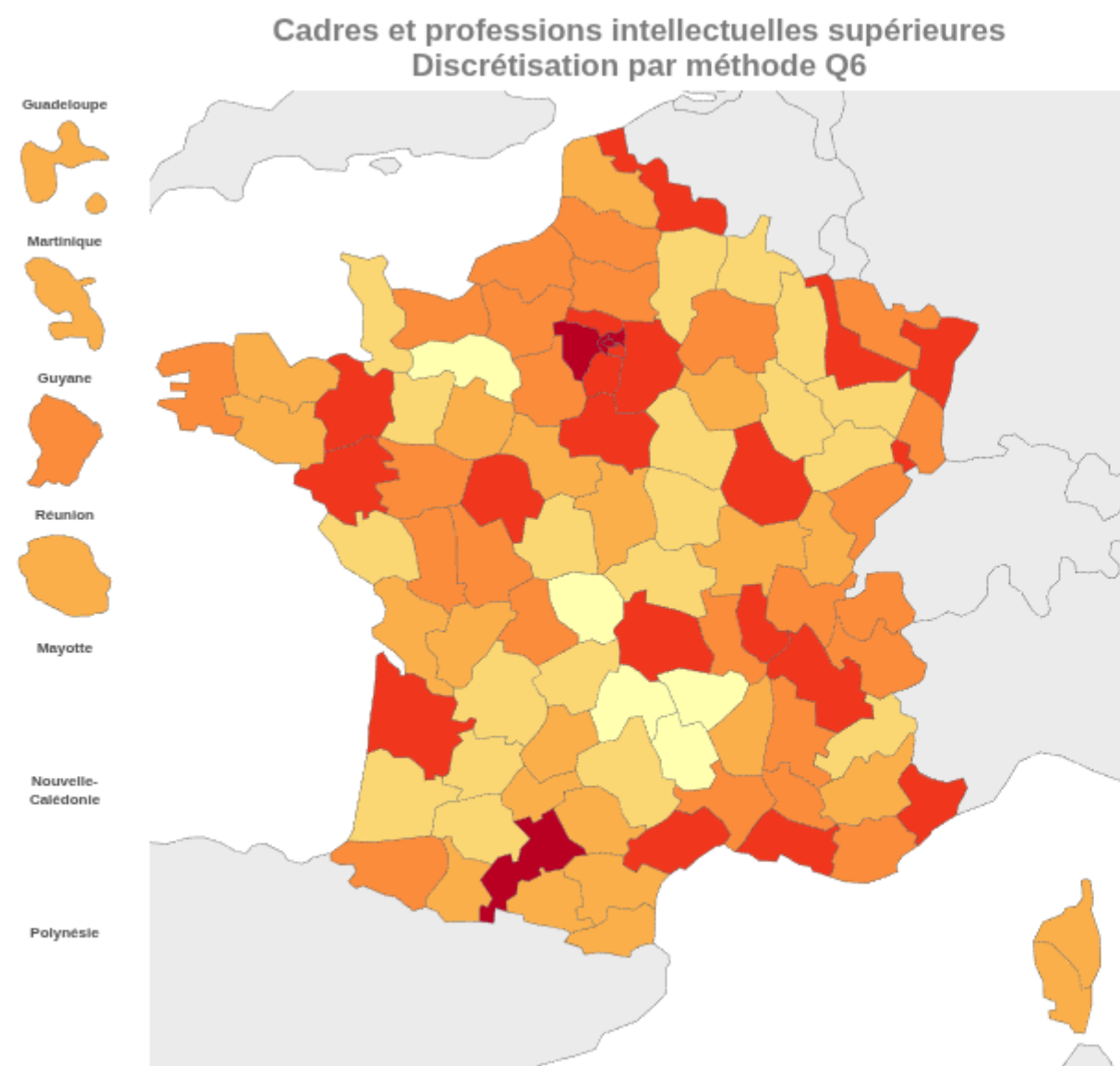
Le message cartographique guide l'ensemble de la production d'une carte.

Il faut toujours avoir conscience des points suivants :

- La carte n'est pas faite pour son auteur mais pour ses **lecteurs**.
- Il faut **adapter son message** aux types de lecteurs et au support de la carte.
- La carte doit être **simple et efficace** au niveau du rendu (pas dans sa conception).
- Le lecteur doit fournir un **minimum d'effort** pour comprendre la carte dès le premier coups d'œil.
- Les cartes sont un ensemble de petits mensonges communément acceptés (pour simplifier la compréhension du message)
- La carte est un outil de **communication très puissant**. Son utilisation doit se faire de manière **honnête et objective**.

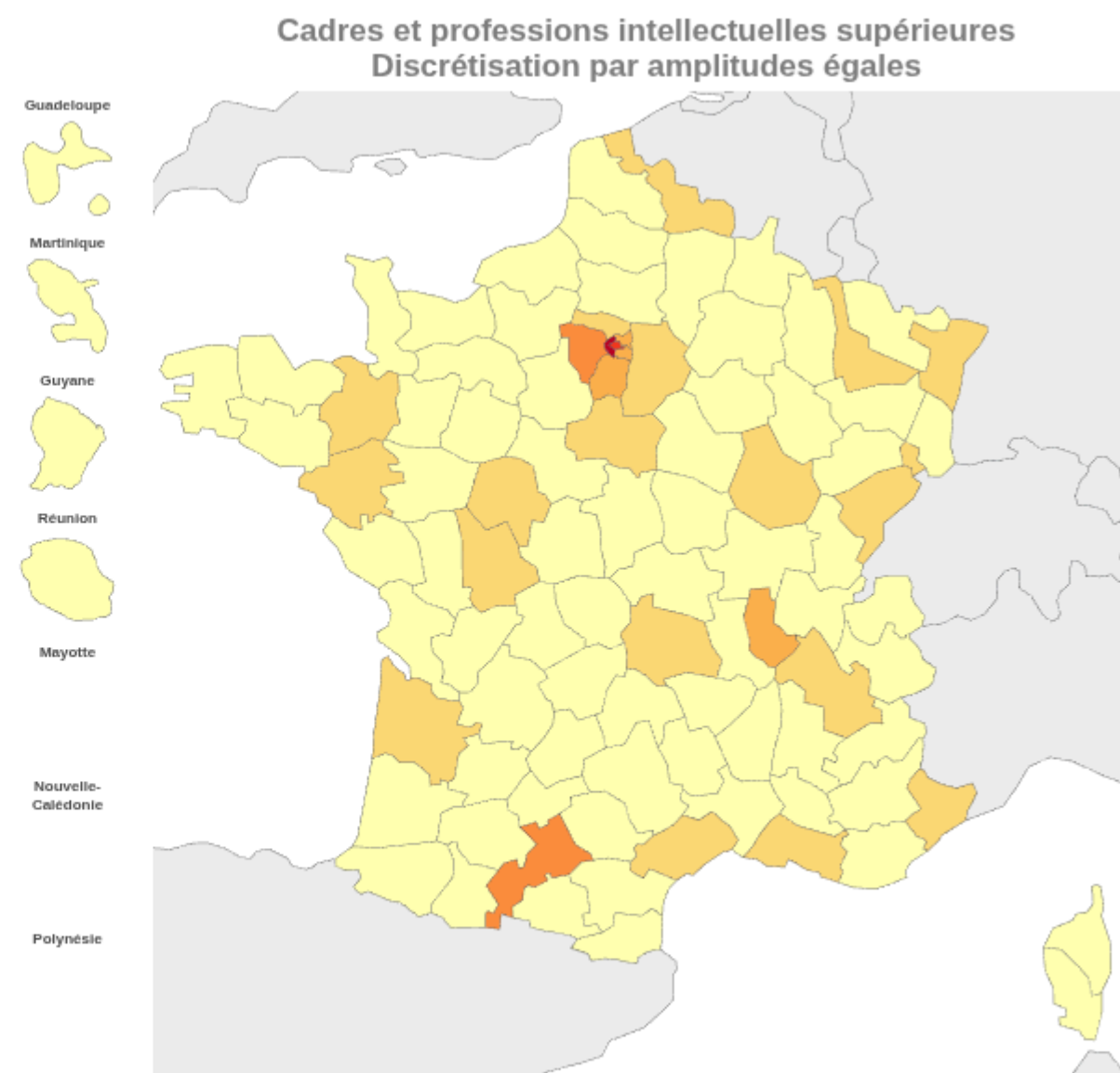
Cela passe par la bonne application des règles de la sémiologie graphique à l'ensemble de ces points Mais aussi dans certains cas un choix judicieux (rarement parfait) d'une discrétisation

IMPACT DE LA DISCRÉTISATION SUR LE MESSAGE CARTOGRAPHIQUE



Part des cadres sur l'ensemble des CSP(%)

■]8.3, 9.1] ■]9.1, 10.5] ■]10.5, 12.0] ■]12.0, 14.6] ■]14.6, 24.8] ■]24.8, 44.1]



Part des cadres sur l'ensemble des CSP(%)

■]8.3, 14.3] ■]14.3, 20.2] ■]20.2, 26.2] ■]26.2, 32.2] ■]32.2, 38.2] ■]38.2, 44.1]

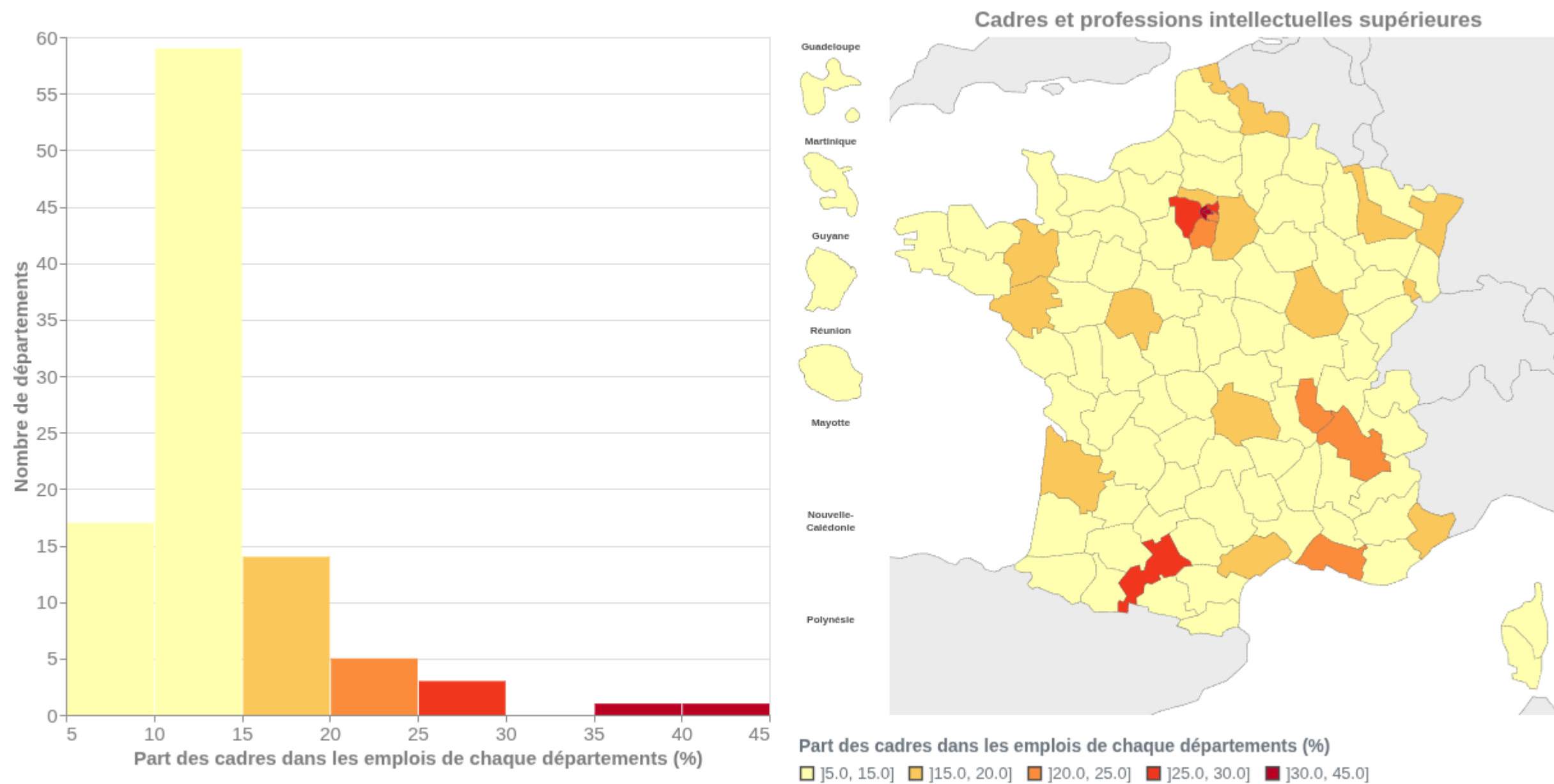
S'APPUYER SUR L'ANALYSE UNIVARIÉE DE LA SÉRIE

Il est essentiel de comprendre **les caractéristiques de la distribution** de la ou des séries de données avec les outils de l'analyse univariée :

- Elle permet de faire un **compromis** entre information statistique, information géographique et la bonne transmission du message.
- Elle permet **résumer** l'information en conservant la forme de la distribution
- Elle permet si besoin de mettre en évidence les valeurs **remarquables** et de les faire apparaître sur la carte
- Elle donne les éléments **scientifiques** pour justifier et reproduire ses choix.

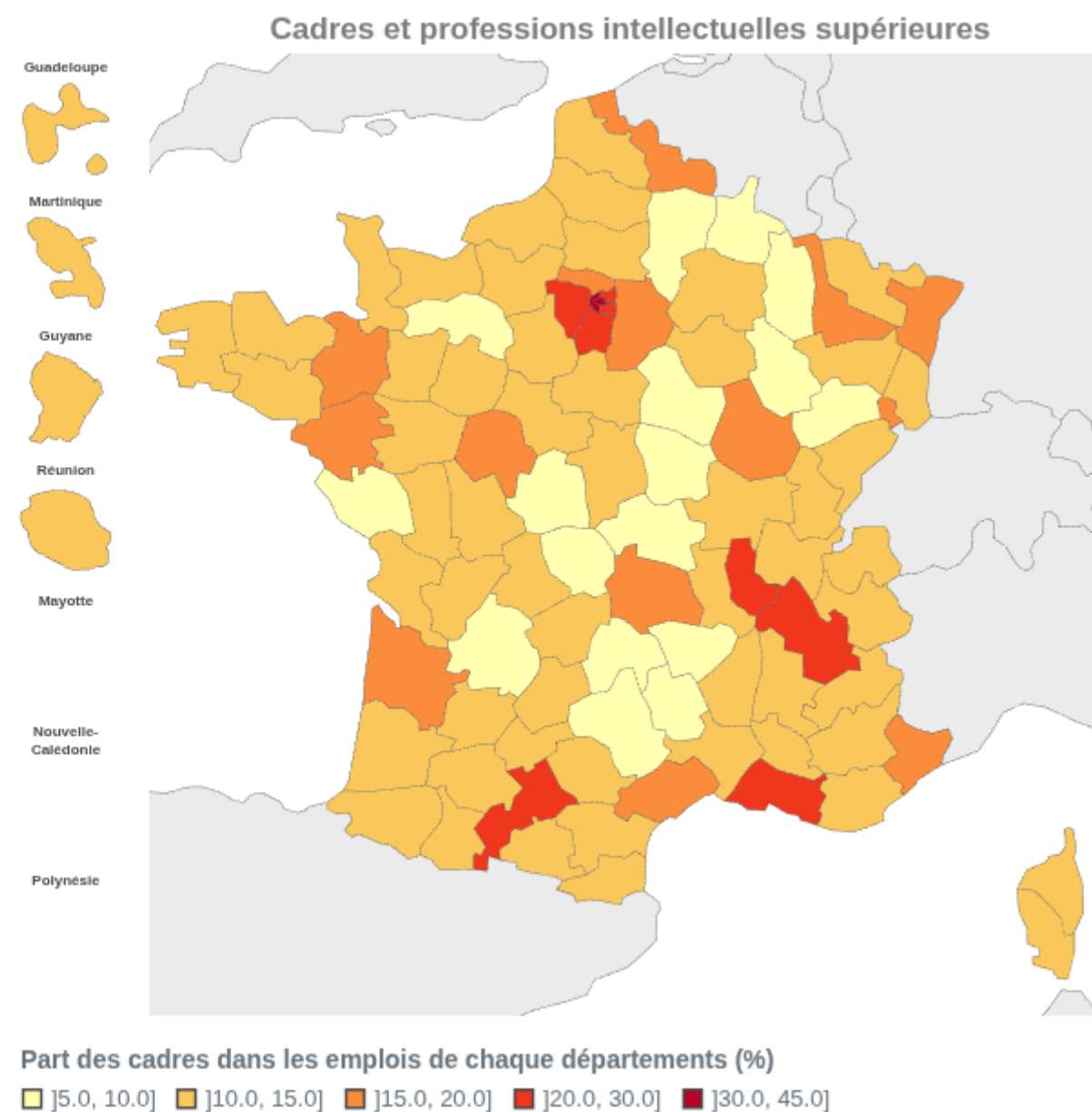
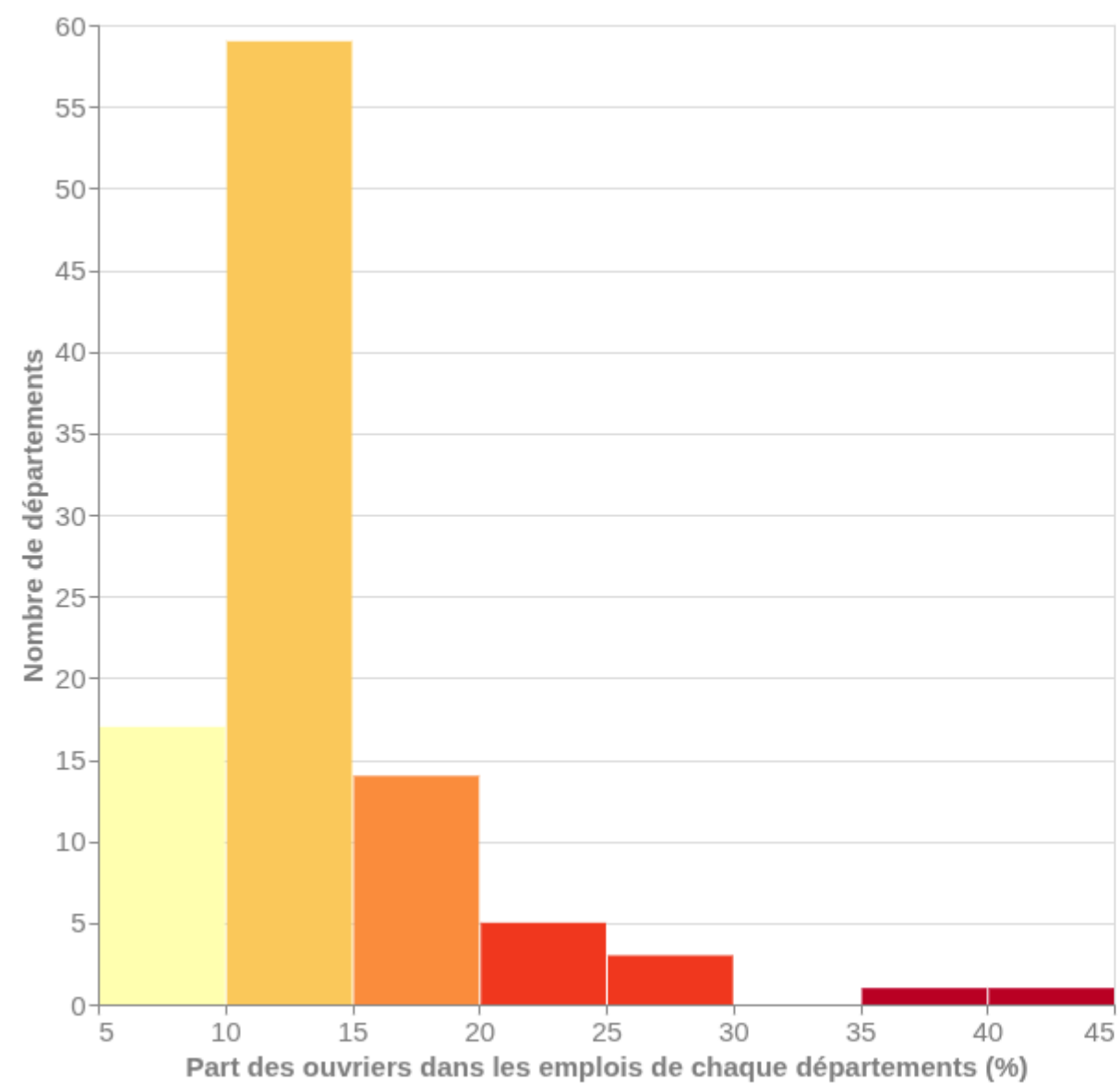
Dans le cas contraire, vous risquez d'avoir une carte n'apportant que très peu d'information, car la discrétisation sera mal adaptée au message cartographique

SUIVRE LA FORME DE LA DISTRIBUTION : EXEMPLE 1

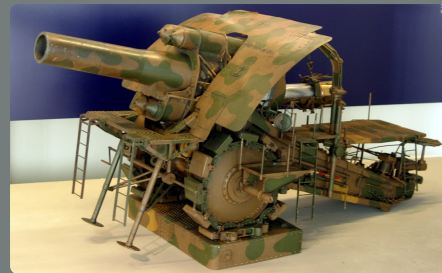


Attention, il faudrait normalement que la première et la dernière classe soient regroupées sur l'histogramme

SUIVRE LA FORME DE LA DISTRIBUTION : EXEMPLE 2



SIMPLE DANS LA PRATIQUE, MAIS...



Certaines contraintes peuvent s'ajouter et complexifier la discrétisation

- besoin de visualiser à un instant t un phénomène (le plus simple).
- Besoin de comparer des données identiques à deux dates.
- Besoin de comparer des données différentes.

LES MÉTHODES DE DISCRÉTISATION



RÈGLES DE DISCRÉTISATION

En cartographie, le découpage en classes d'une série de données suit les mêmes règles qu'en statistique :

- Les classes couvrent l'ensemble de la série statistique
- Elles sont contiguës
- Une valeur ne peut appartenir qu'à une seule classe
- Eviter si possible les classes vides

📊 LES QUANTILES (EFFECTIFS ÉGAUX)

Concept : même nombre d'individus dans chaque classe

Construction : nombre total d'individus (les départements) / nombre de classes souhaités

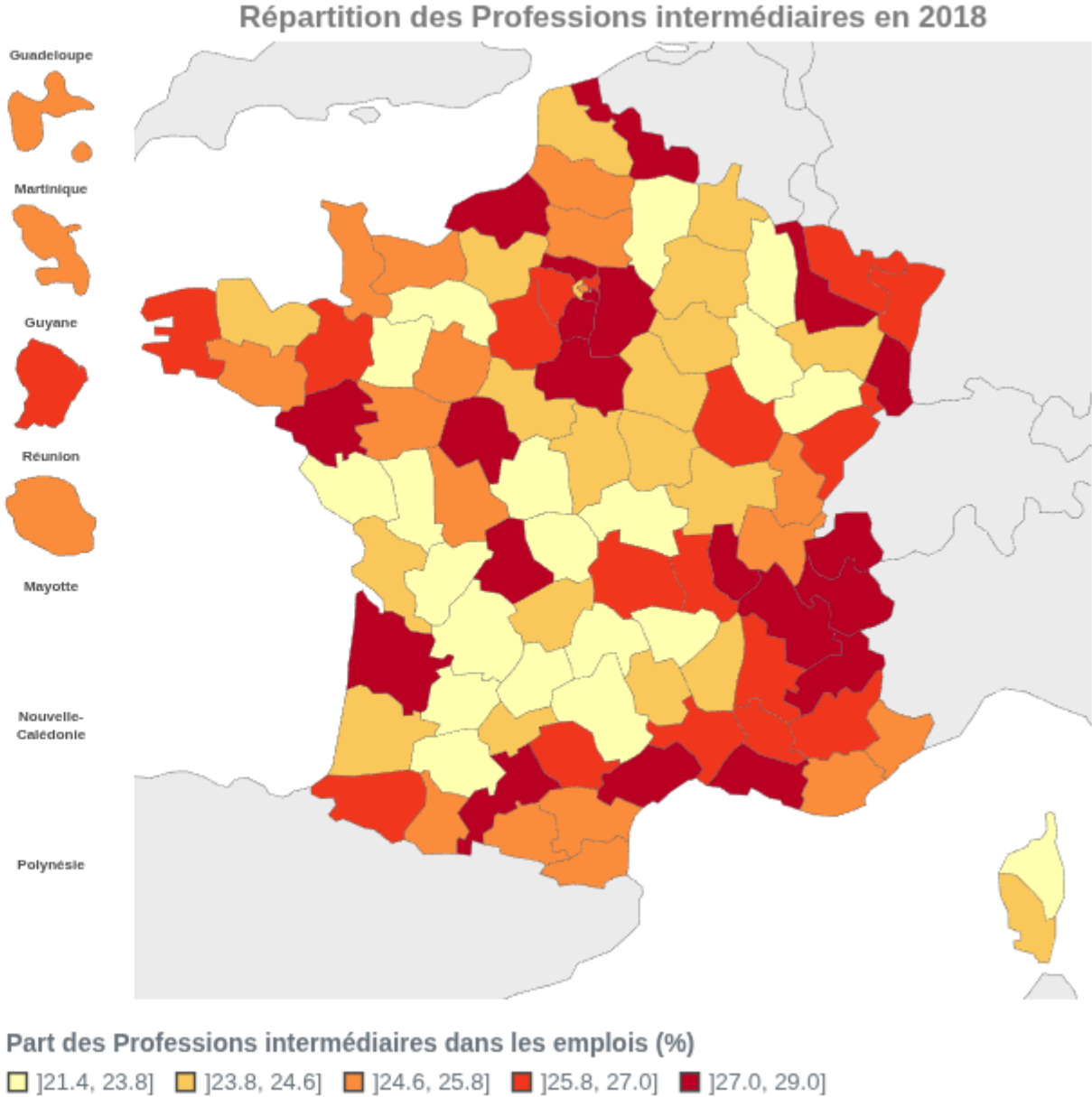
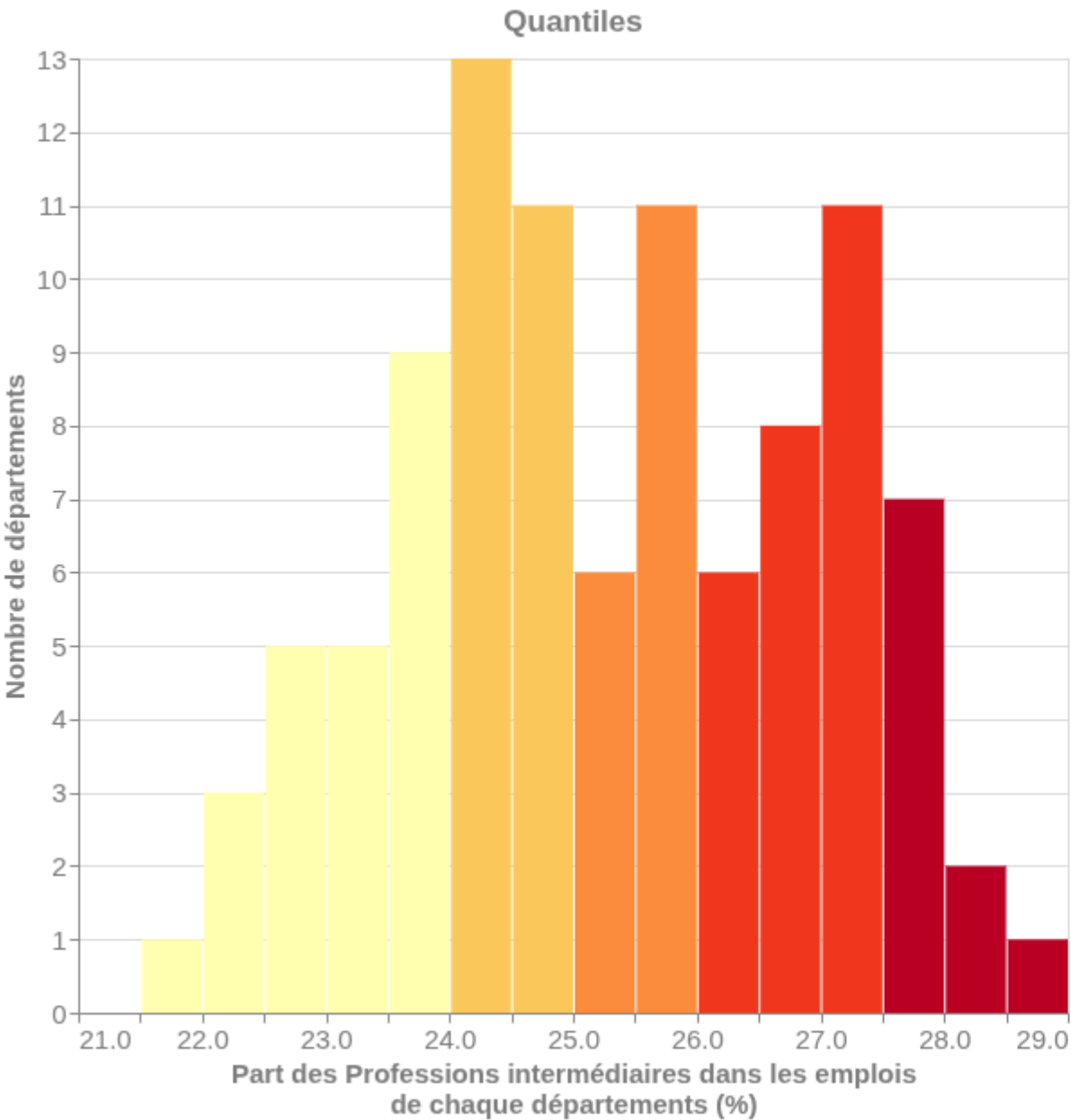
Avantages :

- Très facile à réaliser.
- Facilement compréhensible par le lecteur.
- Permet de comparer la position des individus géographiques dans différentes distributions (ordre de grandeur). Les bornes de classes ne seront pas les mêmes.

Inconvénients :

- Risque de perte d'information sur la forme de la distribution.
- Ne met pas forcément en évidence les valeurs extrêmes (max, min).

% LES QUANTILES (EFFECTIFS ÉGAUX)



📊 LES QUANTILES : VARIANTE Q6

Concept : Pour contourner le problème des valeurs extrêmes non mises en évidence avec les quantiles, Philcarto propose une méthode dite Q6. Ce sont des quartiles, mais la première classe contient les cinq pourcents valeurs les plus petites et non 25%, la dernière classe les cinq pourcents valeurs les plus fortes.

Construction : [Min : 5%[U [5% ; 25%[U [25% ; 50%[U [50% ; 75 %[U [75% ; 95 %[U [95% : max]

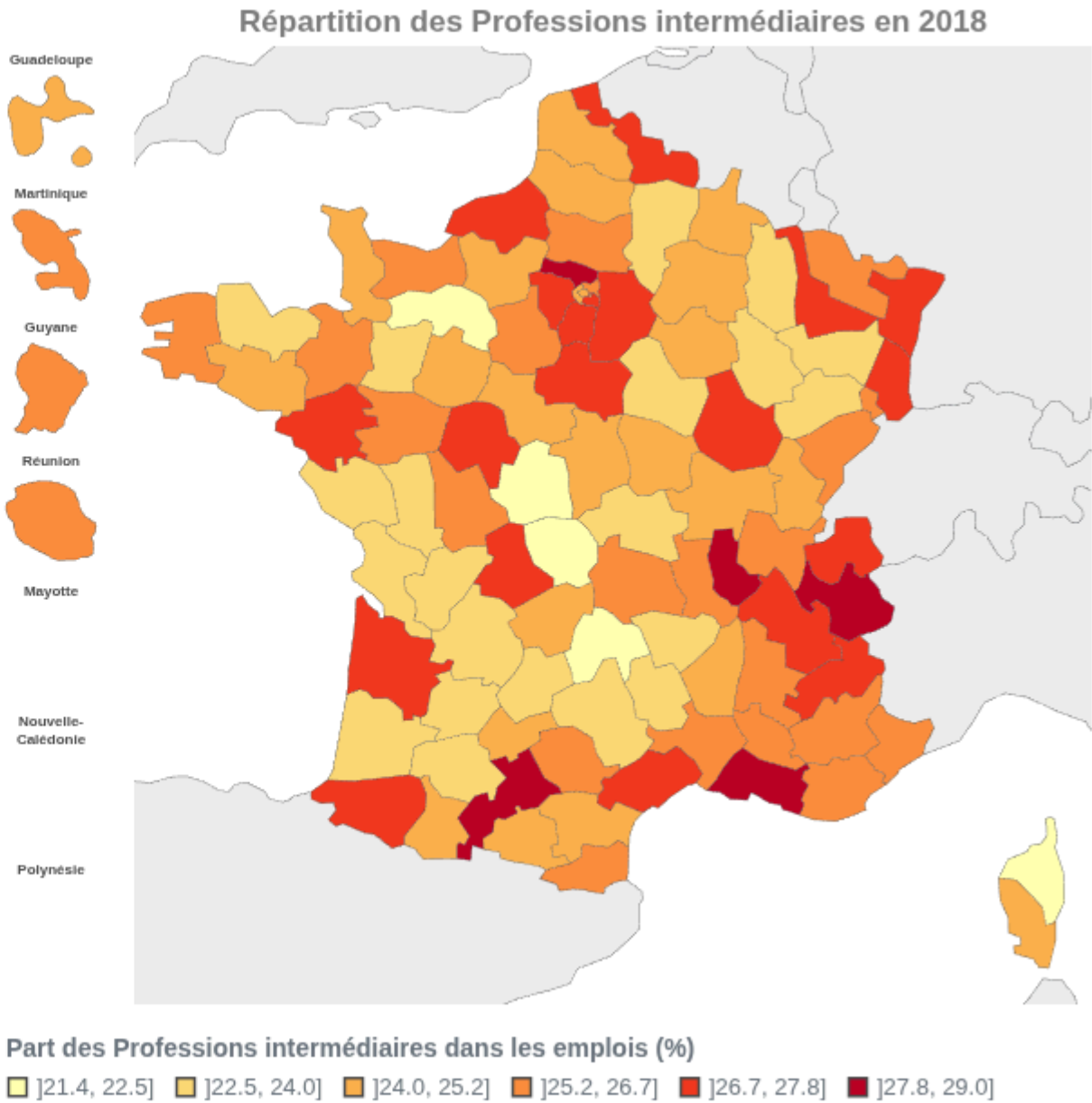
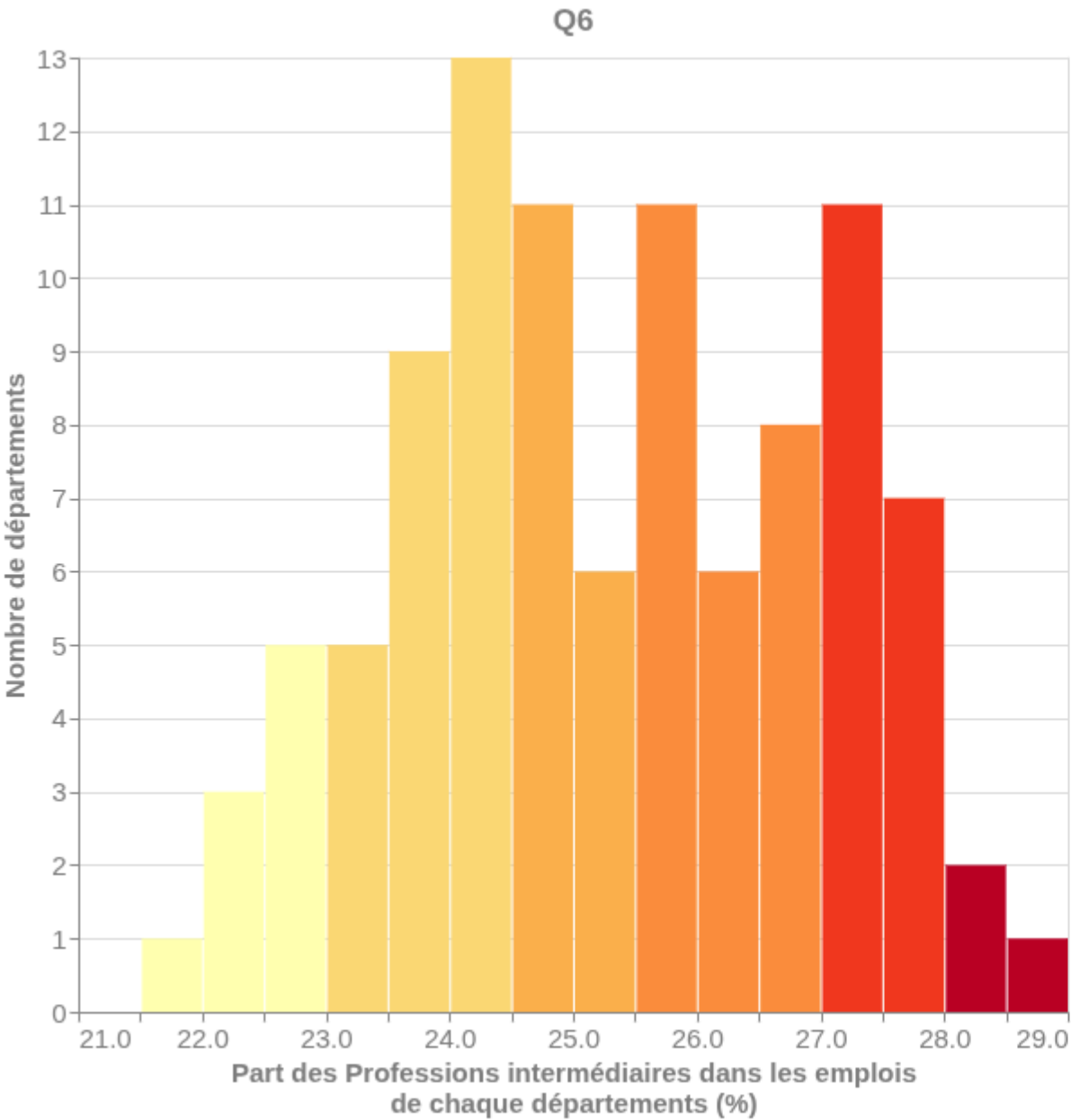
Avantages :

- Facile à réaliser (Quartiles ajustés).
- Mise en évidence des valeurs extrêmes.
- Permet de comparer la position des individus géographiques dans différentes distributions (ordre de grandeur). Les bornes de classes ne seront pas les mêmes.

Inconvénients :

- Risque de perte d'information sur la forme de la distribution (mais moins que pour des quantiles).
- Moins compréhensible par le lecteur que les quantiles (peu utilisées).

%
 LES QUANTILES : VARIANTE Q6



== LES AMPLITUDES ÉGALES

Concept : Les classes ont la même étendue (de 10 en 10, de 5 en 5 etc.)

Construction : $(\text{max} - \text{min}) / \text{nombre de classes souhaités}$

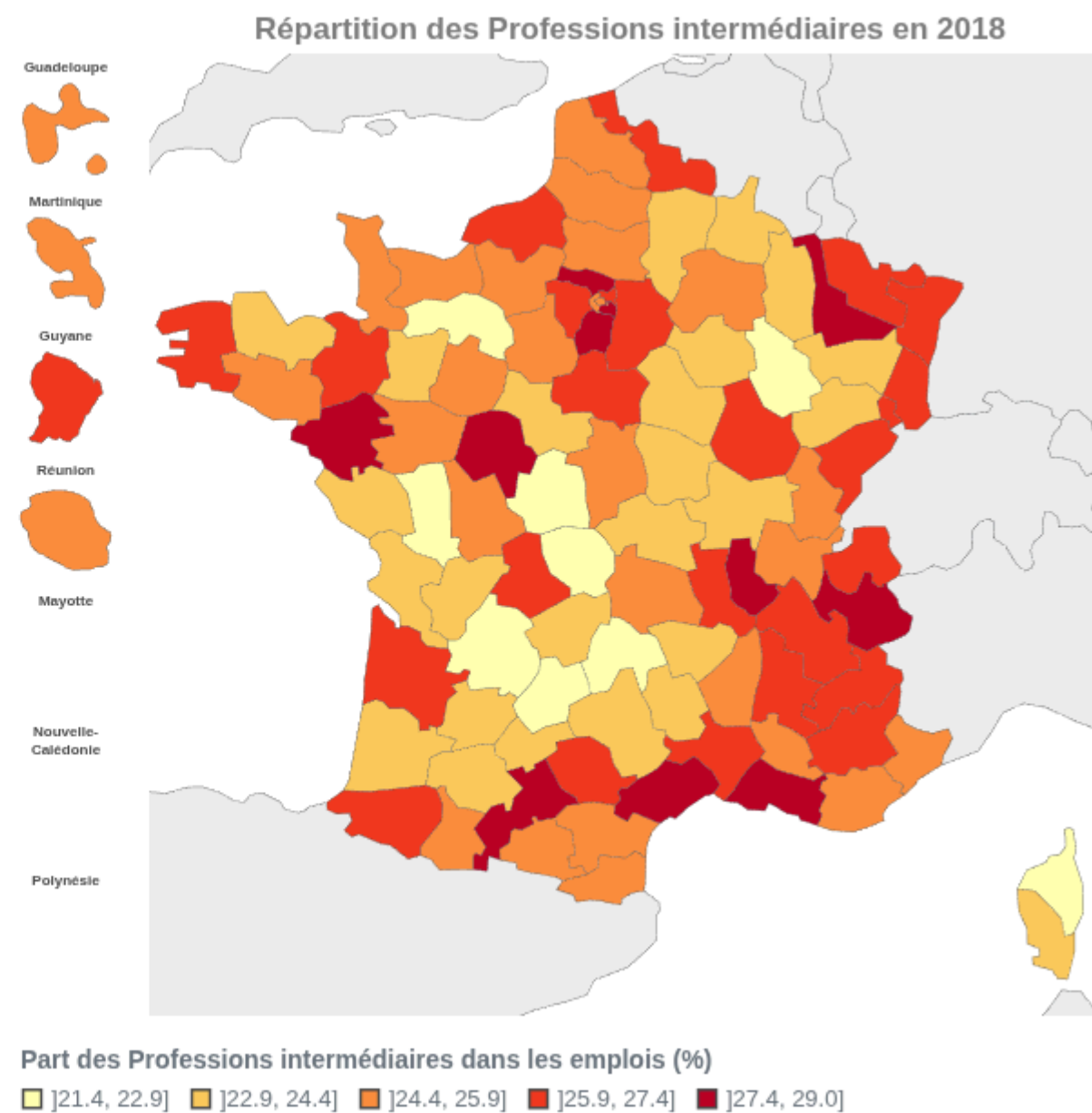
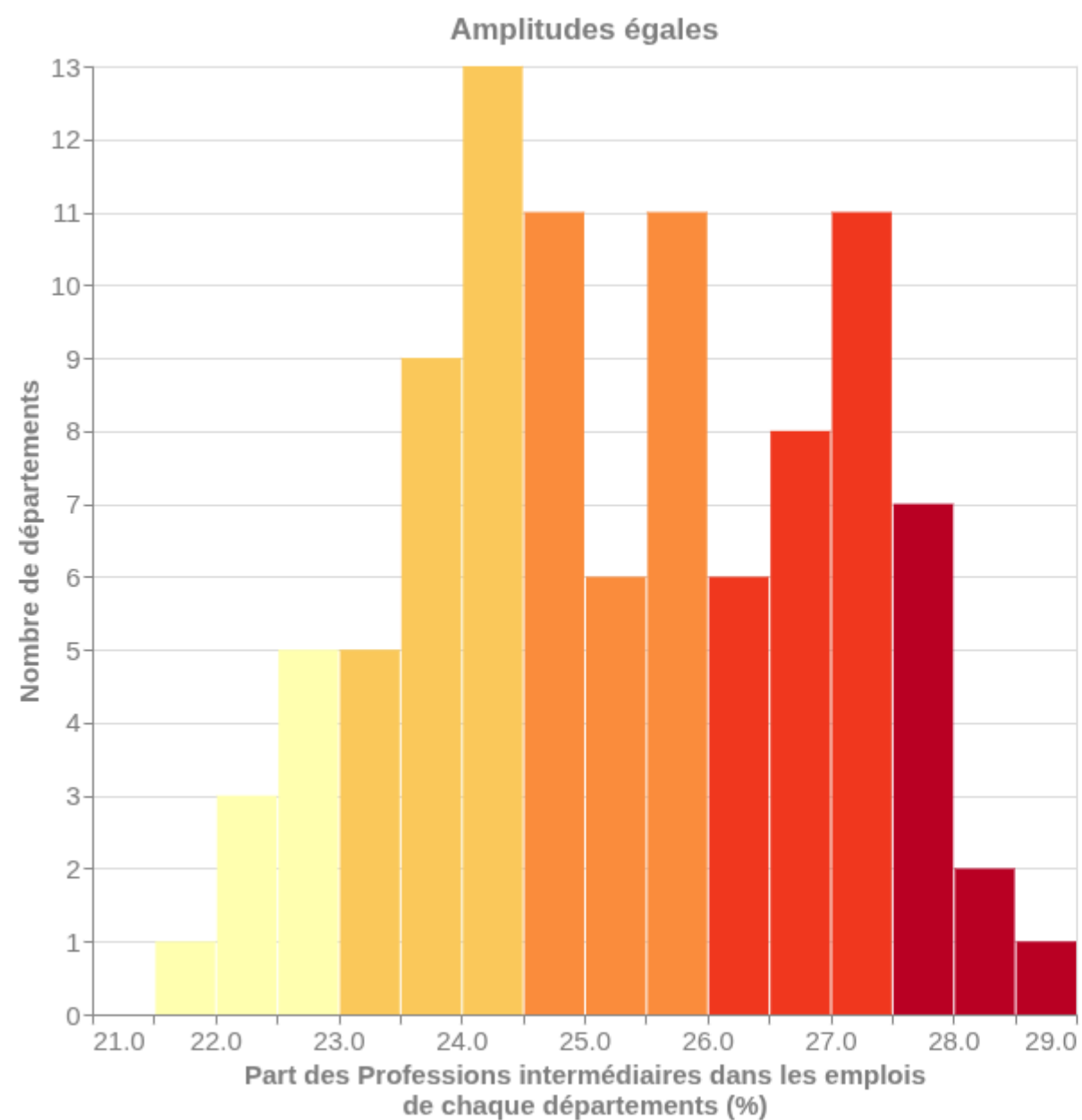
Avantages :

- Très facile à réaliser.
- Facilement compréhensible par le lecteur.
- Efficace sur les distributions uniformes.

Inconvénients :

- Très mal adaptée à une distribution non uniforme.
- Susceptible de créer des classes vides.

LES AMPLITUDES ÉGALES



LA MOYENNE ET L'ÉCART-TYPE

Concept : Les classes se basent sur les propriétés de la loi normale. La moyenne est de préférence au centre d'une classe. L'amplitude de la classe correspond à l'écart type ($0,5 \sigma$, 1σ , $1,5 \sigma$)

Construction : $[\text{Min} ; -1,5 \sigma[\cup [-1,5 ; -0,5 \sigma[\cup [-0,5 \sigma ; +0,5 \sigma[\cup [+0,5 \sigma ; +1,5 \sigma[\cup [+1,5 \sigma ; \text{Max}]$

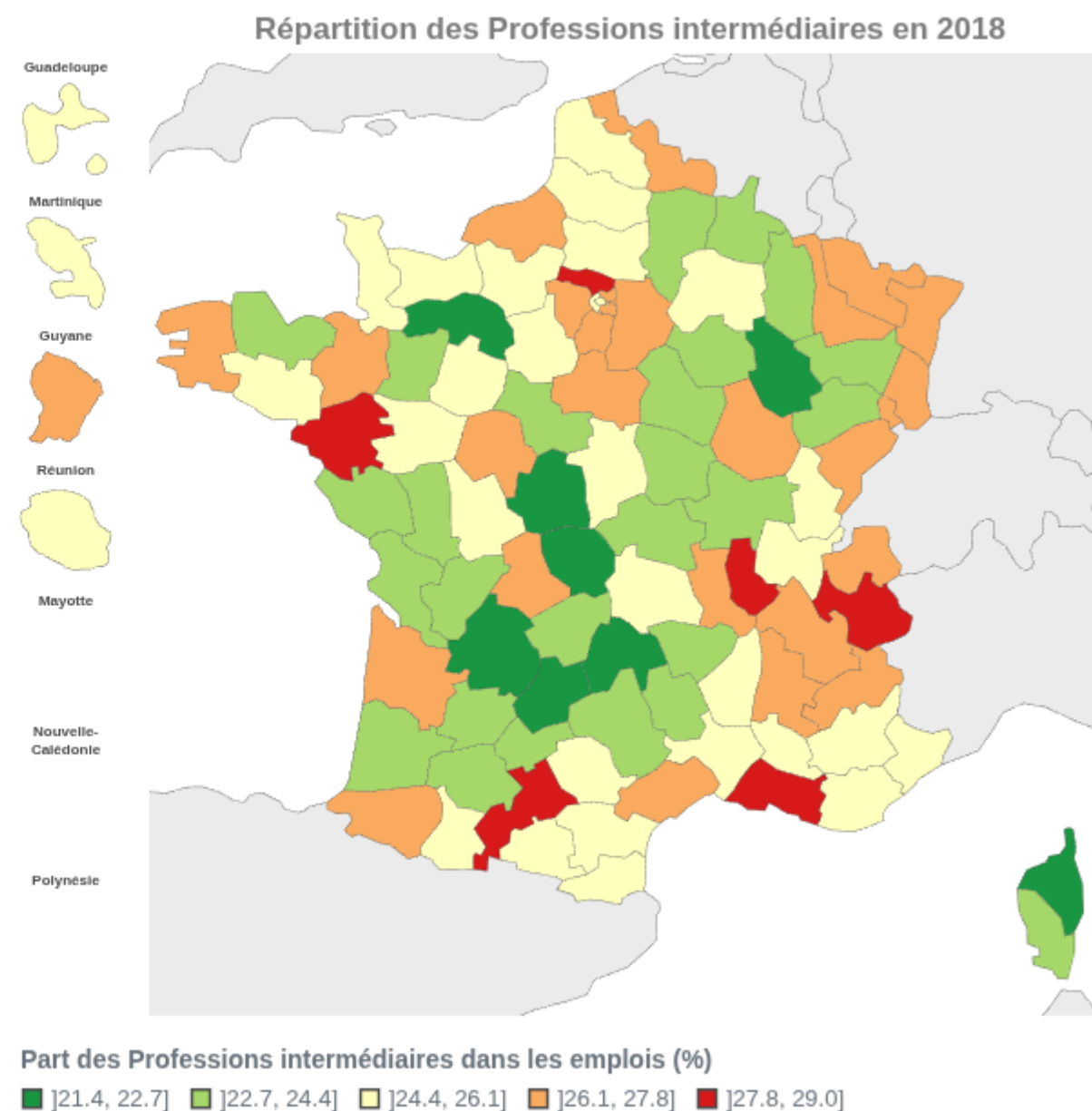
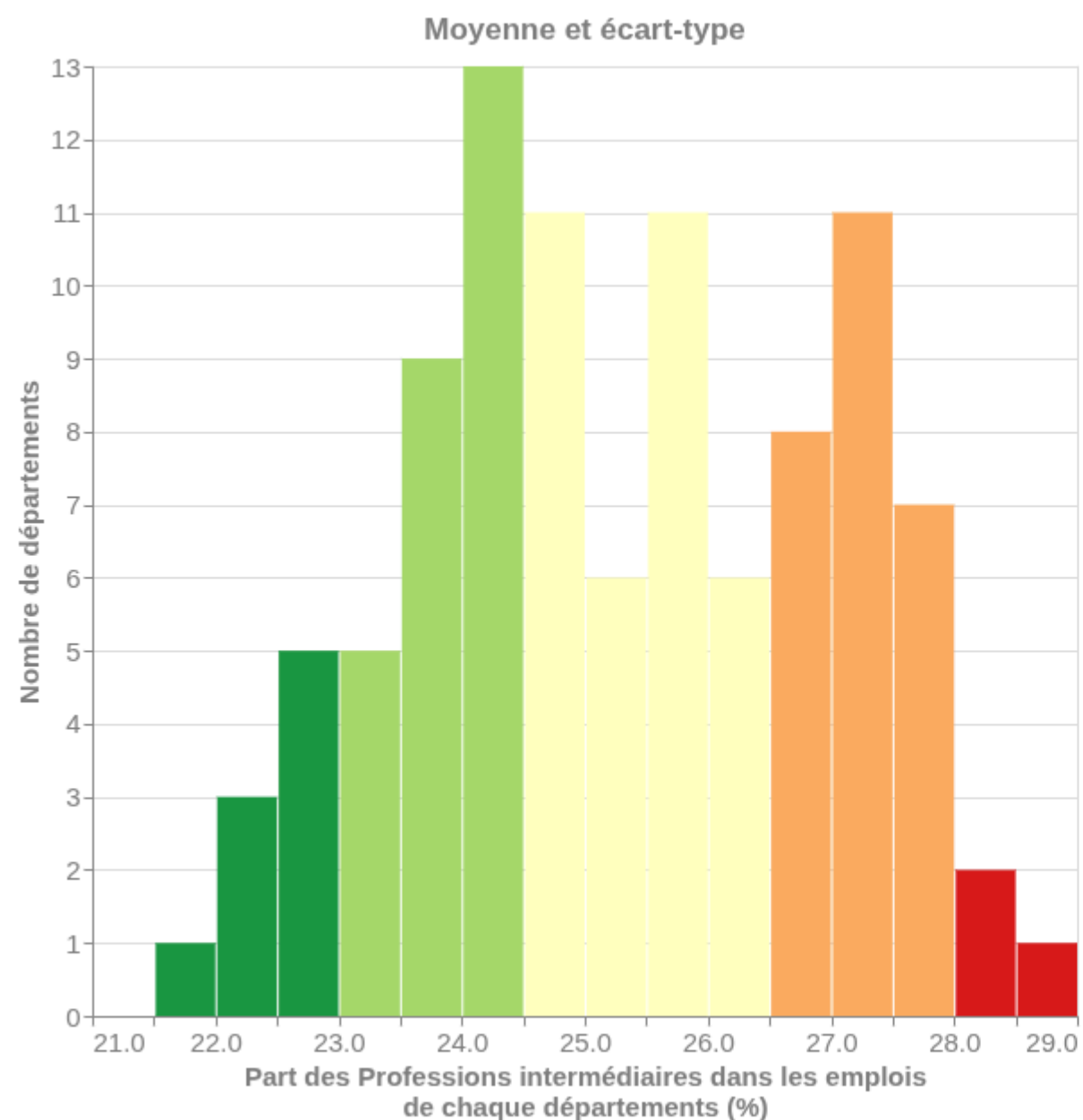
Avantages :

- A un sens sur les distribution gaussienne et permet dans ce cas un bon compromis géographique/statistique. Les classes extrêmes montrent les valeurs anormales, les classes centrales les valeurs proches de la normale.
- Facilement compréhensible par le lecteur initié.

Inconvénients :

- Difficile à comprendre pour le lecteur non initié (propriétés de la loi normale).
- Uniquement pour les distributions normales (transformation possible).

LA MOYENNE ET L'ÉCART-TYPE



ALGORITHME DE JENKS

Concept : les classes suivent au mieux la forme de la distribution, en regroupant les valeurs semblables et en isolant les valeurs extrêmes.

Construction : utilisation de l'algorithme de Jenks, qui minimise la variance intra-classe et maximise la variance inter-classe. Le cartographe peut "suivre" manuellement les coupures de l'histogramme, mais au prix d'une forte subjectivité (on parle de seuils naturels)

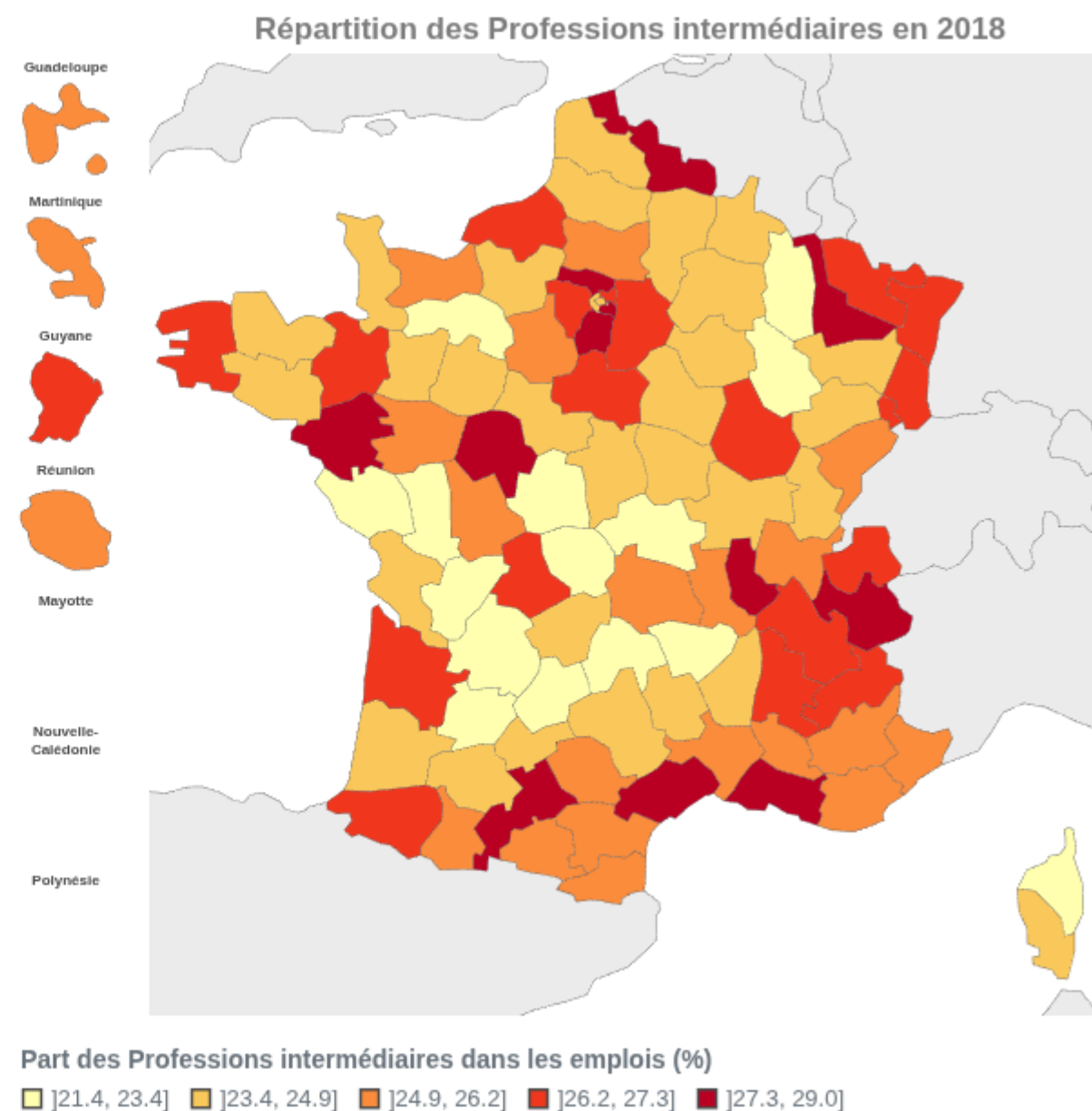
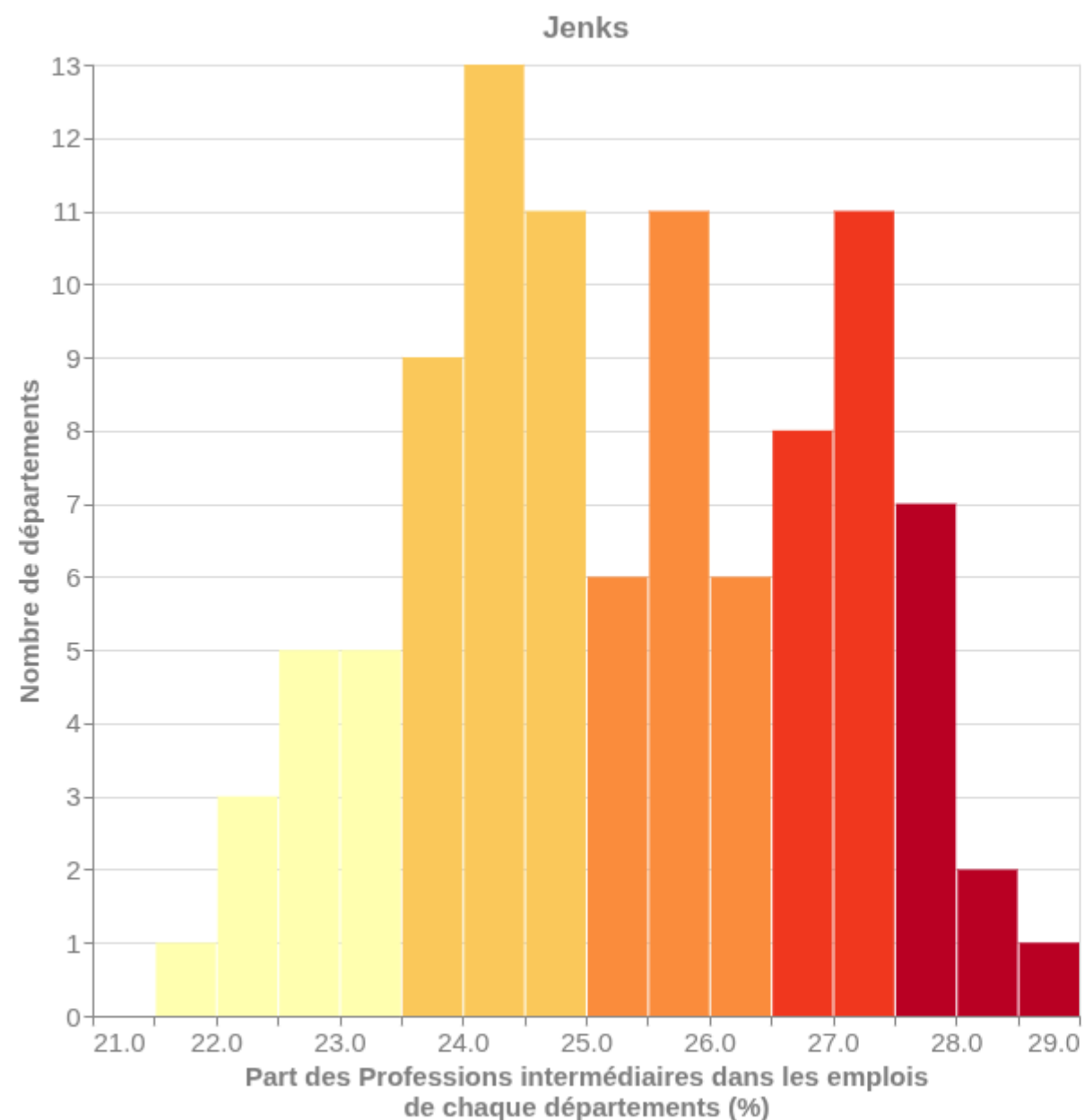
Avantages :

- Permet un excellent compromis entre la transmission de l'information et la conservation des caractéristiques de la distribution statistiques
- Les classes regroupent en leur sein les valeurs les plus semblables (minimise la variance intra-classe)

Inconvénients :

- Ne permet pas la comparaison de cartes si les bornes ne sont pas identiques.
- Subjectif pour les seuils naturels. Deux personnes travaillant sur la même série de données n'auront pas forcément les mêmes résultats.

ALGORITHME DE JENKS



QUESTIONS FRÉQUENTES



? COMBIEN DE CLASSES ?

Pour les données de taux, la transmission du message est en grande partie liée à la discrétisation.

En cartographie, discrétiser une série statistique suppose donc un compromis entre :

- La représentation et la transmission du message cartographique.
- Des biais cognitifs interviendront dans la conception de la carte (vision du cartographe sur ce qu'il observe).

Ce qui conduit souvent à un nombre de classes en cartographie allant de 4 à 7

- En dessous, l'information spatiale sera trop faible
- Au-delà, la carte sera trop complexe à comprendre : trop d'informations visuelles
- La longueur de la variable visuelle valeur ne permet pas à votre œil d'associer les différents niveaux de

? FAUT-IL ARRONDIR LES VALEURS DES CLASSES ?



- A part en science physique, garder 10 chiffres après la virgule n'a pas trop d'intérêt
- Dans la plupart des cas, arrondissez à un chiffre après la virgule, deux au maximum selon l'indicateur
- Mais il faut arrondir en amont de la mise en page. Cela évitera qu'un individu se retrouve dans la mauvaise classe (dans un logiciel de cartographie, changer bornes de classes met à jour automatiquement le rendu. Ce

COMMENT COMPARER DES SÉRIES ?

Soit comparer des données de même nature : comparaison absolue

- Une même valeur (niveau de gris) est associée à un même interval de classe entre les cartes à comparer
- Les bornes de classes doivent donc être identiques

Ou comparer des données de natures différentes : comparaison relative

- On compare la fréquence des individus de chaque classe
- Une même valeur (niveau de gris) est associée à une même fréquence entre les cartes à comparer
- On fait donc en sorte que les effectifs de

COMPARAISON ABSOLUE

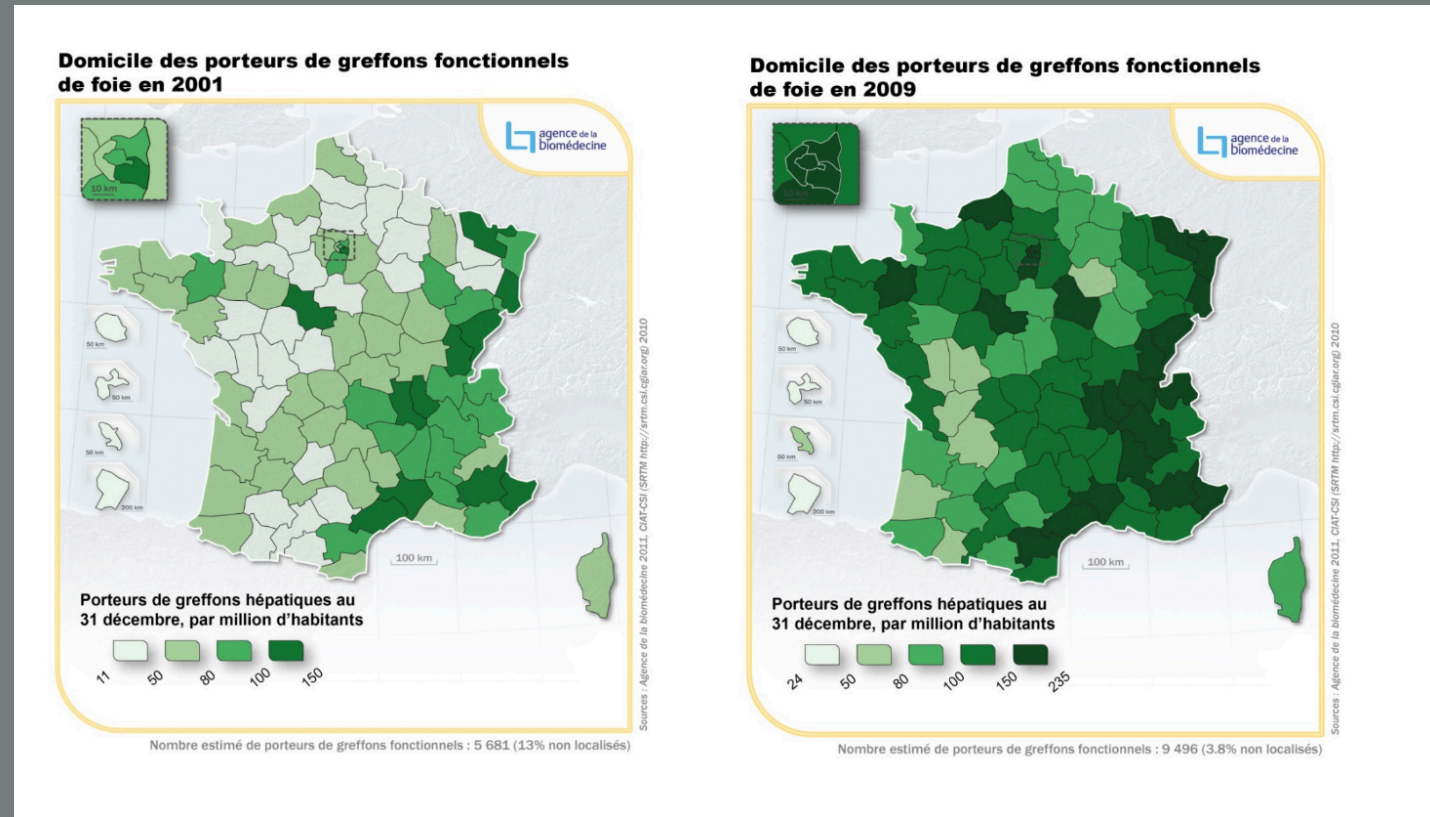
Si on souhaite comparer des données identiques, une solution est de discrétiser avec des bornes de classes identiques entre les cartes : comparaison absolue.

Les même classes avec des bornes identiques et le même niveau de gris se retrouvent sur toutes les cartes

- Amplitude égale :
 - Calcul de l'amplitude de classe à partir des min et max de l'ensemble des séries
- Jenks :
 - Appliquer l'algorithme sur une des séries puis appliquer les bornes de classe calculées aux autres séries
- Tout autre méthode du moment que les bornes de classes soient identiques
 - Toutefois réappliquer des bornes calculées sur des quantiles ou une méthode basée sur la distribution n'a pas grand sens

N'oubliez pas d'ajuster le min et le max de chaque série. Il est également possible d'ajouter ou supprimer des classes si nécessaire

COMPARAISON ABSOLUE : EXEMPLE



Dans cet exemple, une discrétisation Jenks a été appliquée sur les données 2001 puis retranscrites pour 2009 :

- En ajustant le minimum (11 vs 24)
- Et en ajoutant une classe supplémentaire pour 2009

COMPARAISON RELATIVE

Si on souhaite comparer des données différentes, les bornes de classes ne peuvent plus être identiques. On doit alors comparer la position relative des individus géographiques : comparaison relative

Les même classes avec des fréquences identiques et le même niveau de gris se retrouvent sur toutes les cartes

- Quantiles :

- Vous comparer les individus appartenant à chaque n ième classe de la série A avec ceux de la même classe de la série B

- Q6 :

- Même principe, sauf que les classes extrêmes contiennent chacune 5% des effectifs

- Moyenne et écart-type

- Même répartition au sein de la loi normale. Attention, la moyenne et l'écart-type ne doivent pas être significativement différents entre les différentes séries à comparer

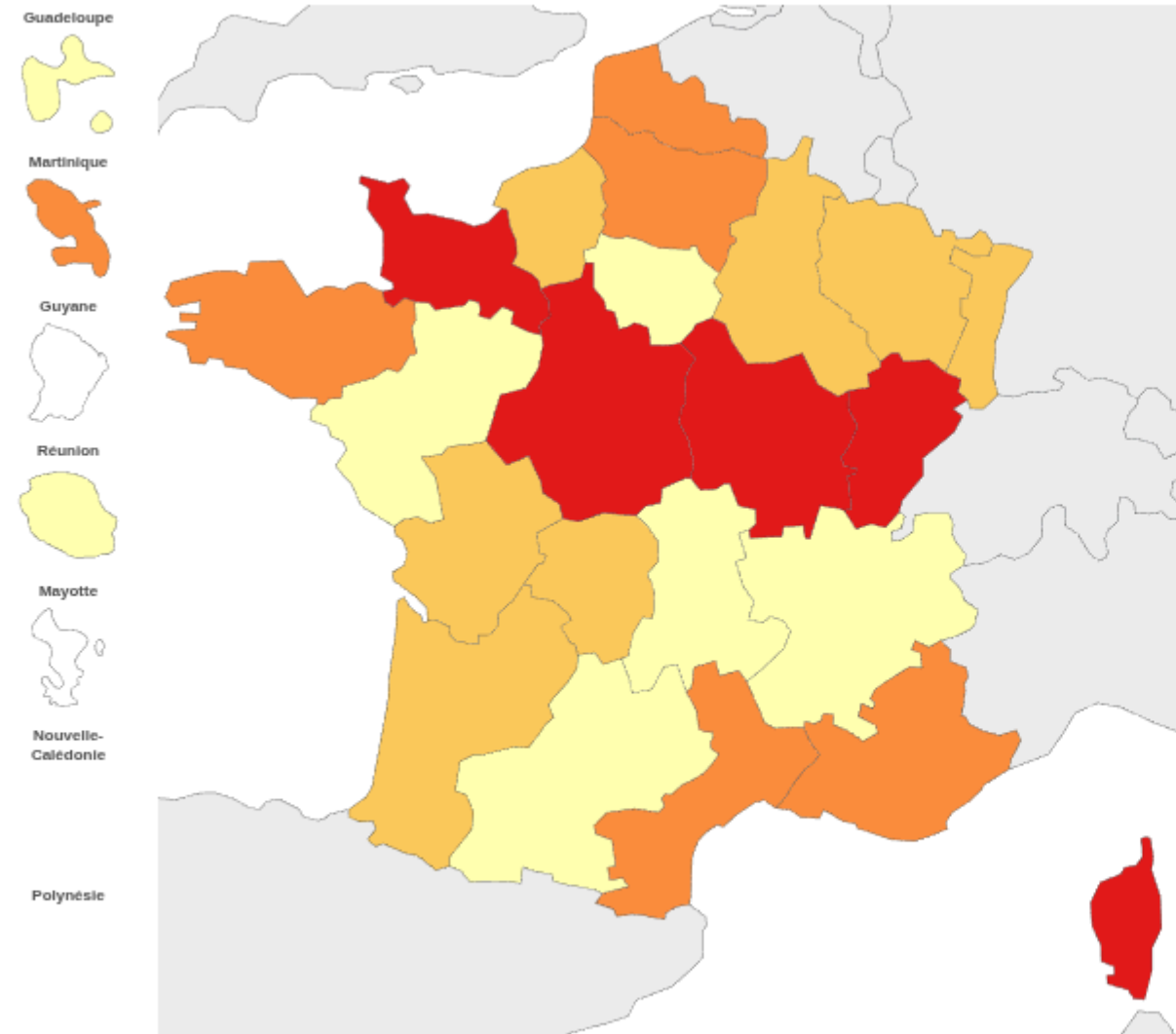
Il est évidemment possible d'utiliser une comparaison relative pour des données de même nature

COMPARAISON RELATIVE : EXEMPLE

Une discrétisation en quartile a été appliquée sur les deux séries de données :

- Les valeurs des bornes sont différentes
- Mais on peut comparer les 25% régions où le taux de sujets recensé est le plus faible V.S. les 25% des régions où les sujets prélevés sont les plus faibles. Idem pour chacune des classes.

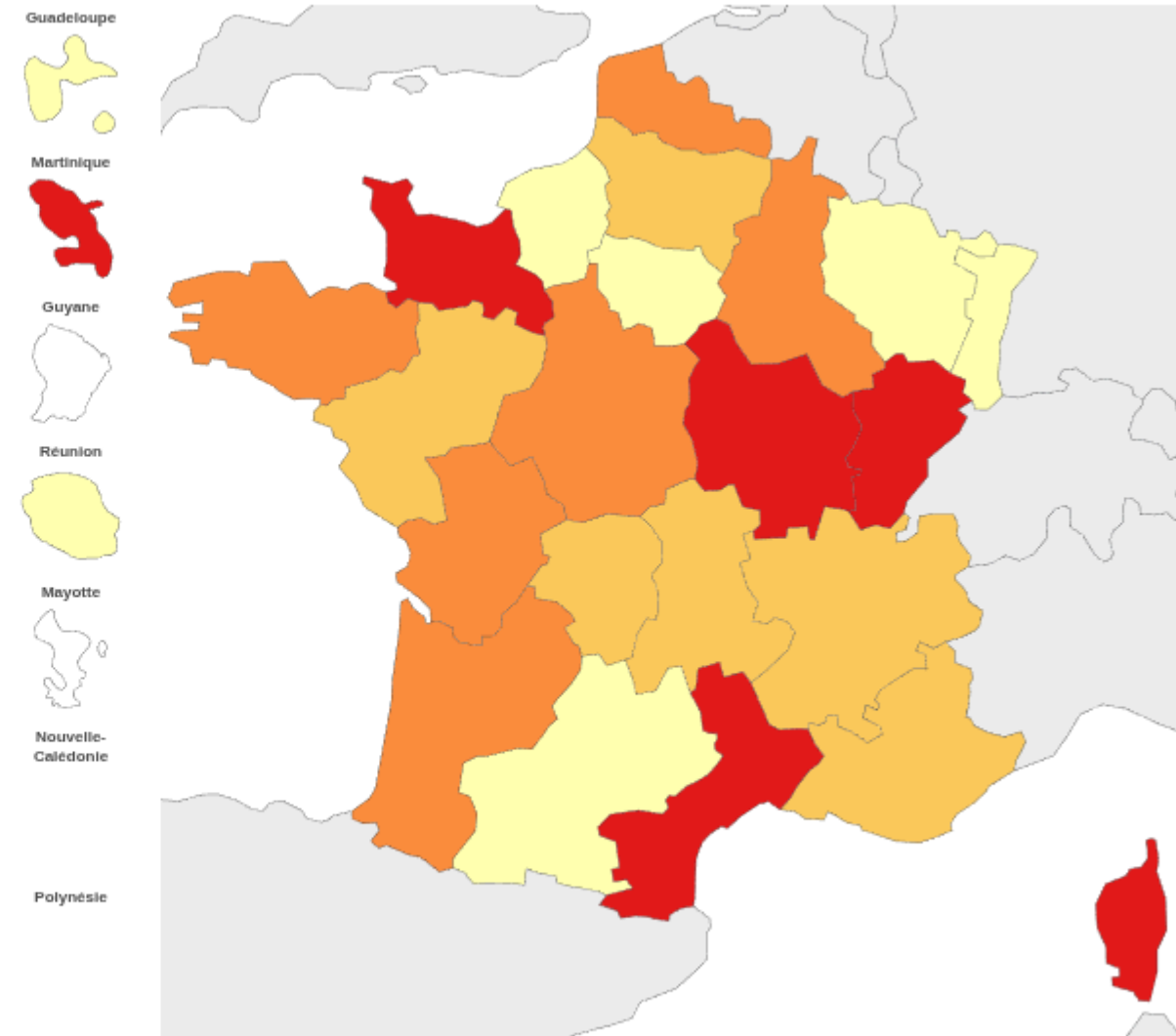
Donneurs en état de mort encéphalique recensés en 2021



Taux de sujet recensés par région (pmh)

■]28.3, 40.0] ■]40.0, 41.9] ■]41.9, 46.0] ■]46.0, 62.1] □ Non calculable

Donneurs en état de mort encéphalique prélevés en 2021



Taux de sujet prélevés par région (pmh)

■]7.7, 18.6] ■]18.6, 21.4] ■]21.4, 25.9] ■]25.9, 33.8] □ Non calculable

CONCLUSION



CONCEPTS-CLÉS (1)

La discrétisation des données de taux est obligatoire en cartographie. Il s'agit d'une limite physiologique, l'œil n'étant pas capable d'associer facilement à plusieurs valeurs de gris plusieurs données.

- On **réduit** donc l'information dans des classes pour que l'œil n'ait qu'un petit (4 à 7) nombre de niveaux de gris à analyser.
- Cette réduction implique une **simplification de l'information statistique**. Pour ne pas fausser le message cartographique, il faut veiller à utiliser une méthode adéquate.
- En s'intéressant au message cartographique (public, format, support).
- En décrivant la série grâce à l'analyse univariée (forme, résumé, dispersion, valeurs extrêmes).

De nombreuses méthodes de discrétisation existent et le choix final dépend évidemment des étapes précédentes.

N'oubliez pas que vous pouvez faire des ajustements manuels sur la discrétisation (bornes des classes) si cela est justifié : **soyez pragmatiques !**

CONCEPTS-CLÉS (2)

Vous ne voulez pas que l'on vous accuse d'avoir manipulé la discrétisation ?

- Utilisez des quantiles (mais vous risquez de ne pas suivre la répartition statistique des données et d'avoir des classes hétérogènes).
 - Q6 permet de conserver les extrêmes dans des classes à part.
-

Vous ne souhaitez pas comparer votre carte à une autre et voulez suivre au mieux la forme de la distribution ?

- Utilisez les seuils naturels avec l'algorithme de Jenks
 - mais vous ne pourrez pas comparer votre carte dans le temps sans ajustement car les bornes des classes des deux cartes seront différentes.
-

Vous devez faire une carte pour le grand public ?

- Privilégiez les amplitudes égales avec si possible une amplitude arrondies (5 en 5, 100 en 100).
- Conservez néanmoins bien le vrai minimum et le vrai maximum.

CONCEPTS-CLÉS (3)

Votre serie de données suit une loi normale et vous souhaitez montrer les individus géographiques « anormaux » ?

- Utilisez la discrétisation en moyenne écart-type.
- Vous pourrez ainsi mettre en évidence les n% individus en queues de distribution

Vous voulez comparer des données de même nature ?

- Appliquez les amplitudes égales sur l'ensemble des séries, puis reportez les mêmes bornes de classe sur les cartes
- Une autre alternative est d'appliquer du Jenks sur l'une des séries et d'appliquer les mêmes bornes de classes sur les autres séries.
- Si la légende n'est pas commune, pensez à appliquer les vrais minimum et maximum.

Vous voulez comparer des données de différentes natures ?

- Appliquez des quantiles/Q6 afin de pouvoir comparer les 20 premiers % individus (pour des quintiles) de la première série au 20 premiers % individus de la seconde série.
- Si vos deux séries sont normales (avec si possible des écart-types proches), vous pouvez aussi utiliser une discrétisation en moyenne écart-type.

EN RÉSUMÉ

La discrétisation des données de taux est obligatoire en cartographie. Il s'agit d'une limite physiologique, l'œil n'étant pas capable d'associer facilement à plusieurs valeurs de gris plusieurs données.

- On **réduit** donc l'information dans des classes pour que l'œil n'ait qu'un petit (4 à 7) nombre de niveaux de gris à analyser.
- Cette réduction implique une **simplification de l'information statistique**. Pour ne pas fausser le message cartographique, il faut veiller à **utiliser une méthode adéquate**.
- En s'intéressant au message cartographique (public, format, support)
- En décrivant la série grâce à l'analyse univariée (forme, résumé, dispersion, valeurs extrêmes).

De nombreuses méthodes de discrétisation existent et le choix final dépend évidemment des étapes précédentes.

**NE SOYEZ PAS PRISONNIER DES STATISTIQUES.
N'OUBLIEZ PAS QUE VOUS POUVEZ FAIRE DES
AJUSTEMENTS MANUELS SUR LA DISCRÉTISATION SI
CELA EST JUSTIFIÉ
SOYEZ PRAGMATIQUES**
