

**Mieux faire parler vos données statistiques
sur une carte :
méthodes et conseils pratiques**



mercredi 4 avril 2012



- Eric Mauvière - 26 chemin de Bellevue - 31140 Pechbonnieu - France

www.geoclip.fr - e.mauviere@emc3.fr - tél : +33(0)5 34 27 45 17

Présentation générale

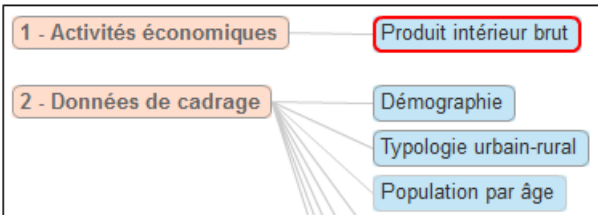
Le but de cet atelier est de visualiser des cartes thématiques, d'en comprendre les paramètres de construction et de jouer avec eux.

Etape n°1	Afficher de une à deux cartes en vis-à-vis avec l'Observatoire des Votes en France (OVF)
Etape n°2	Les deux grands types d'analyse thématique : choroplèthe ou à symboles proportionnels
Etape n°3	Paramétrer une analyse choroplèthe : classes et couleurs
Etape n°4	Paramétrer une analyse choroplèthe : définir des seuils de classes avec les quantiles

Pause

Etape n°5	Paramétrer une analyse choroplèthe : les autres méthodes de discrétisation
Etape n°6	Comprendre les limites d'une analyse choroplèthe. Réduire le bruit

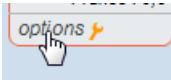
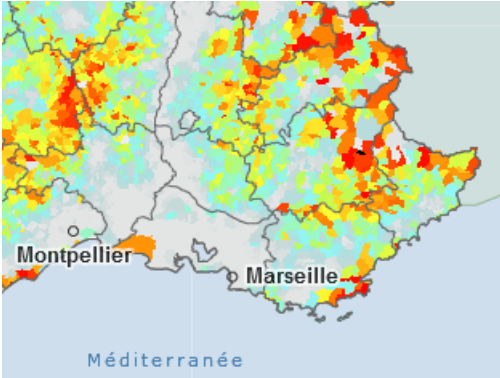
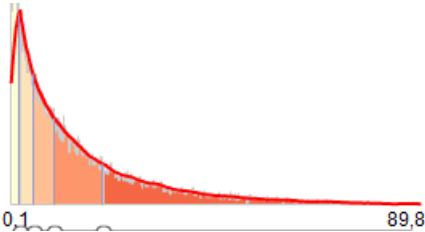
Etape 1 - Afficher de une à deux cartes en vis-à-vis avec l'Observatoire des Votes en France

Démarche	Manipulation
Accéder à l'outil	- Connectez-vous avec un navigateur sur le site http://franceo3.geoclip.fr/ovf
Ouvrir le panneau de choix d'indicateurs	- Cliquez sur le bouton « Voir toute la liste » - ou « Ajouter un autre indicateur » <div style="text-align: center;"> >> VOIR TOUTE LA LISTE Ajouter un autre indicateur... </div>
Parcourir l'arborescence des domaines et thèmes	- Double cliquer sur un thème pour accéder à la liste des indicateurs du thème Choisissez un thème :  <p>Détails de l'arborescence : - 1 - Activités économiques → Produit intérieur brut - 2 - Données de cadrage → Démographie, Typologie urbain-rural, Population par âge</p>
Faire une recherche par mot-clé	- Taper un mot-clé (ex : pib, bac...) Rechercher : <input style="width: 150px;" type="text"/> x <input type="checkbox"/> compatible avec cette carte
Filtrer par type d'indicateur	- Afficher seulement les ratios, ou seulement les indicateurs de type « Additif » - <input type="checkbox"/> Typologies <input checked="" type="checkbox"/> Ratios <input type="checkbox"/> Additifs <input checked="" type="checkbox"/> Ponctuels
Changer d'échelon géographique	- Cliquer au-dessus de la carte sur le bouton « Vues... », puis choisir une autre vue, avec un autre échelon géographique. <div style="text-align: center;"> Vue 1/10 : France métropolitaine par commune ▼ </div>
Afficher une ou plusieurs cartes	- Cliquer sur le bouton « 2 cartes » <div style="text-align: center;"> > 2 cartes </div>
Réinitialiser l'interface	- Ctrl + clic sur le logo

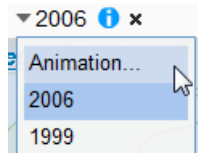
Etape 2 - Les deux grands types d'analyse thématique : choroplèthe ou à symboles proportionnels

Démarche	Manipulation
Comprendre les règles de choix	<ul style="list-style-type: none"> - La représentation par symboles - La représentation par coloration des unités géographiques (choroplèthe)
Superposer deux analyses complémentaires	<ul style="list-style-type: none"> - Choroplèthe : <i>densité de population</i> - Symbolique : <i>population</i> (Thème Données de cadrage >> Démographie) <p style="margin-left: 20px;">densité de population, 2009 (hab/km²) population, 2009</p> <ul style="list-style-type: none"> - <i>Masquer et réactiver chaque analyse par clic sur son titre</i>
Comparer les deux représentations pour la même grandeur observée	<ul style="list-style-type: none"> - La représentation la plus simple porte sur les chiffres bruts (nombre, solde...) - Le calcul d'un ratio, d'un % ou d'une densité produit un indicateur dérivé, qui facilite les comparaisons, mais masque les effectifs sous-jacents - <i>Exemple avec le vote Balladur / Chirac au 1^{er} tour de la Présidentielle 1995</i> <div style="margin-left: 20px;"> <pre> graph LR A[3 - Élection présidentielle] --- B[Présidentielle 2007] A --- C[Présidentielle 2012] A --- D[Présidentielle 2002] A --- E[Présidentielle 1995] style E stroke:#f00,stroke-width:2px </pre> </div> <ul style="list-style-type: none"> - http://franceo3.geoclip.fr/ovf/#i=xd_pres1995_t1.solde_chirac_balladur - http://franceo3.geoclip.fr/ovf/#v=map12;i=xd_pres1995_t1.dif_chirac_balladur

Etape 3 - Paramétrer une analyse choroplèthe : classes et couleurs

Démarche	Manipulation
<p>Représentation continue et classification en tranches</p>	<ul style="list-style-type: none"> - Quand l'intensité d'une couleur doit traduire l'importance d'un chiffre, on peut choisir une correspondance continue, ou une discrétisation en tranches. - Exemple de représentation continue, avec le taux de résidences secondaires : http://franceo3.geoclip.fr/index.php?profil=FD#i=log_ev.pt_nb_rsecocc <p>Cliquer dans la légende le texte « options »</p>  <p>puis l'onglet avancé et cocher la case :</p> <p><input type="checkbox"/> Coloration continue</p>  <p>Limites : cette représentation ne fonctionne bien que si la plage des valeurs est très large, avec beaucoup de valeurs faibles et une minorité de valeurs élevées.</p> 
<p>Intérêt du découpage en classes</p>	<ul style="list-style-type: none"> - Construire des cartes plus lisibles, avec des zones se distinguant clairement les unes des autres - Lisser les aléas statistiques - Faciliter les comparaisons temporelles. - Exemple avec l'indice de vieillissement : http://franceo3.geoclip.fr/ovf/#i=pop_age5h.ind_vieillissement

Dans la liste des années disponibles, choisir « Animation »



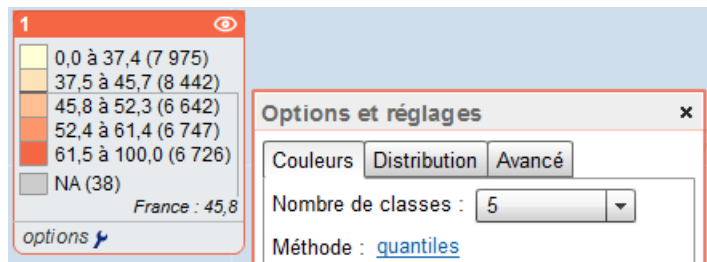
Les classes restent les mêmes, quelle que soit l'année, on voit très bien les communes passer d'une classe à une autre au fil du temps.

Sortir du mode « Animation » en fermant le panneau :



Faire varier le nombre de classes

- Cliquez le bouton « Options » dans la légende et modifiez le nombre de classes :



Jouer avec les palettes de couleurs

- Changez les couleurs par un clic :

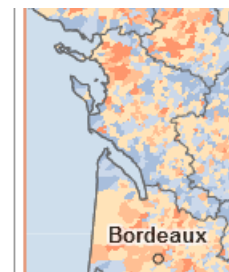


Comprendre la notion de seuil de divergence

- Indicateur avec des valeurs positives et négatives : *taux d'évolution de la population dû au solde naturel*

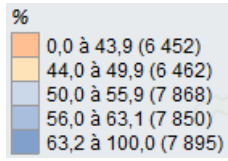
http://franceo3.geoclip.fr/ovf/#i=pops_ev.tvar_nat

- Les palettes de couleur à deux teintes soulignent mieux la différence de signe

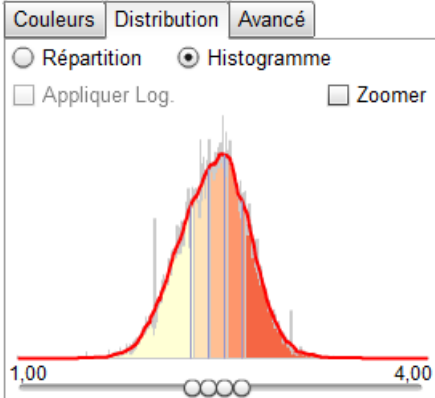



- Un seuil de divergence peut valoir 50 (votes) ou 100 (taux standardisé). Exemple avec le vote Chirac au second tour de la Présidentielle 1995 :

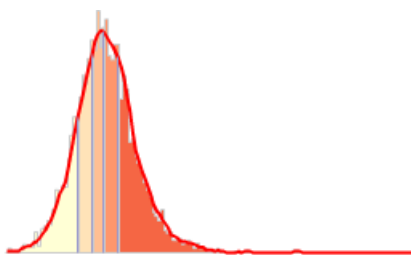
http://franceo3.geoclip.fr/ovf/#i=xd_pres1995_t2.pt_chirac



Etape 4 - Paramétrer une analyse choroplèthe : détermination des seuils de classes avec les quantiles

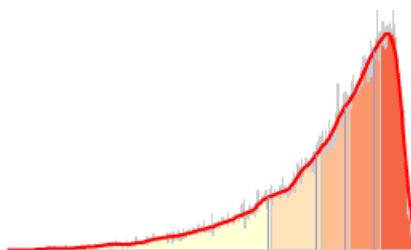
Démarche	Manipulation
<p>Le nombre de classes recommandé</p>	<p>Plutôt proportionnel au logarithme du nombre d'objets. Mais idéalement ne dépasse pas 7 pour préserver la lisibilité. 5 est un bon compromis par défaut.</p>
<p>Les différentes méthodes de calcul des seuils ont une efficacité qui dépend de la forme de la distribution statistique</p>	<p>Les méthodes qui partent de la moyenne des valeurs (écarts-type, moyennes emboîtées...) ne fonctionnent bien que si la distribution est gaussienne et/ou si la moyenne des valeurs est proche de la valeur moyenne. Or de telles distributions se rencontrent rarement en analyse spatiale.</p> <p>C'est la raison pour laquelle on privilégie souvent des méthodes plus robustes comme le calcul des quantiles ou l'algorithme de Jenks-Fischer (seuils naturels).</p> <p>Quelques exemples de distribution :</p> <ul style="list-style-type: none"> - Centrée, exemple avec le nombre moyen d'occupants par résidence principale par commune : http://franceo3.geoclip.fr/index.php?profil=FD#i=fam_ev_men.nb_moy_res_princ <div style="text-align: center;">  </div> <ul style="list-style-type: none"> - Asymétrique à gauche, exemple avec le taux de natalité par commune : http://franceo3.geoclip.fr/index.php?profil=FD#i=etat_civil.taux_natalite;l=fr;s=2010;v=map12 <div style="text-align: center;">  </div> <p>Dans ce cas, l'application d'un logarithme est en général efficace pour normaliser la distribution :</p> <p><input type="checkbox"/> Appliquer Log.</p>

La forme de la distribution dépend du maillage géographique. L'indicateur ci-dessus, projeté sur un découpage cantonal donne une distribution un peu moins serrée :



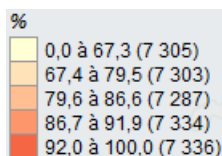
- Asymétrique à droite, exemple avec la part des résidences principales par commune :

http://franceo3.geoclip.fr/index.php?profil=FD#i=log_ev.pt_nb_rp



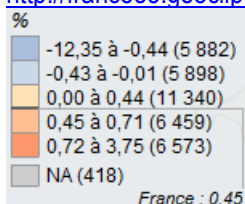
L'intérêt de la méthode des quantiles

Déterminant des classes d'effectifs voisins, c'est la plus robuste statistiquement. Elle ne dépend que de l'ordre des valeurs. Elle est peu sensible à la présence de valeurs aberrantes ou extrêmes. Elle produit des représentations thématiques aux tons équilibrés. Note : les effectifs ne peuvent être rigoureusement égaux si les valeurs sont arrondies (par exemple comme souvent à une ou deux décimales).



Elle permet plus facilement d'intégrer des seuils imposés comme un seuil de divergence (0) et/ou la valeur de référence pour l'indicateur. Dans l'exemple suivant, deux seuils sont pré-déterminés, 0 du fait du changement de signe, et la valeur moyenne en France. La méthode des quantiles est appliquée de part et d'autre de ces seuils.

http://franceo3.geoclip.fr/index.php?profil=FD#i=pops_ev.tvar_nat



- Exercice pratique : dans l'onglet « Distribution » après clic dans la légende sur le bouton « options », agir sur la case à cocher suivante :

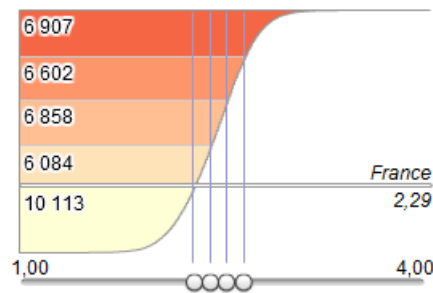
quantiles

La valeur moyenne est un des seuils

L'importance de la valeur de référence

La valeur de référence est différente de la moyenne des valeurs

- Exemple avec le nombre moyen d'occupants par résidence principale : http://franceo3.geoclip.fr/index.php?profil=FD#i=fam_ev_men.nb_moy_res_princ



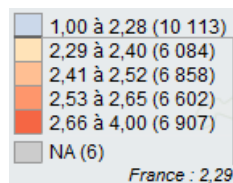
Le nombre moyen d'occupants est plus faible dans les grandes villes, si bien que la valeur moyenne en France est tirée vers le bas.

La valeur de référence fournit une clé de lecture parlante pour une analyse choroplèthe. Elle met immédiatement en évidence les zones au-dessus ou en dessous du comportement moyen.

L'usage d'une palette de couleurs à deux teintes souligne davantage encore ce seuil :

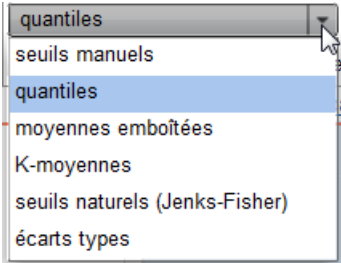
Palettes de couleurs :

- plus contrastées
 une teinte deux teintes

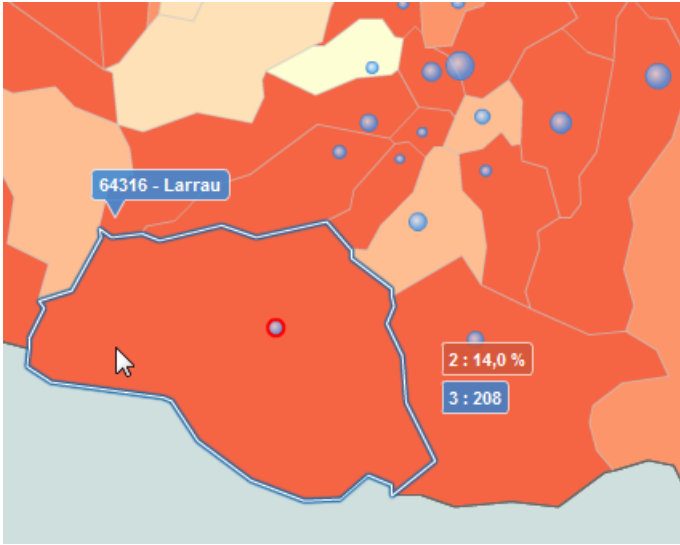



PAUSE !

Etape 5 - Paramétrer une analyse choroplèthe : les autres méthodes de discrétisation

Démarche	Manipulation
Méthode de Jenks-Fischer et K-moyennes	<p>Procédure itérative qui conduit à constituer des classes les plus homogènes possibles. Elle peut repérer des ruptures de pente, des discontinuités dans la distribution. En pratique, elle produit des classes extrêmes d'effectif plus réduit que ceux issus des quantiles.</p> <p>La méthode de Jenks-Fischer, plus coûteuse, conduit à la partition optimale (variance intra-classe minimale et inter-classes maximale).</p> <p>La méthode des K-moyennes est plus rapide, mais dépend d'un scénario d'initialisation et ne garantit pas l'optimum. Elle conduit toutefois souvent à un résultat proche de celui de la méthode de Jenks-Fischer quand le nombre d'objets à traiter est important.</p> <p>Ces méthodes produisent des classes centrales plus étendues, et atténuent donc le contraste global de la carte thématique. Elles sont moins efficaces pour repérer des ruptures quand la population statistique est nombreuse et la distribution par nature plus continue. On les utilisera plus volontiers pour une carte de France par département que par commune.</p>
Moyennes emboîtées	<p>Méthode séduisante, l'algorithme est dichotomique, l'effectif des classes décroît à mesure que l'on s'éloigne de la moyenne, mais moins fortement qu'avec Jenks.</p> <p>Intuitive si la moyenne des valeurs est assez proche de la valeur moyenne. Inconvénient : elle impose un nombre de classes puissance de 2, en pratique 4 ou 8.</p> <p>On peut l'étalonner en l'initialisant avec la moyenne des valeurs ou la valeur moyenne.</p>
Écarts-type	<p>Elle s'appuie sur la moyenne des valeurs et détermine les seuils adjacents en fonction de multiples de l'écart-type. Convient si la distribution est gaussienne.</p> <p>Pb : trouver le bon multiple de l'écart-type... Méthode peu souvent valable et difficile à étalonner.</p>
Amplitudes égales	<p>La plus simple à appliquer, mais trop sensible aux valeurs aberrantes. Elle est donc déconseillée par les statisticiens. Nous ne l'utilisons pas.</p>
Exercices	<ul style="list-style-type: none"> - Exemple avec le nombre moyen d'occupants par résidence principale par commune : http://franceo3.geoclip.fr/index.php?profil=FD#i=fam_ev_men.nb_moy_res_princ - Dans l'onglet Distribution, changer la méthode 

Etape 6 - Comprendre les limites d'une analyse choroplèthe. Réduire le « bruit »

Démarche	Manipulation
<p>Limites visuelles</p>	<p>L'empreinte visuelle d'une couleur dépend aussi de la surface de l'objet qui la porte. Or cette surface n'a pas forcément de lien avec la grandeur représentée.</p> <p>L'intensité d'un ratio ne dit rien du nombre de personnes, par exemple, sur lequel il porte.</p> <p>Ainsi, une commune peu peuplée, de vaste superficie, avec un ratio représenté élevé va transmettre une information forte et un peu abusive.</p>  <p><i>Exemple en zoomant sur une commune de grande étendue, comme Larrau, dans les Pyrénées :</i></p> <p>http://franceo3.geoclip.fr/index.php?profil=FD#i2=pop_age.pxx_pop;i=pop_age5h.pt_pxx_pop75p;s2=2008;v=map12;i=fr;s=2006;z=-152139,5366502,117995,87742</p>
<p>Densité d'écart à la moyenne</p>	<p>Conçue à l'Insee, elle vise à corriger ces effets de bord. L'effet de clarification peut s'avérer spectaculaire.</p> <p><i>Dans l'onglet « Avancé », cocher la case :</i></p> <p><input checked="" type="checkbox"/> Densité d'écart à la moyenne</p> <p><i>Exemples :</i></p> <ul style="list-style-type: none"> - http://franceo3.geoclip.fr/index.php?profil=FV#i=xd_pres2007_t2.pt_sarkozy - http://franceo3.geoclip.fr/index.php?profil=FD#i=fam_men.pt_cxx_menfseul - http://franceo3.geoclip.fr/index.php?profil=FD#i=fam_men.pt_cxx_menfam

	<p>La méthode s'appuie sur la valeur de référence (tauxref) et s'applique à un taux, dont on connaît les composants :</p> <p style="text-align: center;">$var1 / var2$ devient $(\text{tauxref} * var2 - var1) / \text{superficie}$</p> <p style="text-align: center;">ou $var2 * (\text{tauxref} - \text{taux}) / \text{superficie}$</p> <p>Cette densité prend des valeurs positives ou négatives, selon que le taux observé est supérieur ou inférieur à la valeur de référence. Il est donc intéressant de lui appliquer une palette de couleur à deux teintes.</p>
<p>Lissages</p>	<p>Le lissage par disque mobile remplace le ratio observé sur une commune par exemple, par celui calculé sur la commune plus l'ensemble des communes adjacentes. Les différences entre territoires voisins sont donc atténuées, ce qui réduit le « bruit » statistique. De surcroît, l'influence d'une commune peuplée proche sur une commune moins peuplée est prise en compte.</p> <p>Une version étendue de cette méthode consiste à calculer le ratio lissé non pas sur les seules communes adjacentes, mais sur toutes celles distantes de moins de p km, p étant la portée du lissage. Les valeurs voisines sont alors pondérées selon l'inverse de la distance (plus précisément le carré de l'inverse). Ce ne sont pas les ratios qui sont ainsi « moyennés », mais le dénominateur et le numérateur séparément, si bien que la population respective de chaque territoire est bien prise en compte.</p> <p>- Dans l'onglet « Avancé », cocher la case :</p> <p style="margin-left: 20px;"><input checked="" type="checkbox"/> Lissage sur les voisins immédiats </p> <p style="margin-left: 40px;"><input checked="" type="radio"/> simple <input type="radio"/> double (14 km)</p> <p>Après un lissage, la valeur affectée à chaque commune est une donnée d'ambiance, qui peut s'écarter sensiblement de la valeur réelle. Les cartes lissées doivent donc être regardées globalement et non localement.</p>