

## **Atelier Archéomatique 17.1 : l'analyse par maille: un outil d'aide à l'analyse et de représentation cartographique quantitative.**

(Sylvain Badey - Inrap / CITERES-LAT)

Elaborer une carte de répartition à l'aide d'une analyse par maille est une opération relativement simple grâce aux logiciels de SIG. Cependant, il est nécessaire de respecter des étapes indispensables qui induisent autant de questionnements et de choix:

- définir la maille (taille et géométrie)
- choisir l'information à représenter (brute ou calculée)
- découper les classes (méthodes de discrétisation)
- cartographier (règles de la sémiologie graphique)

L'atelier est organisé en deux temps :

- Une présentation des concepts, des méthodes et de leur mise en œuvre dans un logiciel SIG (QGIS).
- Une application sous forme d'atelier collaboratif à partir des données apportées par les participants (ou un jeu de données fourni par l'intervenant). Un temps pour la présentation des résultats et une mise en débat de ceux-ci est prévu.

Présentation:

### **Qu'est ce qu'un maillage ?**

“Le maillage c'est une répartition stricte (sans chevauchement ni flou) d'une zone géographique en unités contiguës dont la forme et la taille peuvent être régulières ou irrégulières. C'est une partition de l'espace dont les pièces élémentaires sont des polygones de formes et de surfaces souvent hétérogènes s'emboîtant à la façon de pièces de puzzle.” (Lambert 2016). On parlera de carte choroplèthe quand les mailles sont hétérogènes (répartitions administratives pour les cartes électorales par exemple) et d'analyse par maille ou tessellation quand les mailles seront de même forme (carroyage par exemple).

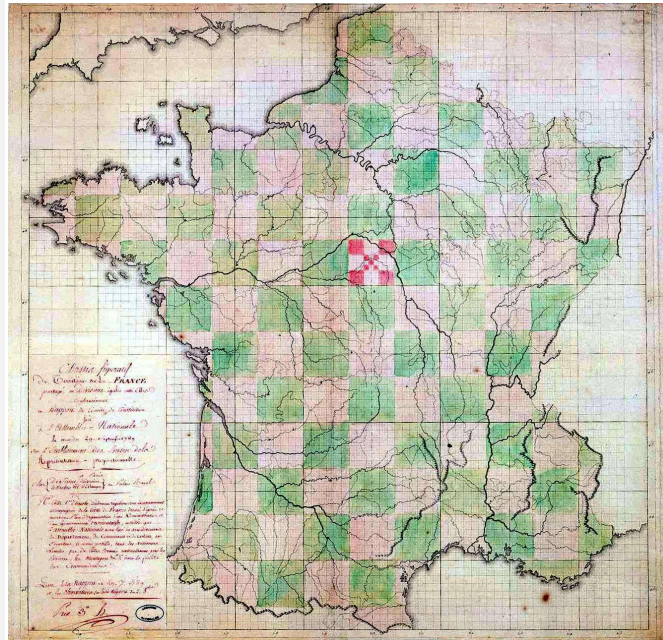
Dans le cadre d'une analyse par maille, choisir le maillage (forme, taille et emplacement) c'est choisir la grille de lecture pour comprendre le phénomène choisi. C'est un canevas qui doit nous aider à analyser la répartition spatiale d'éléments par la généralisation, la simplification de l'information.

Il est évident qu'un changement de forme, de taille ou de positionnement de la maille va influencer les résultats à la fois d'un point de vue statistique mais aussi de la représentation cartographique. C'est pourquoi il est important de bien connaître son objet d'étude et de bien choisir les paramètres de sa maille (taille, forme et position).

### **Pourquoi faire une analyse par maille?**

**Parce qu'on a déjà partitionné le territoire étudié à l'aide d'un carroyage.**

Le carroyage est une technique de division d'un espace géographique en unités spatiales simples très commun. Il est donc parfois utilisé dès la phase d'acquisition des données comme unité de découpage de l'espace étudié.



<http://revolution.1789.free.fr/dossiers-speciaux/decembre-1789/ChassisFiguratif-sieyes-thouret%2029-09-1789.jpg> (sieyes1789.jpg) En 1789 quand l'assemblée nationale décide de diviser la France en départements, plusieurs projets sont proposés. Celui de Lally-Tolendal qui propose de découper le royaume en districts peuplés chacun de 150.000 habitants. L'abbé Sieyès lui, présente son projet qui consiste en une division géométrique du territoire en 89 départements formé chacun de 9 "communes", Paris possédant un régime particulier.

Dès le terrain le carroyage est un moyen simple de division de l'espace à la fois pour se repérer et pour calculer des densités. On l'utilise couramment en écologie pour quantifier la répartition d'espèces (végétales ou animales) mais aussi bien évidemment en archéologie.

En archéologie, le carroyage implanté sur le terrain est encore couramment utilisé pour les périodes anciennes (comptage d'artefacts dans une maille allant du quart de mètres carrés au mètre carré) ou encore pour étudier des secteurs particulièrement denses en mobilier. C'est donc une méthode courante et par conséquent c'est une "grille de lecture" usuelle et compréhensible pour l'archéologue.


exemple image: comptage des morts par carroyage.



[http://www.images-archeologie.fr/userdata/icono\\_fiche/5/5668/670x510\\_5668\\_vignette\\_100-0141.JPG](http://www.images-archeologie.fr/userdata/icono_fiche/5/5668/670x510_5668_vignette_100-0141.JPG) Système de carroyage, Épièdes-en-Beauce (Loiret), 2010. Un carroyage de 1 m de côté a été mis en place sur la totalité de l'enclos (400 m<sup>2</sup>) afin de fouiller la couche de rejet de bûcher funéraire. Il permet la localisation des prélèvements, ainsi que de la céramique en vue des remontages. Les collages entre carrés permettant de comprendre le dépôt.

Présentation du jeu de données d'exemple d'Ozans et du choix de l'analyse par maille. Diagnostic sur une grande surface avec des tranchées mécaniques continues régulières qui induisent une grille de lecture. L'implantation d'un maille a pour objectif de distinguer des concentrations de faits archéologiques en s'affranchissant au maximum de la grille de lecture créée par l'implantation des tranchées.

La ZAC d'Ozans, sur la commune d'Etrechet (36), est un grand projet d'aménagement d'une plateforme multimodale qui devrait au final occuper 800 ha. Ce projet a fait l'objet d'une première phase de prescriptions de diagnostics archéologiques, divisée en cinq tranches, de la part du SRA. Suite à ces diagnostics qui font l'objet de cette étude, trois fouilles ont été prescrites et effectuées en 2011. 412 tranchées ont été ouvertes pour une longueur totale cumulée d'environ 124 km et une surface totale décapée de 248 032 m<sup>2</sup>. Le taux global d'ouverture est donc de 11%.

Ouvrir le projet QGIS:  [Projet\\_Ozans.qgs](#)

Ce projet QGIS est composé:

- d'une image raster de fond: *Ozans\_Bing\_Sat.tif*
- d'une couche de polygone représentant l'emprise du projet d'aménagement: *Emprise\_Projet.shp*
- d'une couche de polygone contenant les emprises des 5 tranches de diagnostics de la 1ère phase ainsi que l'emprise totale du diagnostic: *Emprises\_Diagnostic.shp*
- d'une couche de polygones contenant les contours de tranchées mécaniques (y compris les fenêtres d'extension): *Tranchees\_Diagnostic.shp*
- d'une couche de polygones représentant le contour des faits archéologiques repérés au cours des diagnostics: *Faits\_Diagnostic.shp*

QGIS: présentation de l'interface:

Les couches sont affichées dans la **fenêtre couches** et leur contenus dans le **canevas** (partie centrale de l'interface)

Dans, la fenêtre "Couches" à gauche: Afficher/Désafficher les couches en les cochant/décochant.

Vous pouvez aussi organiser la superposition des couches par glisser-déposer.

**La barre d'état** affiche: les coordonnées de la souris, l'échelle d'affichage (+zoom), l'angle de rotation de la carte par rapport au nord (0 par défaut donc le Nord en haut) et le SCR (Système de coordonnées de référence).

Vérifier le SCR du projet en bas à droite: est-il bien en 2154 (Lambert93)?

**La barre d'outils** contient les outils les plus courant sous forme d'icônes. Pour afficher/désafficher une barre d'outil on fait un clic droit sur celle-ci et on coche/décoche. Déplacer les barres d'outils à votre convenance en cliquant sur les pointillés grisés. : par ex. les outils de contrôle des couches à gauche de l'écran et les outils de navigation à droite.

Pour agir sur une couche (étiquette, style mais aussi SCR, jointures...), généralement on fait un clic droit sur son nom dans la fenêtre "couche" > Propriétés

## Définir la maille: forme, taille et position

### Quelle forme de maille choisir ?

Si l'on choisi de créer une maille régulière et homogène *a posteriori*, la première question est celle de la forme. Deux formes géométriques sont couramment utilisées pour faire des analyses par maille.

#### La maille carré:

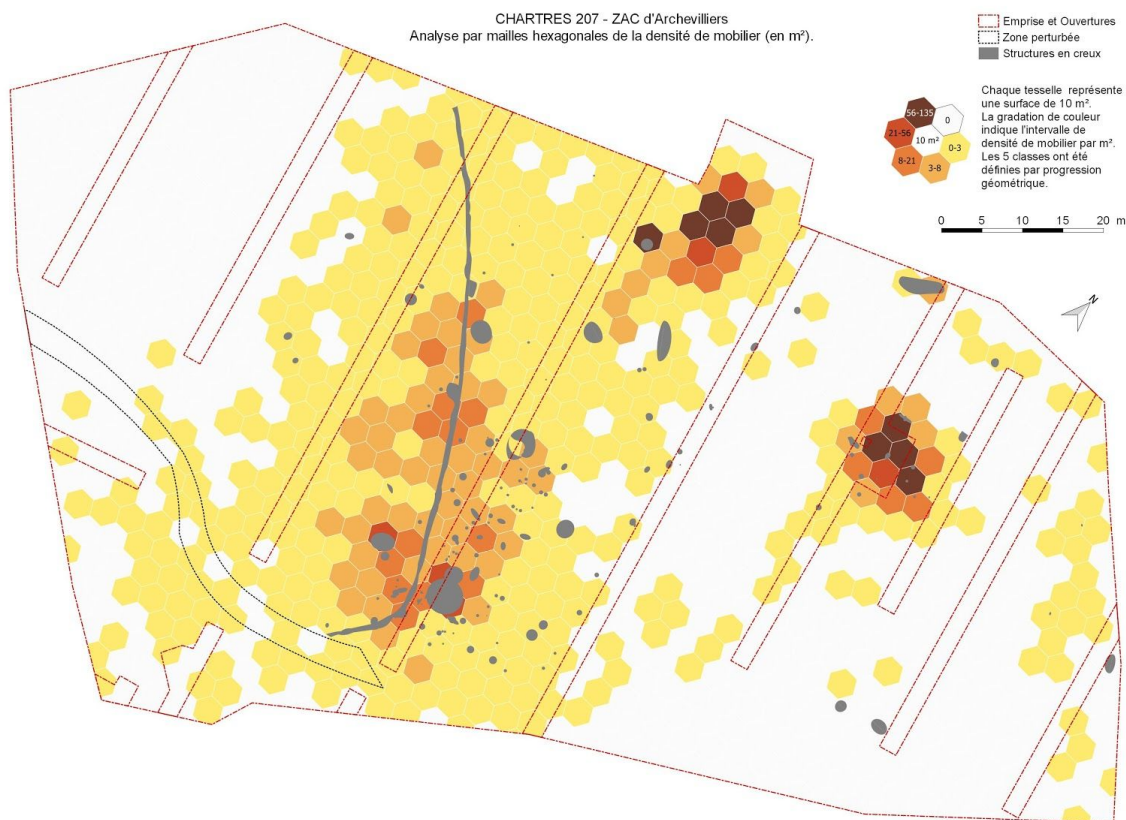
C'est la plus simple et la plus facile à interpréter.

Elle correspond parfois à "une grille de lecture" utilisée sur les terrains archéologiques (carroyages pour prélèvement de silex ou de mobilier dans des zones denses).

#### La maille hexagonale:

Cette forme géométrique comme unité de base pour une analyse par maille n'est quasiment pas utilisée en archéologie alors qu'elle l'est plus fréquemment dans d'autres disciplines comme l'écologie ([exemple1](#),[exemple2](#)), la géographie ([exemple1](#),[exemple2](#),[exemple3](#)).

L'usage d'une maille hexagonale peut être envisagée dans certains cas notamment quand on veut adapter au plus près le maille à l'emprise étudiée qui a une forme complexe.



(CH207\_Maille\_H10.jpeg) Pour l'étude des concentrations de mobilier de la fouille Chartres 207 - ZAC d'archevilliers (RO: M.F. Creusillet) à l'échelle de l'emprise une première analyse par maille a été effectuée sans distinction de type de mobilier avec une maille hexagonale dans laquelle chaque tesselle fait 10m<sup>2</sup>.

En outre, l'hexagone a des propriétés géométriques intéressantes d'une part c'est la forme qui correspond à un *optimum* isopérimétrique c'est-à-dire que le pavage hexagonal est la partition du plan en surfaces égales ayant le plus petit périmètre permettant ainsi un découpage idéal de l'aire étudiée.

D'autre part dans l'avantage d'une maille hexagonale sur la maille carrée traditionnelle est que la distance entre le centre de chaque cellule hexagonale et le centre de tous les six hexagones adjacents est constante. (Par comparaison, dans une grille carrée, la distance du centre de chaque cellule carrée au centre des quatre cellules diagonales adjacentes avec lesquelles elle partage un coin est supérieure à la distance au centre des quatre cellules adjacentes, avec qui elle partage un bord).

Enfin, dans un maillage hexagonal, toutes les cellules voisines partagent toujours des bords; Il n'y a pas deux cellules en contact avec un seul point.

d'après [https://en.wikipedia.org/wiki/Hex\\_map](https://en.wikipedia.org/wiki/Hex_map)

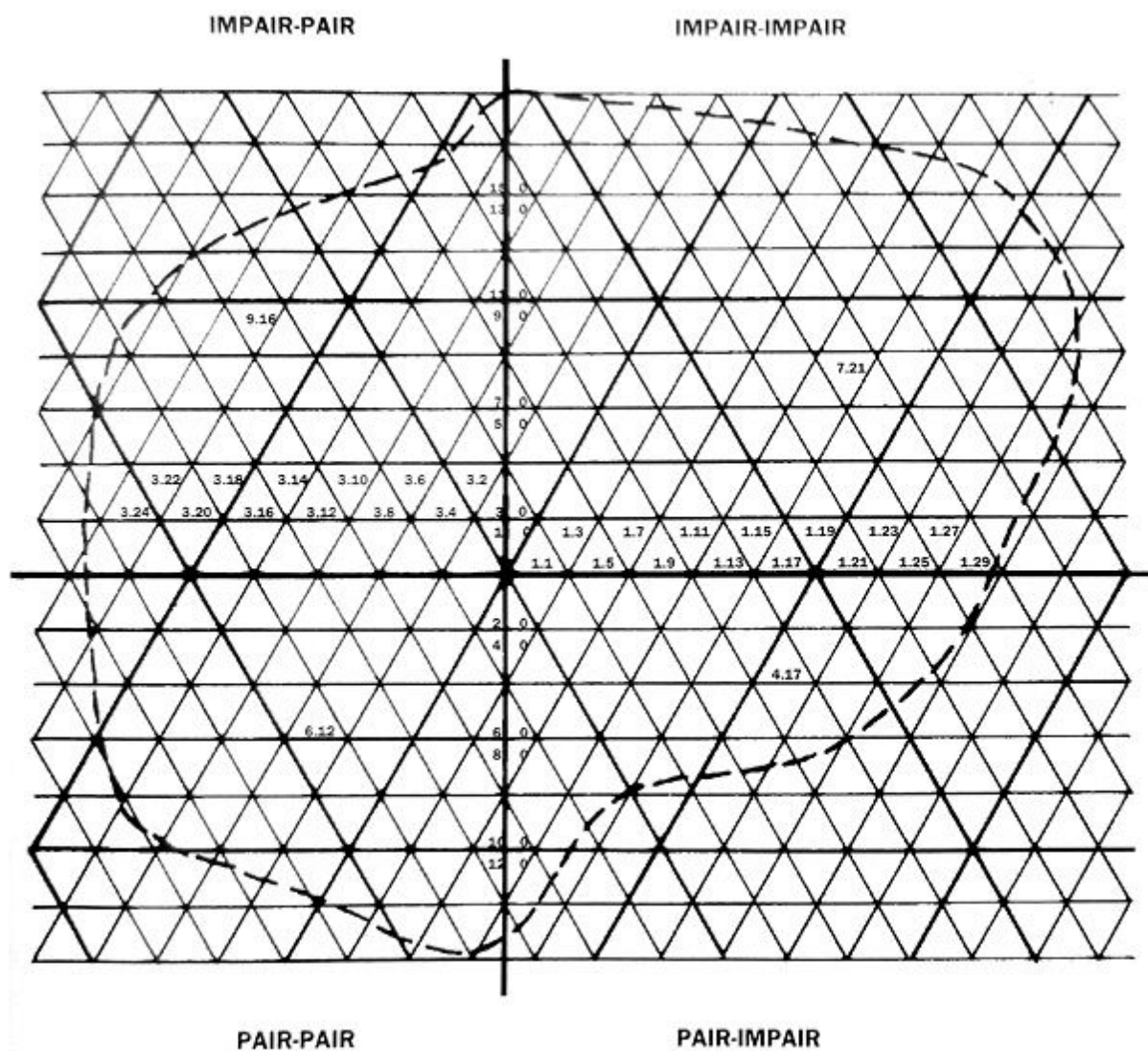
Les analyses par mailles hexagonales sont aussi appelées hex grid ou hex map.

Si l'on utilise une maille hexagonale il sera nécessaire de fixer la taille de la maille sur sa surface de façon à ce qu'elle soit facilement compréhensible/interprétable (10 mètres carrés par exemple) plutôt que sur la longueur d'un côté. On n'utilisera pas de maille hexagonale si l'objet d'étude a été étudié dès le terrain avec une autre maille (carroyage déjà implanté sur le terrain) ou si la grille de lecture usuelle est la maille carrée (locus de silex habituellement étudiés avec des carrés d'1m<sup>2</sup> ou d'1/4m<sup>2</sup>).

### Une autre forme géométrique ?

On évitera d'autres formes géométriques pour effectuer une analyse par maille notamment le rectangle.

Le triangle pourrait être utilisé dans le cas où il aurait servi de système de repérage comme c'est le cas parfois en archéologie subaquatique et d'unité de prélèvement.



[http://www.archeologiesenchantier.ens.fr/IMG/jpg/figure2\\_web.jpg](http://www.archeologiesenchantier.ens.fr/IMG/jpg/figure2_web.jpg)

Repérage triangulaire mis en oeuvre pour la fouille d'installations de l'An Mil du Lac de Paladru (COLARDELLE M. et VERDEL E. 1993). L'unité de base est un triangle équilatéral de 5 mètres de côté divisé en 25 petits triangles métriques. Le repérage se fait à partir d'un point 0 central et d'un axe nord-sud. NDLR: Si une analyse par maille devrait être faite elle pourrait se faire soit sur la base du triangle voire d'un hexagone regroupant 6 triangles équilatéraux.

### Quelle taille de maille choisir ?

“Géo-agrégation et densité: Le calcul d'une densité à partir d'un semis de points impose un changement d'échelle par agrégation dans une surface. L'utilisation d'un carroyage régulier

permet de s'affranchir des biais liés à l'utilisation d'un découpage administratif pour le calcul de densité. Mais plus la distribution des points est concentrée, plus l'influence de la taille de la maille est importante." in SOURIS M., 2011

La taille de la maille est le deuxième paramètre à définir. Définir la taille de la maille c'est définir l'échelle d'analyse, ou par analogie avec les images numériques la résolution de l'image obtenue.

La taille de la maille dépend:

- de l'échelle d'analyse choisie: que veut-on étudier ? A quelle "résolution" ? (On pourra faire plusieurs analyses par maille sur un même terrain d'étude selon que l'on veuille distinguer des concentrations à l'échelle du site ou des organisations dans une concentration de mobilier par exemple cf. CH207)
- de la taille des objets ou du phénomène étudié: elle ne doit pas être trop petite car elle créerait trop de mailles vides ni trop grande car elle simplifierait trop l'information. (Dans l'exemple des tranchées de diagnostic d'Ozans, la maille ne doit pas être plus petite que la distance entre les tranchées car elle ne permettrait pas de s'affranchir de la grille de lecture créé par celle-ci, ni trop grande car elle ne permettrait pas de distinguer de distinguer des concentrations de faits discontinus).
- des usages de la discipline: encore une fois, les préhistoriens ont l'habitude d'étudier des concentrations de mobiliers en  $\frac{1}{4}$  de m<sup>2</sup> ou en m<sup>2</sup>.

### Comment positionner le maillage ?

Dans un logiciel de SIG, la position du maillage se fera automatiquement avec la même orientation que celle de l'affichage cad dans le sens EO-NS.

L'orientation aura assez peu d'importance dans le cas d'une maille hexagonale mais le sera dans le cas d'une maille carrée.

Il n'y a pas de règles définies si ce n'est que l'on va tenter d'adapter au mieux l'orientation du maillage aux objets étudiés afin par exemple d'avoir le minimum de mailles vides et ou découpés par les limites de l'emprise du site.

**Démonstration:** QGIS: création d'un maille avec choix de la taille (longueur du côté ou surface)

Pour faire une maille carrée:



- 1) Menu Vecteur > Outils de recherche > Grille vecteur
- 2) Présent aussi dans la boîte de traitement > Géotraitement QGIS > Outils de création vecteur > Grille vecteur OU Créer une grille (plus d'options mais pas de possibilité de définir selon la surface de la maille voulue)
- 3) Extension MMQGIS > Create > Create Grid Layer (équivalent avec d'autres options mais pas de possibilité de définir selon la surface de la maille voulue)
- 4) Qmarxan (cf. maille hexagonale)

Pour faire une maille hexagonale:

- Installer le plug-in Qmarxan

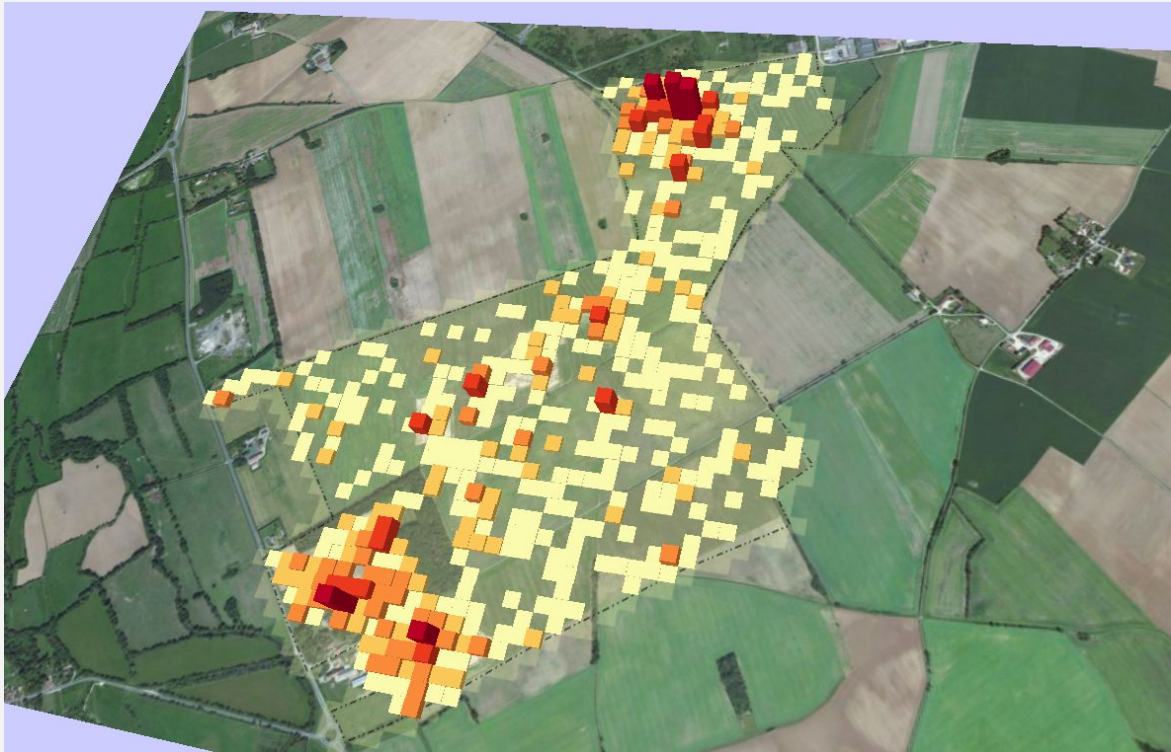
- Boîte de traitement > QMarxan ToolAlgorithms > Data Preparation > Create Planning Unit Layer
- **Note:** l'option "Clip new planning unit layer to this layer [optionnel]" permet de ne créer des mailles que quand elles recouvrent les objets d'une autre couche.

#### Positionner la maille:

Il est toujours possible de créer un maillage puis de le sélectionner afin de le déplacer à la main (à l'aide de l'outil "Déplacer l'entité" ) et même de l'orienter à la main (outil "Pivoter l'entité" )

**Exercice:** QGIS: Création d'une grille avec choix de l'emprise adaptée à l'étendu de la couche d'entités à dénombrer (points ou polygones) ou de la couche d'emprise (polygone)  
**Note:** penser à supprimer les champs inutiles (ouvrir la table attributaire, mettre la couche en mode édition et utiliser l'outil suppression de champs)

#### et une maille 3D ?



(Qgis2threejs.PNG) Répartition des surfaces de faits mis au jour lors d'un diagnostic. La variable hauteur (basée sur la même variable quantitative utilisée pour définir les classes de valeurs) est ici utilisée à la fois pour accentuer visuellement les concentrations et pour pouvoir comparer les valeurs absolues de la variables (différence de hauteur entre deux mailles de la même couleur). par Sylvain Badey.

Ce type de représentation est à éviter si la maille est trop hétérogène (carte choroplèthes surtout) ou quand leur nombre est trop grand.



Il est possible pour accentuer la visualisation des concentrations d'ajouter la variable visuelle taille (la hauteur) qui sert à l'extrusion des polygones de maille. On utilise alors la même variable quantitative que celle mise en classe pour la construction de l'analyse par maille mais en conservant les véritables valeurs.

**les +:**

- on peut visualiser des différences entre deux mailles proches l'une de l'autre de la même classe.
- cela permet de donner rapidement un aperçu de l'ordre de grandeur d'un phénomène spatial.

**le -:**

- l'oeil ne peut pas comparer deux hauteurs éloignées l'une de l'autre.
- à éviter si les mailles sont trop nombreuses ou trop hétérogènes (cartes choroplèthes).

**Facultatif:** QGIS + extension **Qgis2threejs**: création d'un maillage 3D à partir d'un raster de fond et de la maille:

- installer l'extension Qgis2threejs (elle se trouve dans le menu Internet)
- supprimer toutes les mailles vides (NUMPOINTS = 0)
- Internet > Qgis2threejs
- DEM > Display > Layer image, choisir *Ozans\_Bing\_Sat*
- Polygone : ne cocher que la couche de maillage
- puis, Style > Height, choisir le champs "NUMPOINTS" et éventuellement un facteur d'agrandissement: 3
- Bouton [Run]

## Choisir l'information à représenter: brute ou calculée

### Quelles valeurs utiliser ? Qu'est ce que l'on compte ?

Après avoir défini la maille (forme, taille, position) il est temps de s'attaquer à l'information à représenter dans la maille. Quel est l'objet d'étude et comment le répartir dans chaque maille ? Par un simple dénombrement ou par une valeur calculée ?

### Le carroyage adressé:

La manière la plus simple et la plus courante est le simple dénombrement d'objet à l'intérieur de chaque maille. C'est la méthode que l'on utilise quand l'objet d'étude est représenté sous la forme de points. Si ce n'est pas le cas, généralement des polygones, il est possible d'opérer une transformation de géométrie (généralement on transforme le polygone en point en calculant son centroïde).

QGIS: Ozans: transformation des polygones en centroïdes puis comptage des points dans la maille.

**Créer une couche de points d'après le centroïdes de polygones:**

Menu Vecteur > Outils de géométrie >  Centroides de polygones

Note: On retrouve cet algorithme dans la Boîte à outils de traitements sous l'intitulé "Centroides de polygones"

Note: On retrouve cet algorithme dans la Boîte à outils de traitements sous l'intitulé "Décompte de points dans un polygone (pondéré)"

Regarder le résultat dans la table attributaire.

Il est aussi possible de récupérer de l'information attributaire comme valeur à comptabiliser dans chaque maille. On pourrait par exemple non pas représenter le nombre de mobilier dans une maille mais son poids total. Dans ce cas cette valeur doit forcément être une valeur quantitative absolue cad que la somme de ses valeurs doit avoir un sens, (par exemple le poids, la taille, etc.) *a contrario* d'une valeur quantitative relative dont la somme n'aura pas de sens (par exemple un taux de fragmentation ou un pourcentage quelconque).

QGIS: Ozans: Dans ce cas on utilise la jointure spatiale c'est à dire que l'on va récupérer dans chaque maille (polygone) un résumé (somme, minimum, etc..) des attributs des points qui sont contenus dans celle-ci.

**Attention pas de colonne quantitative pour l'exercice !**

Menu Vecteur > Outils de gestion de données >  Joindre les attributs par localisation

Note: On retrouve cet algorithme dans la Boîte à outils de traitements sous l'intitulé "Joindre les attributs par localisation"

Il suffit de:

- Choisir la Couche vecteur cible = Couche dans laquelle on veut récupérer la donnée = la couche de maillage.
- Choisir la Couche vecteur à joindre = la couche de points qui contient les données.
- Choisir le prédicat géométrique = la relation topologique de la couche cible par rapport à la couche à joindre = contient
- Choisir le résumé de l'attribut = ici, "Prendre un résumé des entités intersectées"
- Choisir les statistiques que l'on veut récupérer = écrire "sum"

Attention! cette option ne sera pas prise en compte si on a choisi "Prendre les attributs de la première entité localisé" dans l'option précédente !

- Choisir de garder (ou non) les mailles ne contenant pas de points = "Conserver tous les enregistrements" Ou "Ne conserver que les enregistrements correspondants"
- Enfin, choisir le dossier et le nom d'enregistrement de la couche créée.

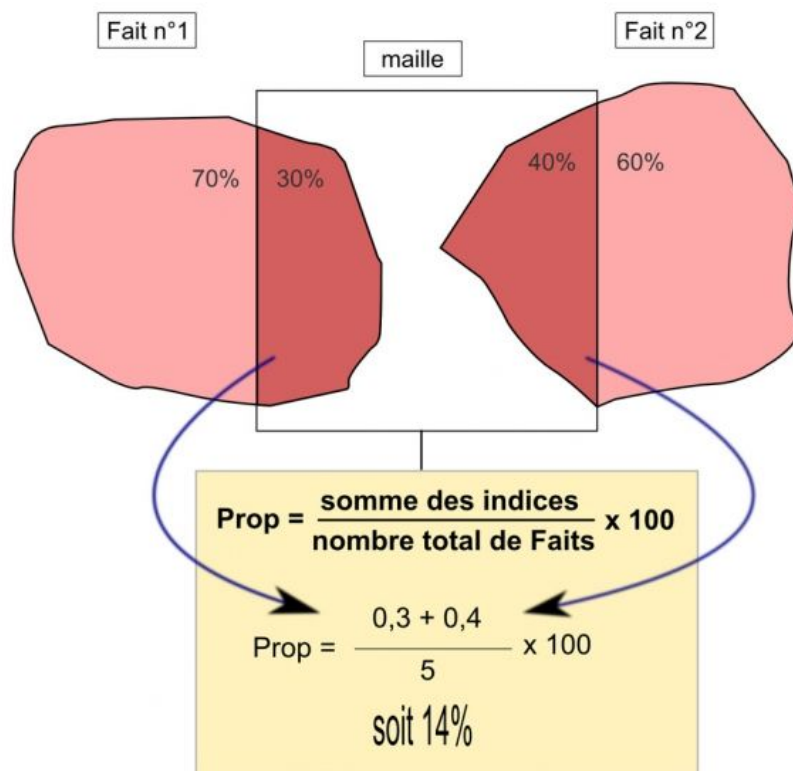
### Le carroyage ventilé:

Dans certains cas, le "simple" carroyage adressé cad "rapatrier" le nombre d'entités géographiques dans chaque maille ou une valeur de celle-ci n'est pas pertinent.

C'est le cas quand le maillage découpe les objets géographiques étudiés et que ceux-ci **intersectent** (*recoupent*) plusieurs mailles. C'est le cas dans l'exemple du diagnostic d'Ozans, notamment avec des structures archéologiques de grande taille comme certaines

fosses ou linéaires (murs, chemins, fossés). On ne peut évidemment pas comptabiliser les objets étudiés plusieurs fois il faut donc faire un choix: soit les adresser dans une seule maille soit les répartir selon un critère.

Le cas le plus courant consiste à découper les objets étudiés (sous forme de polygones) par la maille utilisée puis d'utiliser la somme des surfaces comme valeur ou de s'en servir pour pondérer une autre valeur quantitative propre à chaque objet d'étude.



<https://racf.revues.org/docannexe/image/2202/img-4-small580.jpg>

Pour l'analyse par maille du diagnostic d'Ozans l'objectif était de représenter les concentrations de faits archéologiques représentés "classiquement" sous forme de polygones. Il a été choisi de représenter dans chaque maille une quantité relative de surface de faits cad un pourcentage de surface de faits rapporté au nombre total de faits. Cet indice permettant de pouvoir comparer des analyses par mailles par périodes entre elles malgré la représentation différentielle du nombre de faits archéologiques selon les périodes (Badey S. 2014).

**Facultatif:** QGIS: Ozans: découper les faits avec la maille puis faire un calcul (lequel?) avec la calculatrice de champs.

Calculer la surface de chaque faits:

- avec la calculatrice de champ, créer un champs "surface" (nombre décimal, 10, 10) et utiliser la fonction \$area

Découper les faits selon la maille:

- Problème avec l'outil de QGIS à cause d'erreurs topologiques dans la couche de faits
- Utiliser: Traitement > SAGA > Vector polygon tools > Intersection
- Layer A = faits, Layer B = grille, enregistrer sous *faits\_intersect.shp*
- Décocher split part
- avec la calculatrice de champs, créer un champ "surf\_inter" (nombre décimal, 10, 10) et utiliser la fonction \$area
- avec la calculatrice de champs, créer un champ "indice" et mettre la formule "surf\_inter/surface"

Récupérer la donnée dans chaque maille avec une jointure par localisation:

- Vecteur > Outils de gestion de données > Joindre les attributs par localisation
- Couche cible = grille, Couche à joindre= faits\_intersect, Prédicat= intersecte, "Prendre un résumé..", sum, "ne conserver..." (ou pas)
- enregistrer sous *maille\_intersect.shp*
- supprimer les champs inutiles (toutes les sum.. dont on n'a pas besoin, garder seulement *sumindice*:  
en mode édition > icône "supprimer un champ"

Créer l'"indice de proportion de faits":

- avec la calculatrice de champs créer un champ "prop" (nombre décimal 10, 2) avec la formule *sumindice/2246\*100*

QGIS: Ozans: Représenter le résultat, premier essai "presse-bouton"

Clic droit sur la couche de maillage créée > Propriétés > Style > Gradué

Puis choisir:

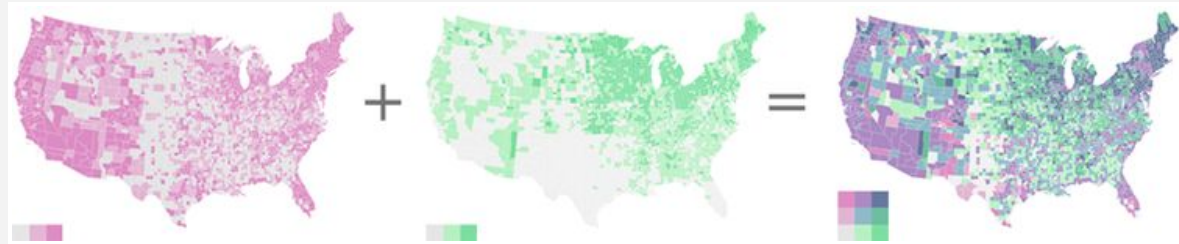
- le champs sur lequel faire la graduation (NUMPOINTS)
- un dégradé (*Inferno* ou *plasma* par exemple ;)
- le mode de partitionnement de classe (Laisser Intervalles Egaux)
- le nombre de classes
- Bouton [Classer]
- Bouton [Appliquer] et [OK] si le résultat est satisfaisant

Remarques: la carte est peu lisible car noyée par les faibles valeurs. Cela est dû à:

- une classe qui commence à 0, on a donc attribuer une même valeur de couleur pour une maille ne contenant aucun fait et une maille en contenant un peu ! => C'est un problème de représentation, heureusement il existe des règles de sémiologie graphique !
- On a laissé le mode "Intervalles égaux" qui partitionne notre série statistique (le tableau contenant chaque maille et le nombre de faits que chacune contient) sans se soucier de la distribution de celle-ci ! => C'est un problème statistique, il faut choisir une mode de partitionnement, une méthode de discrétisation en rapport avec la distribution de la série et ce que l'on cherche à voir/ à montrer avec notre carte.

## Et une analyse par maille bi-variée, c'est possible ?

Il est également possible de "compiler" deux analyses par maille (ou carte choroplèthe) univariées en une seule. Ce type de représentation a pour objectif d'aller plus loin que ce que montrerai les 2 analyses univariées séparément. L'objectif est alors de chercher/montrer où ces 2 variables ont tendance à être concordantes ou à l'inverse discordantes.

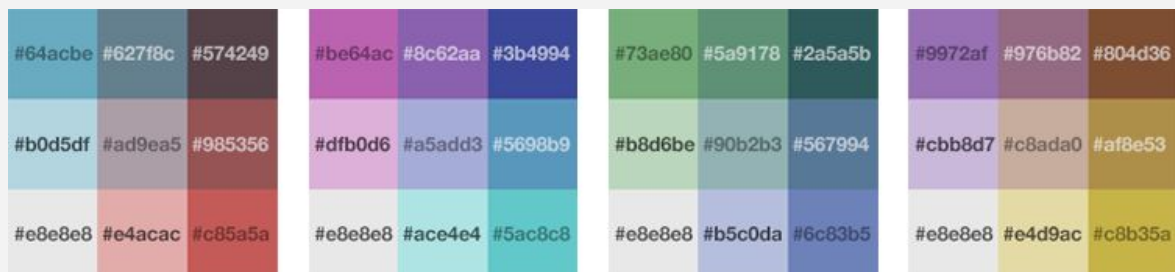


[http://www.joshuastevens.net/images/js\\_choroMath.png](http://www.joshuastevens.net/images/js_choroMath.png)

Ici, Joshua Stevens agrège 2 cartes choroplèthes univariées avec 3 classes chacune en une carte choroplèthe bivariable à 9 classes.

Il sera alors nécessaire:

- de vérifier la pertinence de la compilation des deux variables.
- de limiter le nombre de classes pour chaque variable (3 classes par variables idéalement) tout en vérifiant leur pertinence.
- de bien choisir le double camaïeu de couleur.



[http://www.joshuastevens.net/images/js\\_bivariatePalettes.png](http://www.joshuastevens.net/images/js_bivariatePalettes.png)

Propositions de palettes de couleurs pour une carte choroplèthe bivariable par Joshua Stevens.

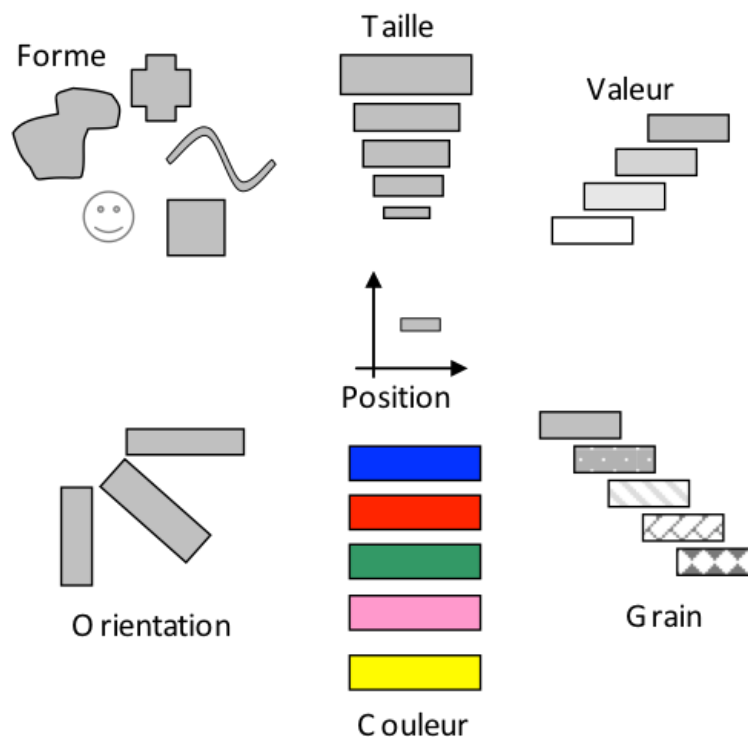
La maille ayant été choisie, la manière d'y rapporter une valeur correspondant aux objets étudiés aussi, reste à faire des choix quant à la représentation cartographique: quelle valeur représenter ? combien de classes de valeurs représenter ? définir les classes à représenter ?

Cette étape s'appuie sur plusieurs principes: les règles de la sémiologie graphique et la description statistique de la distribution de nos valeurs.

## Cartographier: les règles de la sémiologie graphique

## La sémiologie graphique: c'est quoi ?

La "Sémiologie Graphique" c'est le titre du livre édité en 1967 par Jacques Bertin (1918-2010), un cartographe français et théoricien. Sur la base de son expérience en tant que cartographe et géographe, ce livre présente le premier et le plus complet fondement théorique de ce que nous appelons aujourd'hui «Visualisation de l'information" ou *dataviz*. Bertin fournit une théorie universellement reconnue des symboles et des modes de représentation graphique qui permettent cartographes de transformer l'information géographique en communication cartographique graphiques. Son travail est basé sur un système clair et logique dans lequel symboles peuvent varier en fonction six variables rétiniennes, qui comprenait la taille, le grain, l'orientation et la forme.



<http://www.cartographie-semantique.fr/wp-content/uploads/2012/11/Variables-visuelles-d%C3%A9finies-par-Jacques-Bertin.png> Les six variables rétinienne définies par Jacques Bertin dans sa "sémiologie graphique"

Ses 6 variables variables visuelles sont appliquées aux types d'implantations: le point, la ligne, la surface et vont être utilisées en fonction de la nature de la donnée à représenter.

En effet les données, de part leur nature peuvent permettre d'exprimer des relations d'association, de sélection (différence), d'ordre ou de quantité.

*(semio\_utilisation\_variables.gif, [kartoweb.nl](http://kartoweb.nl) ?) tableau de synthèse d'utilisation des 6 variables rétinienne pour une représentation cartographique. Remarquez que seule la variable taille permet d'exprimer une quantité brute (absolue) alors que la variable valeur permet d'exprimer soit une sélection soit un ordre.*

### **La variable rétinienne valeur**

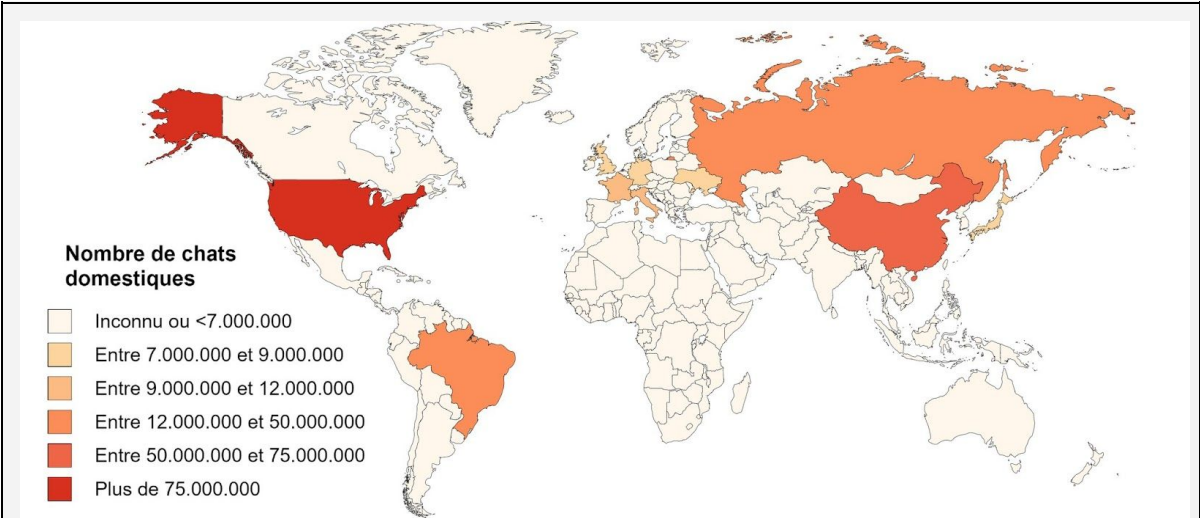
La variable visuelle valeur est le rapport entre la quantité de noir et de blanc perçu sur une surface donnée. C'est donc une gradation en niveau de gris s'échelonnant du gris clair au noir. Le blanc est généralement réservé pour indiquer l'absence d'information.

Cette variable peut aussi être utilisée par l'ajout de noir dans une couleur (teinte) pour former un camaïeu.

Il ne faut pas confondre la couleur (la teinte) qui permet de représenter des associations ou au contraire des différences avec la valeur qui représente un ordre, une hiérarchie.

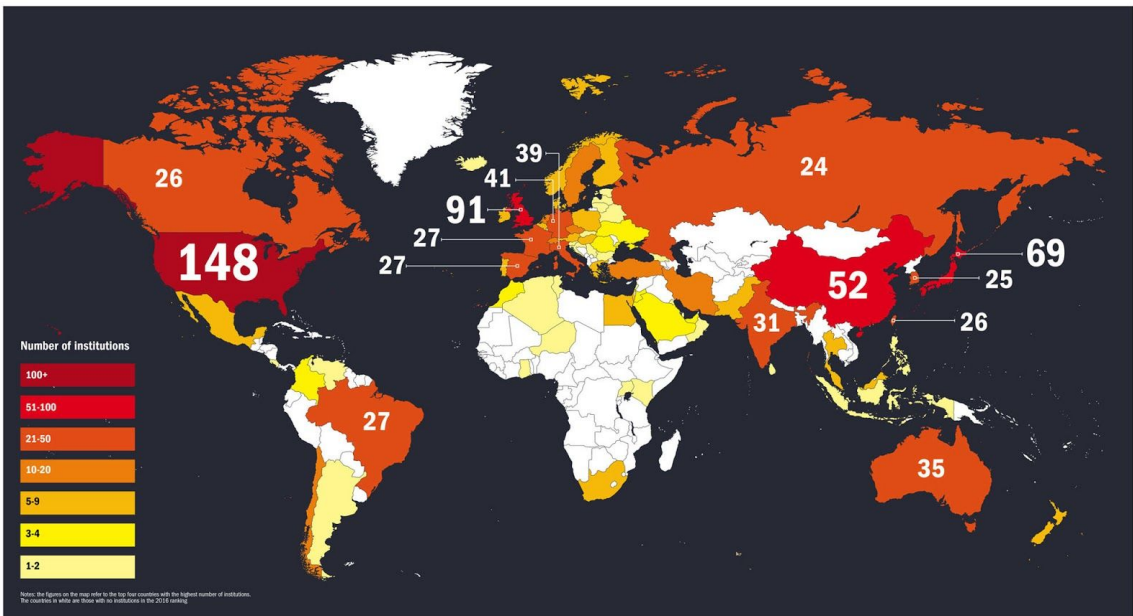
Elle est très efficace pour les implantations zonales (les polygones), et peut s'appliquer aux données qualitatives ordonnées ou aux données quantitatives relatives.

**Pourquoi ne peut-on pas mettre une valeur absolue dans une carte choroplèthe?**



<https://pbs.twimg.com/media/CzodFvbXUAAjKUu.jpg:large>

Cette carte représentant le nombre de chats domestiques (variable quantitative absolue) par pays est fautive du point de vue de la sémiologie graphique: en effet il faudrait rapporter le nombre de chats à la surface ou au nombre d'habitants pour obtenir une valeur quantitative relative (une densité ou un pourcentage de chat par milliers d'habitants par exemple). On appelle cela un #mapfail



**THE TIMES HIGHER EDUCATION WORLD UNIVERSITY RANKINGS**  
www.thewur.com

[https://assets.weforum.org/editor/bJgAvENIqYu-Fdj2d\\_w9ABYN311xslBbobka7eRxFoc.jpg](https://assets.weforum.org/editor/bJgAvENIqYu-Fdj2d_w9ABYN311xslBbobka7eRxFoc.jpg)

Ici encore une fois on a représenté une variable quantitative absolue (le nombre d'université par pays) avec la variable rétinienne valeur. La France et la Russie ont la même couleur alors que si on rapportait ce nombre par habitants ou à la superficie la hiérarchisation visuelle ne serait pas du tout la même. #mapfail

En utilisant une maille régulière l'utilisation d'une variable quantitative absolue (comptage de point dans un carré par exemple) est équivalent à un calcul de densité puisque chaque



maille a la même surface. On préférera cependant utiliser une unité exprimant la densité (nombre de points au mètre carré par exemple).

Outre le fait que l'on ne peut représenter qu'une valeur quantitative relative avec la variable rétinienne valeur, la sémiologie graphique nous impose aussi:

- **l'usage de la seule variable rétinienne valeur** et non de la variable rétinienne couleur qui n'exprime pas un ordre mais une différence.

Note: Évidemment, la teinte la plus claire représentera les valeurs les plus faible et la teinte plus foncée les valeurs fortes.

Note: il est possible (et courant) néanmoins d'utiliser la variable couleur pour exprimer un ordre en utilisation la gradation qui correspond au spectre de la lumière. On partira alors du milieu du spectre (le jaune) et on utilisera les couleurs soit dans le sens des ultraviolets ou dans celui des infrarouges.

- **un nombre limité de classes** pour que le lecteur puisse facilement les différencier (7 maximum).
- **le blanc** sera utilisé préférentiellement pour signifier **l'absence d'information** du phénomène étudié.

QGIS: Ozans: adapter le style de notre représentation cartographique aux règles de la sémiologie graphique:

- choisir un camaïeu de couleur une seul teinte de plus plus foncée (ou correspondant au spectre chromatique) par ex *YIOrRd* ou mieux *Reds*
- mettre 5 classes
- ajouter une classe pour l'absence d'information (NUMPOINTS = 0) et ne pas mettre de remplissage voire pas de contour).
- Changer la borne inférieure de la première classe non vide à 1.
- [Appliquer] et [OK]

Pour la suite on choisit de supprimer toutes les mailles à valeurs nulles:

- Ouvrir la table attributaire
- Outil sélection selon une expression 
- Taper l'expression SQL: "count" IS NULL
- rentrer dans le mode édition  et supprimer les entités sélectionnées 

Remarque: là cela commence à ressembler a quelque chose même si la carte reste un peu fade. Faire un clic droit sur la couche > cocher "montrer le décompte des entités". On comprend bien qu'une écrasante majorité des carrés contient seulement entre 1 et 17 faits (et même seulement 1 fait). Le mode de découpage des classes en "intervalles égaux" n'est pas satisfaisant.

**L'usage d'une double gradation pour exprimer une opposition: négatif/positif.**

Il est possible de faire une double gradation seulement si l'on veut exprimer une relation d'opposition (valeurs négatives/valeurs positives, bien/pas bien). Dans ce cas, cette opposition doit toujours se référer à une valeur qui a un sens



<https://static.openfoodfacts.org/images/misc/nutriscore-a.svg> Un exemple très récent avec le logo 5-C conçu par le professeur Serge Herberg (Inserm), qui donne une note globale aux aliments vendus en supermarchés à partir des éléments à favoriser et de ceux à limiter. Ici on n'utilise pas la variable rétinienne valeur mais une double gradation de la variable couleur ordonnée selon le spectre visible de la lumière en partant du jaune (centre du spectre) d'un côté vers l'ultraviolet (les verts) et l'infrarouge (orange et rouge) pour exprimer une opposition. En outre le créateur joue sur la dimension symbolique et culturelle des couleurs: vert = c'est bien, rouge = pas bien ☺

## Découper les classes de valeurs: la discrétisation

### Le découpage en classes de valeurs: la discrétisation.

Nous avons une série statistique composée d'entités (les mailles) qui ont chacune une valeur. Ces valeurs sont des densités, c'est à dire le rapport entre un nombre et une surface. C'est une données quantitative continue, c'est à dire que sur une échelle donnée elle peut prendre n'importe quel nombre (généralement un chiffre avec des virgules et des décimales derrière), par exemple il peut très bien y avoir 3,98354 faits dans les 2500 m<sup>2</sup> d'une maille.

La discrétisation est le procédé qui consiste à transformer une donnée quantitative continue en variable ordonnée, discrète en définissant des classes de valeurs. On parle aussi de découpage en classes.

Plus simplement l'objectif de la discrétisation est de simplifier l'information d'une série statistique pour la rendre lisible sur une carte.

Pour cela il faut se pencher sur la série statistiques en passant par une série d'étapes:

- résumer la série statistique
- l'analyser
- discrétiser (définir le nombre de classes et les bornes de classes)

### Résumer la série statistique

Résumer une série statistique permet de se l'approprier. Connaître les valeurs minimales et maximale, la moyenne arithmétique des valeurs, quelles sont les valeurs les plus récurrentes sont des moyens quasi-intuitifs quand on se retrouvent face à une série statistique qui permettent de l'appréhender.

Pour résumer une série statistique il existe deux types de paramètres:

- Les paramètres de position ou comment résumer la série par une seule valeur ?
- Les paramètres de dispersion ou comment se situe une valeur par rapport aux autres ou par rapport à une valeur centrale ?

**Les paramètres de position** d'une série statistiques sont composés de valeurs spécifiques (généralement le minimum et le maximum) et de valeurs centrales, calculées ou observées (moyenne, médiane mode).

Le minimum et le maximum sont deux valeurs importantes pour la compréhension d'une série. Leur différence est appelée étendue et dans le cas de petites séries est parfois le seul indicateur valable.

QGIS: Ozans: Résumer l'information de notre série statistique

- cliquer sur l'icône "Montrer le résumé statistique": un panneau "statistiques" apparaît au dessus du panneau "couches".
- choisir le nom de la couche et le champs (NUMPOINTS) sur lequel calculer les statistiques basiques.

Note: l'outil statistique de QGIS nomme l'étendue = *plage*, le mode = *majorité*

La moyenne notée  $\bar{x}$  est le centre de gravité d'une série statistique c'est la valeur calculée la plus courante (parfois à tort) pour la résumer par une seule valeur.

Elle se calcule en divisant la somme de toutes les valeurs par l'effectif.



**Albert**

@amisdalbert



Suivre



## Les limites d'une moyenne : si Bill Gates entre dans un bar, en moyenne, tout le monde y est millionnaire.

*teasing* pour une vidéo de la chaîne youtube "les statistiques expliquées à mon chat" intitulée pourquoi gagnez vous moins que le salaire moyen. <https://t.co/AJOG2djDf5>

La médiane notée  $Q_2$  est la valeur qui sépare la série statistique en deux parties d'effectifs égaux. C'est la valeur la plus proche de l'ensemble des valeurs de la distribution. Elle est moins utilisée que la moyenne alors qu'elle est parfois plus parlante d'autant plus si elle est associée à la moyenne.

La médiane se repère en triant la série statistique par ordre croissant et en prenant la valeur prise par l'individu correspondant à la moitié de l'effectif (dans le cas d'un nombre impair) ou c'est la moyenne entre les deux valeurs prises par les individus

Le mode ou valeur dominante est la valeur la plus fréquente d'une distribution. Dans une série statistique il se peut qu'il n'y ait pas de mode, qu'il y en ait qu'un (la distribution est dite unimodale), deux (bimodale) ou plusieurs (plurimodale).

Le mode se calcule en faisant un dénombrement des modalités c'est à dire pour chaque valeur de la série, combien de fois elle apparaît.

Les **paramètres de dispersions** sont toujours liés à une valeur centrale. Ils permettent d'évaluer la dispersion générale par rapport à celles-ci et de placer la valeur de chaque individu par rapport aux autres.

L'écart-type noté  $\sigma$  est le plus courant, il se réfère à la moyenne. Il est issu d'un autre calcul, celui de la variance: "La variance c'est la moyenne de la somme des carrés des écarts à la moyenne".

"La variance c'est la moyenne de la somme des carrés des écarts à la moyenne" ?!...

On se la fait à l'envers:

- On calcule la moyenne.
- Pour chaque individu on calcule la différence entre sa valeur et la moyenne, on obtient donc une valeur exprimant l'écart à la moyenne. Comme on va se retrouver avec des écarts négatifs et des écarts positifs, on les mets au carré (c'est une technique mathématique bien connue pour supprimer les chiffres négatifs).
- On fait donc la somme de tous ses "écarts au carré"
- On divise le résultat par l'effectif pour obtenir une moyenne ou plus exactement "la moyenne de la somme des carrés des écarts à la moyenne" ou encore la variance (c'est moins long)

Le problème c'est que l'unité de mesure de la variance c'est celle de départ mais au carré ! Alors on en calcule la racine carré et on obtient l'écart-type !

L'intervalle inter-quartile ou *IQR* c'est  $Q_3 - Q_1$  ! Or la médiane c'est  $Q_2$  donc c'est bien qu'il existe un  $Q_1$  et un  $Q_3$  ? En effet de la même façon que l'on calcule la médiane qui divise la série statistique en deux parties d'effectifs égaux, on peut faire la même chose en divisant la série en 4 parties d'effectif égaux. Ainsi le premier quartile noté  $Q_1$  sera la valeur en dessous de laquelle se situe  $\frac{1}{4}$  de l'effectif et les  $\frac{3}{4}$  restant auront une valeur supérieure. Le troisième quartile ou  $Q_3$  c'est l'inverse:  $\frac{3}{4}$  des individus ont une valeur inférieure à  $Q_3$  et  $\frac{1}{4}$  des individus des valeurs supérieures.

Si vous avez compris, vous comprenez que 50% des individus ont une valeur entre  $Q_1$  et  $Q_3$ . Mais la meilleure façon de le comprendre c'est de les visualiser sur un graphique dédié: la boîte à moustache.

R & Rstudio: Présentation des logiciels et de l'interface de Rstudio

### Préparer les données avec QGIS:

- ouvrir la table attributaire et supprimer tous les champs sauf "puid" (Identifiant) et "count" (comptage de faits)
- clic droit sur la couche de maille > Enregistrer sous... > Valeurs séparées par une virgule [CSV]

### Ouvrir Rstudio.

- présentation de l'interface:
- 4 panneaux avec chacun des onglets
- définition du Répertoire du travail
- import d'un csv

Importer le fichier csv à l'aide du bouton [Import DataSet] > à partir d'un CSV

Note: acceptez toutes les propositions d'installation avec le bouton [OK] si nécessaire

- structure d'un tableau (str)

Dans le panneau console tapez str(maille50)

- nouveau script
- historique
- envoi depuis l'historique vers script
- exécution du script
- principe des variables objets avec <-

R & Rstudio: Ozans: Résumer l'information de notre série statistique (calculer les paramètres de position et de dispersion)

### QGIS:

- moyenne

```
mean(maille$count, na.rm=TRUE)
```

- médiane
- dens <- maille\$count

```
median(dens)
```

- étendue

```
max(dens)-min(dens)
```

- écart-type

```
sd(dens)
```

- IQR

```
quantile(dens)
```

- boîte à moustache

### **Analyser la distribution**

Théoriquement en ayant calculé les valeurs correspondants aux paramètres de position et de dispersion, on a déjà une idée de la distributions des données. Mais le plus simple pour analyser la distribution d'une série c'est de s'aider de représentations graphiques pour évaluer la forme de la distribution.

La première représentation graphique qui va nous permettre d'analyser la distribution de notre série statistique est **la boîte à moustache** qui va nous permettre d'analyser la

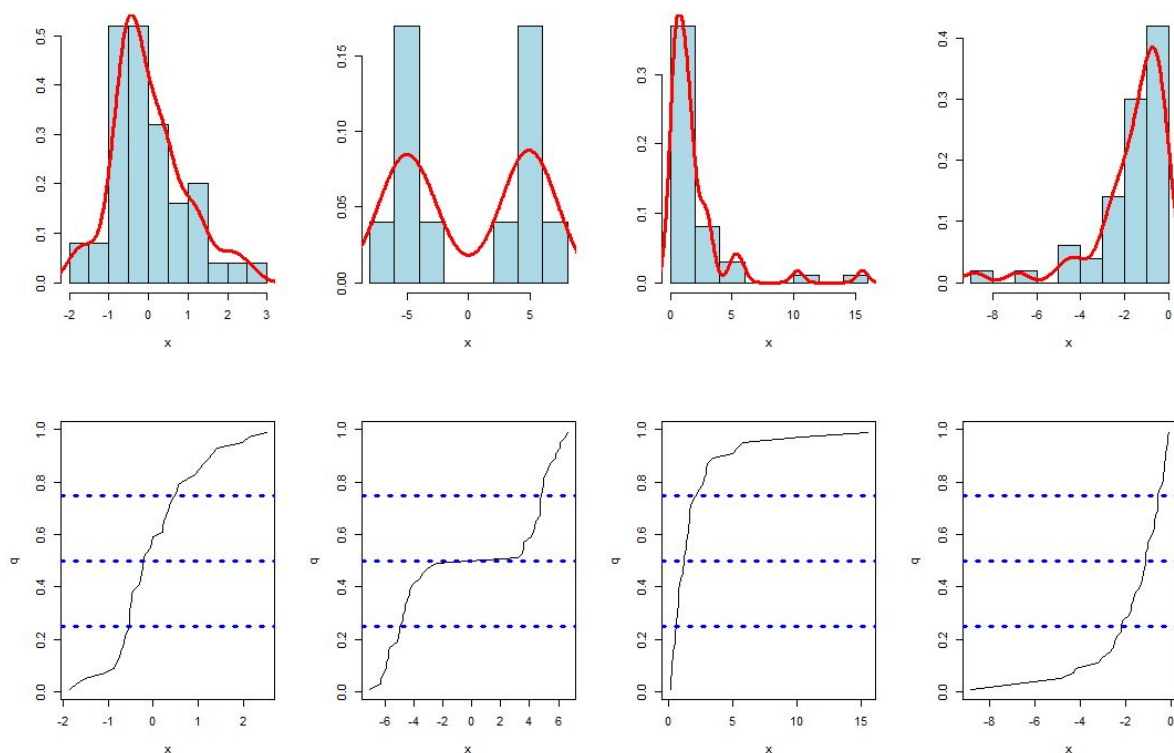
distribution des valeurs autour d'une valeur centrale (la médiane) et des valeurs de dispersion (les écarts-types).

Une autre représentation, très simple à comprendre est le **scalogramme**, dans lequel on indique en abscisse toutes les valeurs et en ordonnées le nombre d'occurrences pour chacune d'entre elles.

Enfin on peut aussi faire un premier histogramme (différence avec diagramme)

Il convient maintenant de définir la distribution de notre série statistique est elle normale, bi-modale ou pluri-modale ou encore dissymétrique ?

### Les formes de distributions



[\(distribution\\_formes.png\)](#) exemple de 4 distributions (de gauche à droite: normale, bimodale, dissymétrique à gauche, dissymétrique à droite). représentées en haut avec des histogrammes (classes d'amplitudes égales) et courbes de densités et en bas avec des courbes de fréquences cumulées.

### distribution uniforme:

Elle est caractérisée par le fait que toutes les valeurs possibles de la variable ont des fréquences égales. Distributions assez rares.

Si l'analyse d'une distribution issue d'une analyse par maille donnait une distribution uniforme, cela indiquerait une répartition régulière ou totalement aléatoire dans l'espace de l'objet d'étude à l'échelle de la maille.

### **distribution normale:**

Le plus grand nombre d'individus se trouvent dans les classes centrales, ce nombre s'amenuise progressivement de part et d'autre de la valeur moyenne.

moyenne=médiane.

En archéologie, les phénomènes suivant une loi normale sont rares.

### **distribution bimodale ou plurimodale:**

Elles correspondent à des distributions où la variable est en fait composée de sous-populations ayant chacune son ordre de grandeur et sa dispersion propre.

Dans ce cas, la plupart des paramètres statistiques (moyenne, écart type...) sont sans signification et sans utilité; seuls des graphiques permettent une analyse correcte.

### **distribution asymétriques:**

Elles traduisent une concentration des individus, plus ou moins accentuée, vers les petites valeurs (dissymétrique à gauche, médiane<moyenne) ou les grandes valeurs (dissymétrique à droite, médiane>moyenne).

Ce type de distribution est très courante en archéologie.

## **Discrétiser**

Exprimer (carto-)graphiquement une information implique souvent sa simplification. Ainsi en découpant en classes une série de données, la discrétisation réduit une variable quantitative en variable ordonnée.

*«Perdre le moins possible d'information, du point de vue de la statistique, reviendrait à cartographier toutes les valeurs, sans discrétisation préalable. Or, nous savons que les règles de la perception visuelle contredisent cette exigence, la lisibilité de l'image cartographique étant d'autant plus grande que le nombre de signes distincts, donc de classes, est petit. Il est donc indispensable de procéder à une «réduction» de l'information statistique.» (BEGUIN & PUMAIN, 1994)*

La discrétisation c'est le fait de découper une série statistique en classes de valeurs. L'objectif, rappelons le est de simplifier/généraliser l'information pour la comprendre et la représenter cartographiquement.

## Combien de classes de valeurs ?

Il existe des formules mathématiques permettant d'estimer un nombre maximum de classes d'après l'effectif mais elles sont souvent adaptées à des distributions normales et généralement elles donnent un nombre de classes supérieur au seuil de perceptions des couleurs. Nous nous tiendrons donc aux règles de la sémiologie graphique et nous ne dépasseront pas 6 classes de valeurs.

## Les règles de la discrétisation

La réduction statistique, appelée discrétisation, doit conserver au mieux l'information tout en la simplifiant. Ces objectifs, a priori contradictoires, peuvent être conciliés en respectant certaines règles:

- Indiquer les descripteurs clé d'une distribution. L'expression graphique fait surtout ressortir les inégalités, les variations dans l'espace. Pour leur interprétation correcte, il convient d'indiquer certains des descripteurs suivants: moyenne, mode, minimum, maximum et écart-type.
- Respecter la forme de la distribution. La plupart des séries de données ressemblent à un type de distribution: uniforme, symétrique (normale), dissymétrique (lognormale ou non). La méthode de discrétisation tiendra compte du «modèle» de distribution que la série épouse.
- Appliquer le principe de ressemblance / dissemblance. Toute discrétisation s'efforce de minimiser la variance intraclasse tout en maximisant la variance interclasse.

## Les règles:

- Les classes doivent couvrir l'ensemble de la distribution et elles doivent être contiguës (jointives).
- Une valeur ne doit appartenir qu'à une classe et une seule.
- Les classes ne doivent pas être vides.
- Les valeurs limites doivent être précises et rapidement appréhendables.
- Éviter de placer dans deux classes distinctes des valeurs non significativement différentes.
- Ne pas définir des seuils avec un nombre de décimales supérieur à celui de la précision des données.

## Les méthodes de discrétisations

QGIS: Ozans: Représenter le résultat, premier essai "presse-bouton"

Clic droit sur la couche de maillage créée > Propriétés > Style > Gradué

Puis choisir:

- le champs sur lequel faire la graduation (NUMPOINTS)
- un dégradé (YlOrRd pour la suite = une gradation monochromatique)
- le mode = méthode de discrétisation) (Laisser Intervalles Egaux)
- le nombre de classes = 5



- Bouton [Classer]
- Cliquer sur l'onglet [Histogramme]
- Cliquer sur le bouton [Charger les valeurs]

Note: malgré le nom ce n'est pas réellement un histogramme... puisqu'en ordonnées c'est le nombre d'occurrence et non pas la densité d'effectif.. mettre un maximum de barres: 100

Note: c'est donc bien un diagramme et non pas un histogramme :c'est la hauteur de la barre qui compte et pas sa surface..

**Observer les bornes de classes en même temps que les slides...**

### **La méthode des seuils observés (ou seuils naturels)**

Cette méthode consiste à observer la distribution d'une série statistique et à choisir comme bornes des classes les trous de la distribution (scalogramme ou histogramme). On peut aussi identifier des ruptures dans la distributions sur une courbe de fréquence cumulée.

#### **Les + :**

Elle est simple à mettre en œuvre si des regroupements apparaissent clairement dans la distribution.

Elle permet de mettre en évidence des valeurs extrêmes ou des regroupements spécifiques.

#### **Les - :**

Comme elle est construite sur les discontinuités de la série statistique, le risque est de produire des classes surchargées et inégales.

C'est la méthode la plus subjective est donc la plus à même de permettre d'orienter la représentation cartographique !

dans l'onglet [Histogramme] tenter de repérer les discontinuités dans la série statistique.

### **La méthodes des intervalles égaux (ou amplitudes égales)**

On divise l'étendue (max-min) par le nombre de classes choisies.

Dans notre exemple  $83/5$  soit des classes d'amplitude 16,6

#### **Les + :**

Facile à lire et à comprendre.

Permet de valoriser les valeurs extrêmes.

#### **Les - :**

Quand la distribution est dissymétrique les effectifs (ici le nombre de carrés) sont très déséquilibrés.

Dans notre exemple, la première classe représente 425 mailles alors que les valeurs extremes sont représentées dans 3 classes différentes !

De la même façon il faut faire attention car certaines classes peuvent être vides !

Ne Permet pas de comparer des cartes.

**Observer les bornes de classes en même temps que les slides...**

### **La méthode des quantiles (ou effectifs égaux)**

On essaye de construire des classes qui ont, dans la mesure du possible, le même nombre d'individus. Les classes s'appellent alors des quantiles (et s'il y a 4 classes des quartiles, 5 classes, des quintiles)

Dans notre exemple il ya 445 mailles / 5 classes = 89. Le but va être de s'approcher au plus près 5 classes de 89 mailles mais attention une valeur ne doit appartenir qu'à une seule classe !

**Les + :**

Permet de comparer des cartes sur des notions d'ordre et non plus de valeurs: ce n'est plus le nombre de faits par maille qui compte mais *"telle classe représente 20% de la surface étudiée contenant les valeurs les plus faibles..."*.

**Les - :**

Noie les valeurs extrêmes.

Observer les bornes de classes en même temps que les slides...

**La méthode standardisée (ou écarts-types)**

Avec cette méthode c'est l'écart-type (à la moyenne) qui permet de calculer l'amplitude des classes. C'est donc l'écart à la moyenne que l'on représente ce qui peut être utile pour pointer du doigt les zones de fortes concentrations par rapport a un « bruit de fond ».

La moyenne est donc une borne de classe ou le centre d'une classe. Les autres bornes seront des fractions ou multiplications d'écarts-types.

**Les + :**

Méthode idéale pour les comparaisons.

Quand on a compris le principe, la carte est facilement compréhensible : on met en valeur les écarts à la moyenne.

Comme pour jenks, les données sont prise en compte.

**Les - :**

Malheureusement elle s'adapte mal aux distributions dissymétriques car les valeurs extrêmes tirent la moyenne vers le bas (ou le haut).

Pas toujours comprise par le lecteur (la légende automatique de QGIS par exemple affiche pour chaque classe le nombre d'écart-type mais pas les valeurs correspondantes, il faut les rajouter à la main)

Attention le nombre de classes ne correspond pas toujours à celui demandé en effet, c'est l'écart-type qui entraine l'amplitude et le nombre de classes. Dans notre exemple, 5 classes correspondent a des paliers de 2 écarts-types mais si on en veut plus le logiciel en créera automatiquement 9 comme dans l'exemple ci-dessous.

Observer les bornes de classes en même temps que les slides...

**La méthode par progression géométrique**

Dans cette méthode, l'amplitude des classes augmente rapidement en fonction d'une progression géométrique.

Méthode utile pour les distributions asymétriques qui comportent une très forte représentation des faibles valeurs et pour les distributions exponentielle et logarithmique décroissantes. Il s'agit de distributions souvent rencontrées en archéologie.

La progression de l'amplitude des classes se fait à raison de  $r = 10$  ( $r = 10$  exposant  $[(\log(\text{valeur max}) - \log(\text{valeur min})) / \text{nb classes}]$ )

Et borne 0 = n (n = valeur min)

Borne 1 = n x r

Borne 2: (n x r) x r

Borne 3: ((n x r) x r) x r

**Attention:** ne peut être appliquée quand la valeur minimale est à 0 !!

Cette méthode de calcul n'est pas présente dans QGIS:

- il faut donc soit la calculer à l'aide d'un tableur (on clacul la raison puis les bornes) soit avec R grâce au package cartography et sa fonction getBreaks.

- récupérer les valeurs de bornes et les rentrer à la main dans QGIS.

Observer les bornes de classes en même temps que les slides...

### La méthode des ruptures naturelles (ou algorithme de Jenks)

C'est une méthode automatique et itérative.

La méthode des seuils naturels revient à décréter que les plus fortes discontinuités observées dans la distribution constituent des paliers marquant qu'il faut séparer en classes.

Cette méthode se rapproche de celle des seuils observés mais c'est un algorithme qui travaille ! L'algorithme de jenks minimise la variance dans les classes et maximise la variance entre les classes.

#### Les + :

Préserve les caractéristiques des données puisque l'algorithme cherche les ruptures dans la distribution des données.

Le résultat est donc toujours harmonieux.

Pour un plan de travail c'est souvent un bon choix visuellement pour du vite fait bien fait mais il faut comparer les résultats du logiciel avec l'observation de la distribution.

#### Les - :

Chaque logiciel donnera un résultat différent.

Plusieurs cartes ne peuvent pas être comparées car les classes sont définies d'après la distribution particulière à chaque ensemble de données.

Observer les bornes de classes en même temps que les slides...

exemple en ligne: [Géovisualisation des discrétisations : une petite application pédagogique](#)

exemple en ligne: [discretisatoR](#)

Logiciels utilisés : liste indicative

- QGIS version 2.14 LTR <http://qgis.com/>
- R (<https://cran.r-project.org/bin/windows/base/>) et RStudio (<https://download1.rstudio.org/RStudio-1.0.136.exe>)

Bibliographie / Webographie: en vrac

OPENSHAW S., 1981, Le problème de l'agrégation spatiale en géographie. Traduit de l'anglais par PUMAIN D., L'Espace Géographique, Tome X, n°1, Doin, pages 15-24.

#taille de la maille #découpage en unités spatiales #agrégation spatiale

LAJOIE G., 1992, Le carroyage des informations urbaines. Une nouvelle forme de banque de données sur l'environnement de Grand Rouen. Rouen, Publications de l'université de Rouen, Collection Nouvelles Données en Géographie, N°177, 238 pages.

#origine de la maille #semis de point #taille de la maille

BARBARINO N., 2005, De la qualité de vie au diagnostic urbain, vers une nouvelle méthode d'évaluation: Le cas de la ville de Lyon. Thèse de doctorat soutenue à l'université Lumière - Lyon 2. *partie 2.2: Entre exigences scientifiques et contraintes géographiques : recherche heuristique d'un système de référence géographique.*

[http://theses.univ-lyon2.fr/documents/getpart.php?id=lyon2.2005.barbarino\\_n&part=91453](http://theses.univ-lyon2.fr/documents/getpart.php?id=lyon2.2005.barbarino_n&part=91453)

#taille de la maille #découpage en unités spatiales #agrégation spatiale #origine de la maille #semis de point

BADEY S., & RODIER X., 2014, Exploitation des données de diagnostics en tranchées mécaniques par l'analyse spatiale, *Revue archéologique du Centre de la France* [En ligne], Tome 53 | 2014

<http://racf.revues.org/2202>

#archéologie #exemple #analyse par maille #maille hexagonale #adressage ventilé

MOREAU G., & LOCHT J. L., 2016, Importance de l'utilisation du Système d'Information Géographique pour les vastes sites de plein air du Paléolithique moyen en France septentrionale. L'Exemple de Caours (Somme, France) et Beauvais (Oise, France).

<http://hdl.handle.net/2268/203471>

#archéologie #exemple #poster #analyse par maille

SOURIS M., 2011, Module SIG-Santé, 12.SIG et analyse spatiale. Outils et méthodes pour l'analyse spatiale. Support de cours pour le Master "Géographie de la santé" de Paris Ouest Nanterre-La Défense.

[www.savgis.org/cours/12\\_MethodesSIGpourAnalyseSpatiale\\_2012.pptx](http://www.savgis.org/cours/12_MethodesSIGpourAnalyseSpatiale_2012.pptx)

#taille de la maille #agrégation spatiale #semis de point

<http://www.sciencespo.fr/cartographie/khartis> pour faire des cartes choroplèthes en ligne

<http://mappemonde.mgm.fr/119geov1/> Géovisualisation des discrétisations : une petite application pédagogique par JEGOU L. #discretisation #classes

<https://archeomatic.shinyapps.io/Discretisator/> Application de discrétisation par [archeomatic](#)

#discretisation #classes

<http://magrit.cnrs.fr/modules> Application en ligne de cartographie quantitative (et thématique) par CNRS - UMR RIATE. Attention ne semble fonctionner qu'à petite échelle (monde?)

<http://www.joshuastevens.net/cartography/make-a-bivariate-choropleth-map/> tuto en anglais très complet pour faire une carte choroplète bi-variée avec pas à pas de la démarche et cameaux d'exemple. #choroplète #bi-varié

<http://www.cartographie-semantique.fr/etat-de-lart/la-cartographie/principe-de-linformation-cartographique/> Principes de la représentation cartographiques, un article synthétique mais complet. #sémiologie graphique #cognition externe #variables visuelles