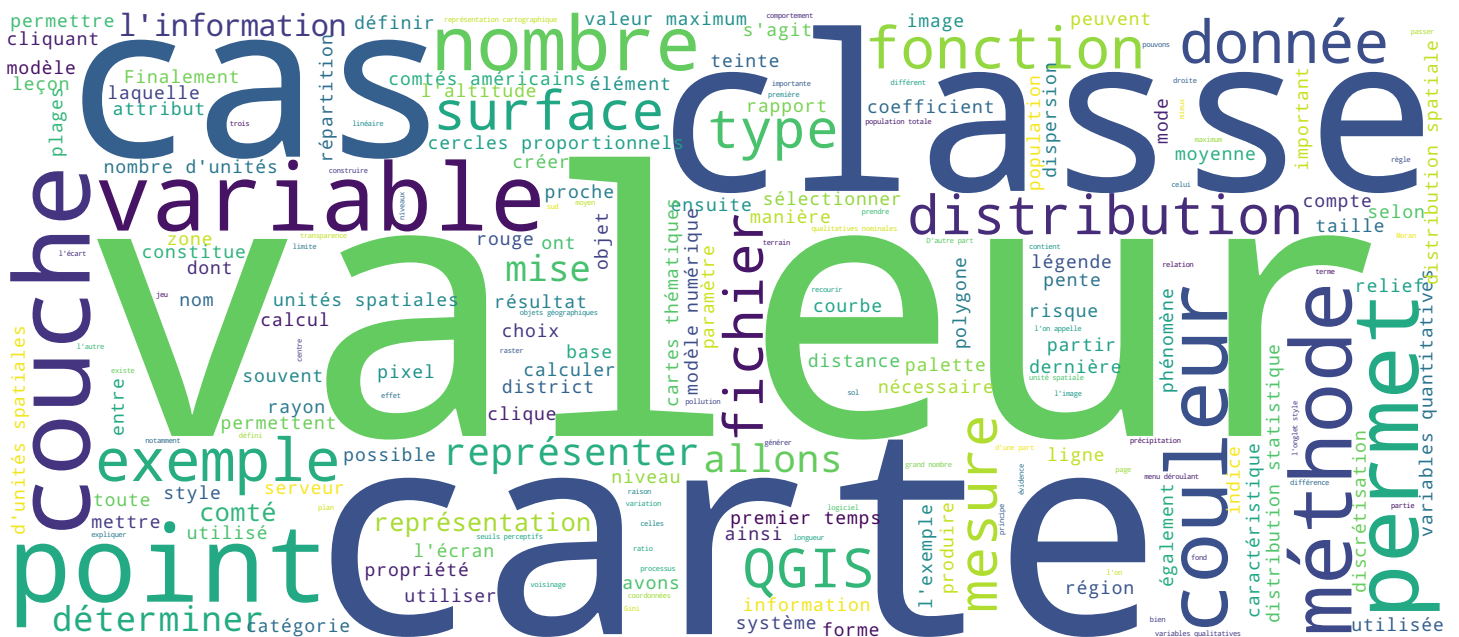


Cartographie thématique et sémiologie graphique

Attributs thématiques et mise en classes

Introduction aux systèmes d'information géographique

Stéphane Joost, Marc Soutter, Fernand Kouamé, Amadou Sall



Search MOOC



Video



Attributs thématiques et mise en classes



Buts de la leçon

- Inculquer les notions de cartographie thématique statistique
- Présenter les types de cartes thématiques
- Expliquer les méthodes de discrétisation

Après cette leçon vous serez capables

- De sélectionner les méthodes de mise en classes et de représentation en fonction des caractéristiques des jeux de données

Introduction aux systèmes d'information géographique

Bonjour et bienvenue dans cette troisième leçon dédiée à la cartographie thématique et à la sémiologie graphique. Nous allons cette fois passer aux aspects pratiques de l'élaboration des cartes thématiques statistiques et examiner comment prendre en compte les différents types d'attributs des objets géographiques à représenter. Les buts de cette leçon sont de vous inculquer les notions nécessaires à la mise en œuvre de la cartographie thématique statistique, de présenter les différents types de cartes thématiques et de vous expliquer le fonctionnement des principales méthodes de mise en classes de l'information statistique. A la fin de cette leçon, vous serez capables de sélectionner les méthodes adéquates de mise en classes et de représentation graphique en fonction des caractéristiques des jeux de données et de construire des cartes thématiques statistiques correctes.

Notes

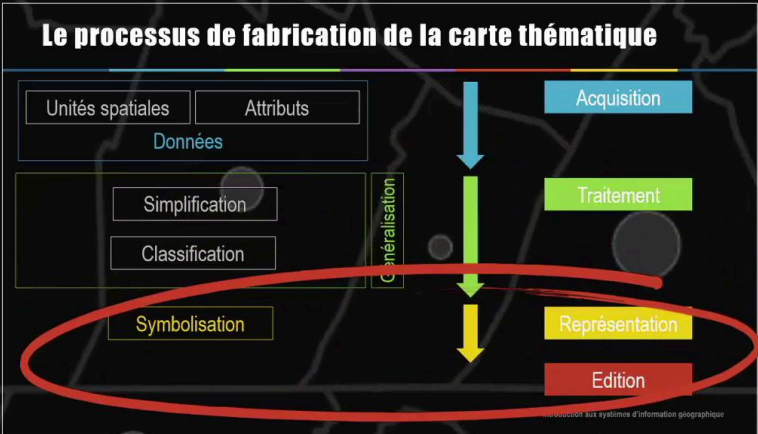
Summary



0m 30s

Notions de base

- Contenu versus contenant
- Traitement



The diagram illustrates the process of creating a thematic map, titled "Le processus de fabrication de la carte thématique". It is organized into three main horizontal sections: "Données" (Data), "Généralisation" (Generalization), and "Représentation" (Representation). The "Données" section includes "Unités spatiales" (Spatial units) and "Attributs" (Attributes). The "Généralisation" section includes "Simplification" and "Classification". The "Représentation" section includes "Symbolisation" (Symbolization). A vertical arrow on the right side of the diagram indicates the flow from "Données" to "Généralisation" to "Représentation". A red oval highlights the "Symbolisation" step, which is the final step in the process. The diagram also includes a small box for "Acquisition" and a box for "Edition" (Editing) at the bottom right. The text "Introduction aux systèmes d'information géographique" is visible at the bottom right of the slide.

Le processus de fabrication de la carte thématique

Unités spatiales Attributs

Données

Simplification

Classification

Généralisation

Symbolisation

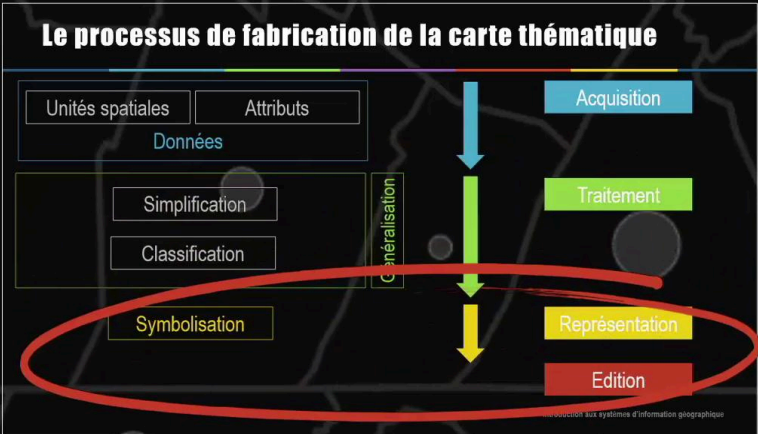
Acquisition

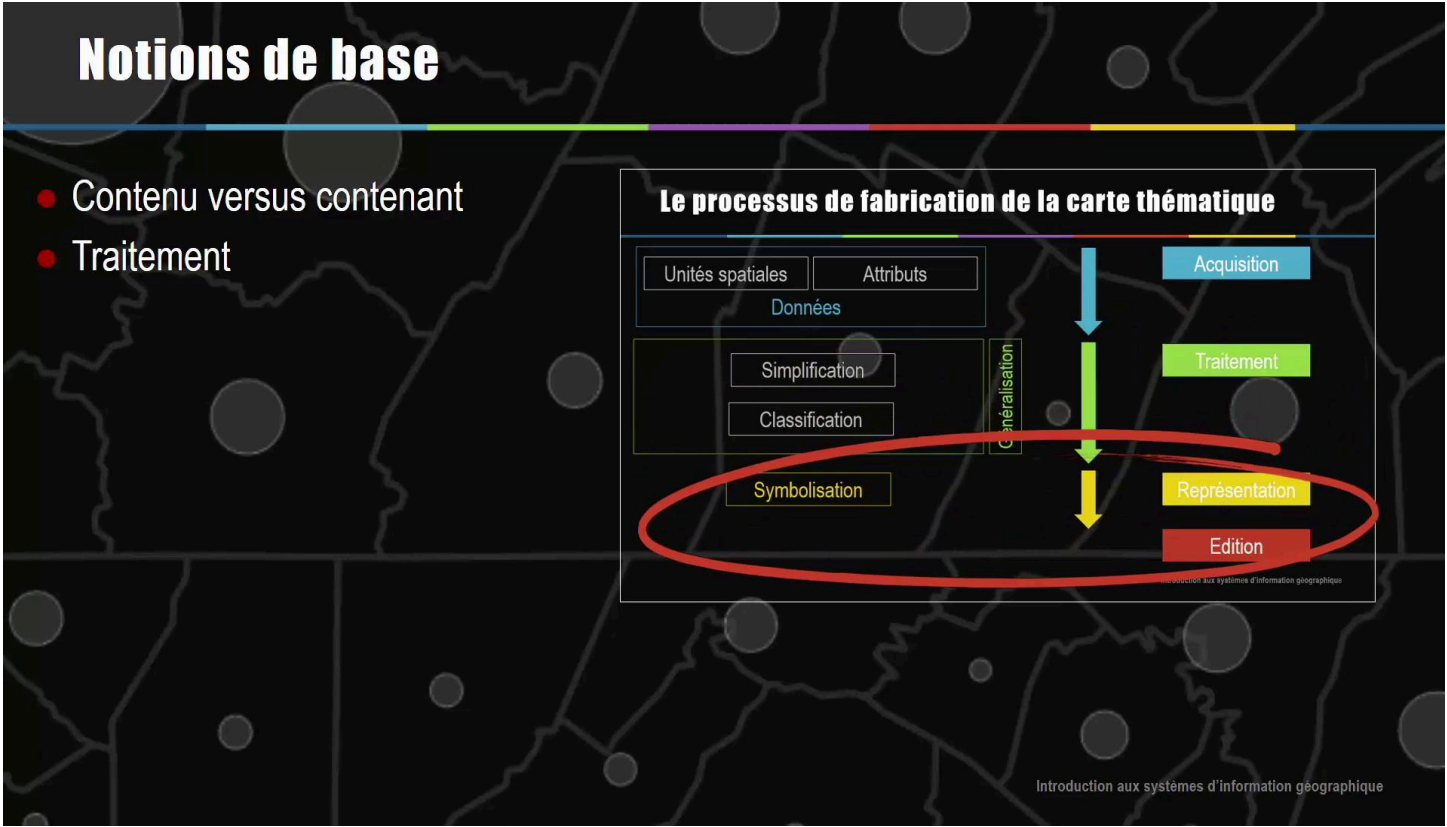
Traitement

Représentation

Edition

Introduction aux systèmes d'information géographique

- # Notions de base
- Contenu versus contenant
 - Traitement
- 
- The diagram illustrates the process of creating a thematic map, titled "Le processus de fabrication de la carte thématique". It is organized into three main horizontal sections: "Données" (Data), "Généralisation" (Generalization), and "Représentation" (Representation). The "Données" section includes "Unités spatiales" (Spatial units) and "Attributs" (Attributes). The "Généralisation" section includes "Simplification" and "Classification". The "Représentation" section includes "Symbolisation" (Symbolization). A vertical arrow on the right side of the diagram indicates the flow from "Données" to "Généralisation" to "Représentation". A red oval highlights the "Symbolisation" step, which is the final step in the process. The diagram also includes a small box for "Acquisition" and a box for "Edition" (Editing) at the bottom right. The text "Introduction aux systèmes d'information géographique" is visible at the bottom right of the slide.
- Le processus de fabrication de la carte thématique
- Unités spatiales Attributs
- Données
- Simplification
- Classification
- Généralisation
- Symbolisation
- Acquisition
- Traitement
- Représentation
- Edition
- Introduction aux systèmes d'information géographique



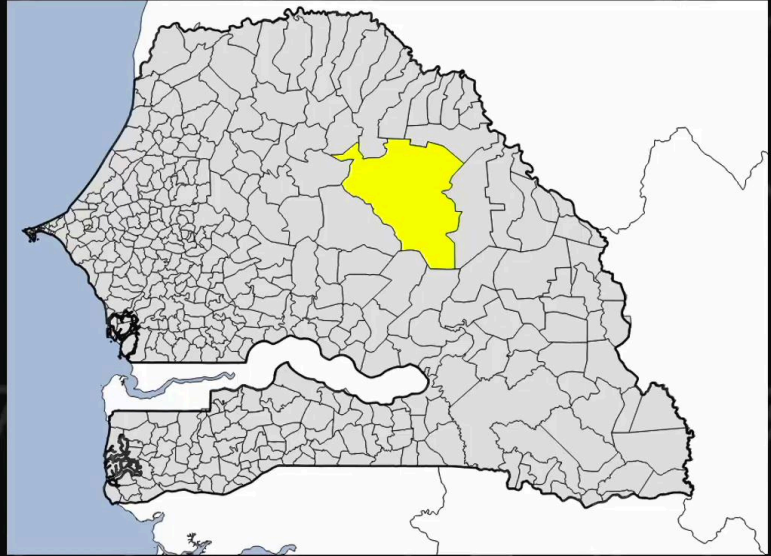
Jusqu'ici, nous avons principalement évoqué les signes utilisés pour représenter l'information géographique, c'est-à-dire le contenant matérialisé par les coordonnées X et Y. Nous allons aborder maintenant les notions nécessaires au traitement du contenu, soit la composante thématique représentée par les caractéristiques ou les attributs des objets géographiques. Vous vous souvenez que la phase de traitement contribue à la généralisation de l'information et qu'elle contient une étape de simplification et de classification, ce qui rend ensuite possible la représentation du modèle de la réalité que constitue la carte. Pour expliquer comment opérer la simplification, la classification puis la représentation de l'information, nous allons utiliser différentes notions.

Summary



Notions de base

- Contenu versus contenant
- Traitement
- Unité spatiale = unité d'observation



Introduction aux systèmes d'information géographique

Il s'agit tout d'abord de celle d'unité spatiale qui constitue une unité d'observation indivisible et géoréférencée, comme le polygone, le point ou la polyligne.

Notes

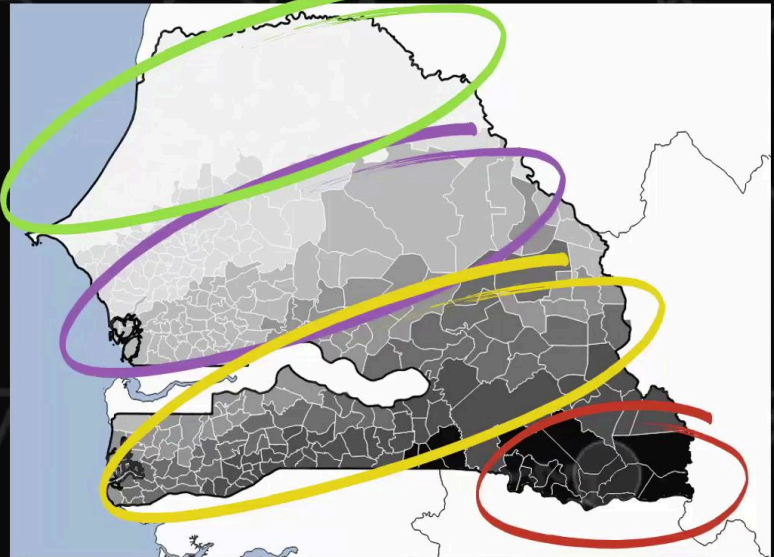
Summary



2m 10s

Notions de base

- Contenu versus contenant
- Traitement
- Unité spatiale = unité d'observation
- Distribution spatiale = répartition
- Structure spatiale = organisation



Introduction aux systèmes d'information géographique

Ensuite, la distribution spatiale concerne la répartition de l'information. C'est la fréquence à laquelle un phénomène apparaît dans l'espace géographique. La structure spatiale, quant à elle, est liée à l'organisation de l'information, à la localisation de chaque unité spatiale par rapport aux autres, par rapport à leur ensemble.

Notes

Summary

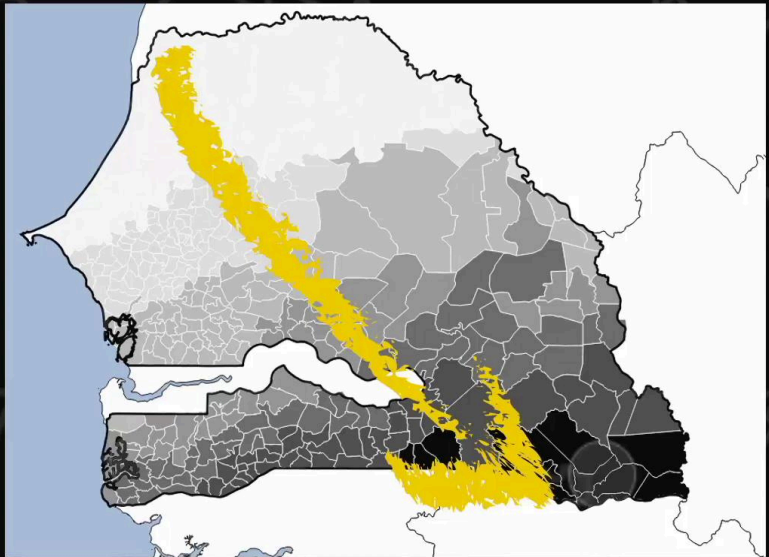


2m 21s

Notions de base

- Contenu versus contenant
- Traitement
- Unité spatiale = unité d'observation
- Distribution spatiale = répartition
- Structure spatiale = organisation
- Processus spatial = mécanisme

- Précipitations



Introduction aux systèmes d'information géographique

Enfin, ce que l'on appelle le processus spatial a trait aux mécanismes qui engendrent les structures spatiales des distributions.

Notes

Summary



2m 39s

Tableau d'information

id	var1	var2	var3	var4	var5	Var6	var7
1	23	bleu	5.5	11.5	1	ouv	pet
2	21	vert	9.7	12.2	2	ferm	grd
3	16	vert	2.3	11.0	4	ouv	moy
4	24	gris	5.9	10.8	2	ferm	grd
5	18	rose	6.2	9.9	1	ouv	moy
6	22	gris	2.3	9.8	3	ferm	grd
7	17	vert	4.4	11.5	3	ferm	grd
8	24	bleu	4.8	10.4	4	ouv	grd
9	23	gris	5.2	11.3	1	ouv	moy
10	20	rose	5.6	10.8	2	ouv	grd
11	19	vert	4.9	11.2	4	ferm	grd
12	24	bleu	1.5	10.4	3	ouv	grd



ation géographique

Pour stocker l'information, on va recourir à un tableau à double entrée. Les lignes horizontales de ce tableau représentent les unités spatiales étudiées. Les colonnes verticales du tableau sont utilisées pour stocker les attributs thématiques. Les unités spatiales sont toujours obligatoirement identifiées de façon univoque. C'est une variable spéciale qui remplit ce rôle, un identifiant numérique, ici en jaune, la colonne ID. Cet identifiant numérique permet de s'assurer que chaque unité spatiale est bien unique et elle permet de faire le lien avec sa géométrie stockée dans un autre fichier. La cartographie thématique statistique va nous permettre d'évaluer l'évolution des variables thématiques dans l'espace. Elles peuvent également évoluer dans le temps et dans ce cas, on devra recourir à plusieurs colonnes représentant plusieurs états temporels. Ces variables thématiques peuvent être une mesure directe, soit un descripteur chiffré précis ou qualitatif expert, ou alors un indicateur dérivé et indirect.

Notes

Summary



2m 48s

Types de données

- Variables **quantitatives**
- Grandeurs numériques
- **Discrètes**: nombre fini de valeurs
- **Continues**: nombre infini de valeurs
- Variables **qualitatives**
- Ses modalités sont des mots

Introduction aux systèmes d'information géographique

Il existe deux types principaux de variables thématiques, soit les données qualitatives et les données quantitatives. Une variable est quantitative si les valeurs qu'elle peut prendre sont des nombres et qu'ils reflètent une notion de grandeur. Ces grandeurs numériques proviennent de comptages, de mesures ou de calculs effectués sur des comptages ou des mesures. Les variables quantitatives peuvent être discrètes ou continues. Elles sont discrètes quand l'information qu'elles représentent prend un nombre fini de valeurs entre deux valeurs quelconques. Les valeurs des variables discrètes sont des nombres précis, isolés et souvent entiers. Les variables quantitatives sont continues quand l'information qu'elles représentent prend un nombre infini de valeurs entre deux valeurs quelconques. Dans ce cas, les valeurs sont des nombres issus d'un intervalle de nombres réels. En général, les variables quantitatives discrètes dénombrent et les variables quantitatives continues approximent. Une variable qualitative ou catégorielle est une variable pour laquelle l'information qui caractérise chaque unité spatiale ne représente pas une quantité. Les différentes valeurs que peut prendre cette variable sont appelées les modalités.

Notes

Summary

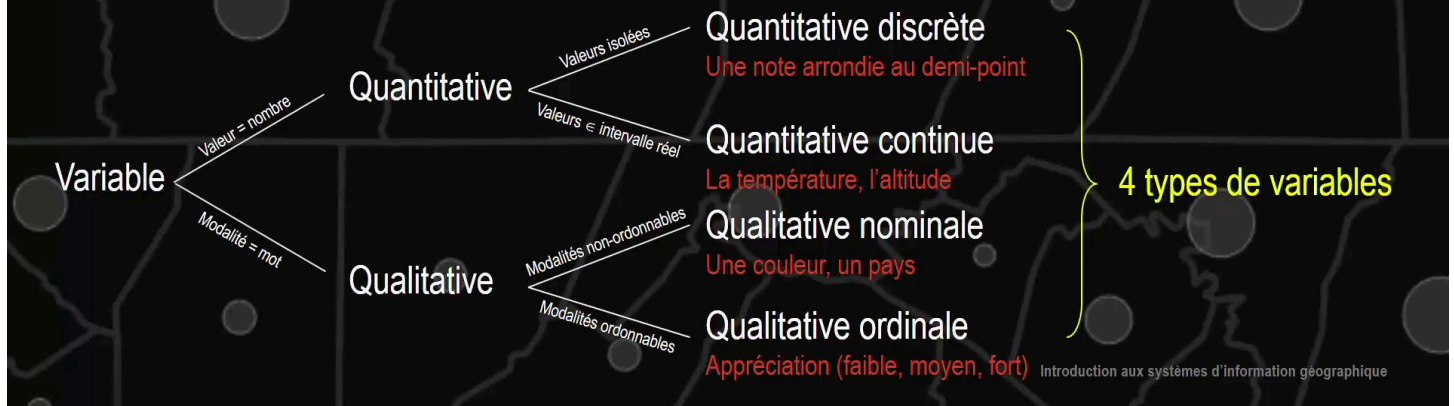


3m 50s

Types de données

- Variables **quantitatives**
- Grandeurs numériques
- **Discrètes**: nombre fini de valeurs
- **Continues**: nombre infini de valeurs

- Variables **qualitatives**
- Ses modalités sont des mots
- **Nominales**: non-ordonnables
- **Ordinales**: ordonnables



On distingue les variables qualitatives nominales dont les modalités ne peuvent pas être ordonnées des variables qualitatives ordinales dont les modalités peuvent être ordonnées selon leur sens. Il est nécessaire de toujours utiliser deux termes pour caractériser précisément une variable parmi les quatre types existant. Ce type peut être quantitatif discret, quantitatif continu, qualitatif nominal ou qualitatif ordinal.

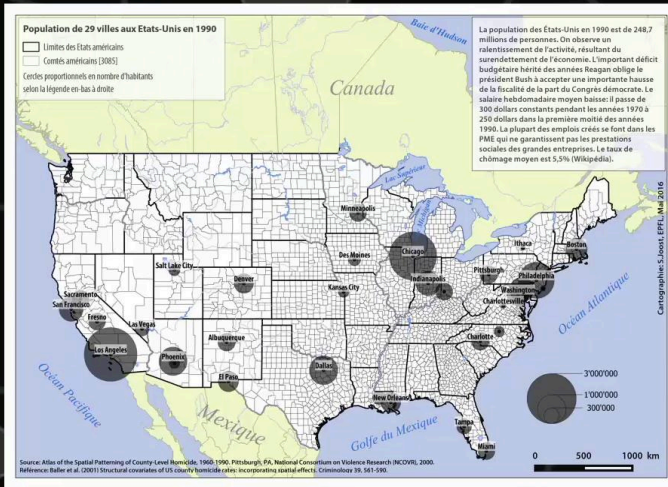
Notes

Summary



5m 02s

Données quantitatives absolues et relatives



- Variable quantitative **absolue**, brute
- Variable visuelle = taille
- Exemple: nombre d'habitants
- Symboles (cercles) proportionnels

Introduction aux systèmes d'information géographique

Une règle importante doit être appliquée aux variables quantitatives : lorsque la variable thématique est brute et qu'elle est issue d'un comptage ou d'un dénombrement, il est obligatoire d'utiliser la variable visuelle de la taille qui représente directement ces variations quantitatives par les variations de la surface d'un symbole. Ainsi, un nombre d'habitants, une quantité de tonnes ou un nombre de véhicules sera toujours représenté par des cercles proportionnels par exemple, centrés sur la localité correspondante ou sur le centroïde d'un polygone. C'est le sujet que nous aborderons dans un premier temps.

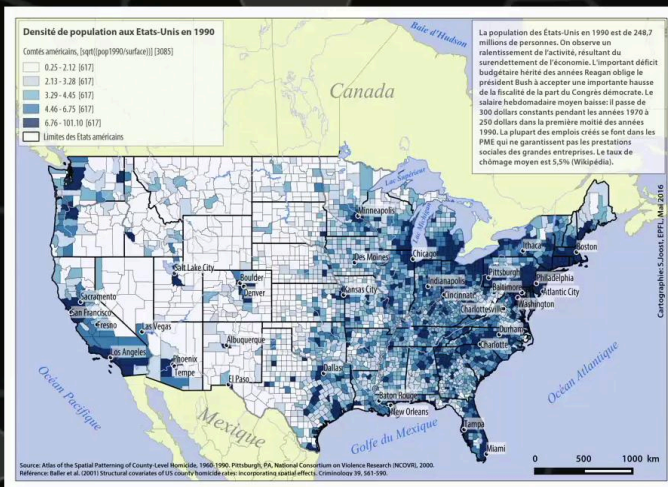
Notes

Summary



5m 30s

Données quantitatives absolues et relatives



- Variable quantitative **absolue**, brute
- Variable visuelle = taille
- Exemple: nombre d'habitants
- Symboles (cercles) proportionnels
- Variable thématique **relative**
- Valable en tout point des surfaces représentées
- Variables visuelles = valeur et/ou couleur
- Exemple: densité de population

Introduction aux systèmes d'information géographique

Par contre, lorsqu'une variable statistique est relative et qu'elle a été transformée par un calcul comme c'est le cas sur la carte affichée à l'écran, elle est valable en tout point de la surface des unités géographiques considérées. On pourra donc utiliser la variable graphique de la valeur et de la couleur qui est adaptée à la représentation des informations ordonnées. Les moyennes, les taux, les indices sont des valeurs relatives qui sont toujours définies par rapport à une autre valeur. C'est le cas de la densité de population ou de la part de personnes sans emploi dans la population active. La représentation cartographique de ce type de variables fait partie de la catégorie des cartes en plages de couleurs que nous aborderons dans un deuxième temps.

Notes

Summary



6m 03s

Cartes en symboles proportionnels

- Relation mathématique entre valeurs et surface des symboles
- Connaître la distribution statistique
- Maximum, minimum, **étendue**

Introduction aux systèmes d'information géographique

La représentation cartographique d'une valeur absolue en symboles proportionnels revient à appliquer une relation mathématique entre la distribution statistique des valeurs enregistrées et la surface des symboles proportionnels à représenter sur le fond de carte. C'est la surface des symboles qui est visualisée et qui constitue le paramètre qui varie proportionnellement. Dans un premier temps, il est nécessaire de connaître la distribution statistique de la variable choisie et notamment les valeurs maximum et minimum qui définissent l'étendue.

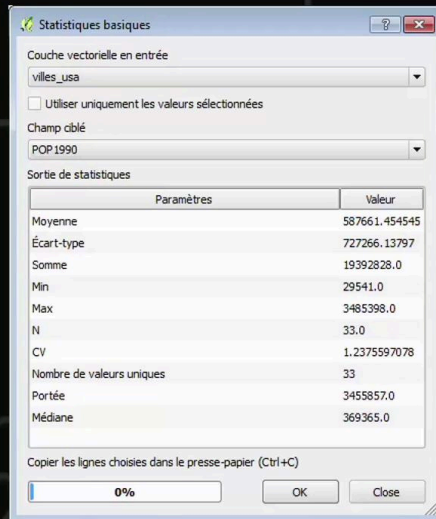
Notes

Summary



6m 53s

Cartes en symboles proportionnels



Paramètres	Valeur
Moyenne	587661.454545
Écart-type	727266.13797
Somme	19392828.0
Min	29541.0
Max	3485398.0
N	33.0
CV	1.2375597078
Nombre de valeurs uniques	33
Portée	3455857.0
Médiane	369365.0

- Relation mathématique entre valeurs et surface des symboles
- Connaître la distribution statistique
- Maximum, minimum, **étendue**
- Dans QGIS, «Statistiques basiques» ou extension «Statist»

Introduction aux systèmes d'information géographique

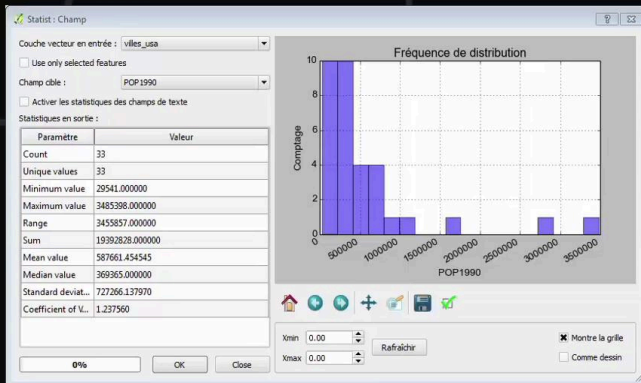
Dans QGIS, l'outil d'analyse vecteur "statistiques basiques" ou le plugin "Statist" permettent de rapidement prendre connaissance des statistiques descriptives de la distribution et même de visualiser la dispersion de la variable dans Statist.

Notes

Summary



Cartes en symboles proportionnels



- Relation mathématique entre valeurs et surface des symboles
- Connaître la distribution statistique
- Maximum, minimum, **étendue**
- Dans QGIS, «Statistiques basiques» ou extension «Statist»
- **Dispersion** faible, symboles de même taille ☹
- Dispersion forte, très petits versus très gros symboles ☹
- Cas extrêmes: transformer le ratio

Introduction aux systèmes d'information géographique

On pourra se servir de ces valeurs dans le processus de construction des cercles proportionnels comme on le verra plus loin. L'étendue est importante car c'est sur cette base que nous pourrions choisir un ratio entre la valeur de la variable et la surface du symbole proportionnel qui permettra de représenter correctement toutes les unités spatiales. Les caractéristiques de la dispersion et de la distribution sont également importantes. Si la dispersion des valeurs est faible, le risque est de construire une carte sur laquelle les symboles ont à peu près tous la même taille sauf les extrêmes. Et dans le cas d'une dispersion très importante, le risque est de produire uniquement de petits et de gros symboles sur la carte. Il faut être attentif à ces cas extrêmes et le cas échéant, il sera nécessaire de transformer le ratio valeur/surface.

Notes

Summary



7m 37s

Symboles proportionnels – contrainte spatiale



géographique

La distribution spatiale et la taille des unités spatiales constituent également une contrainte car ces éléments influencent la répartition des points qui vont supporter les symboles dans le plan. Le problème majeur à résoudre est de dimensionner les plus gros symboles afin qu'ils ne masquent pas le fond de carte et afin que leur emprise n'écrase pas les plus petits symboles. Ensuite, une grande densité d'unités spatiales comme ici avec la population totale dans la région de New York va nuire à la lisibilité de la carte d'autant plus quand, en raison de la proximité avec un centre urbain, les valeurs qu'elle montre sont toutes élevées. D'autre part, les grandes villes au centre des agglomérations possèdent des valeurs de loin supérieures à celles des unités spatiales qui les entourent. Et représenter sur la même carte ces valeurs très différentes et très rapprochées est un réel problème. Dans notre exemple new yorkais, le niveau de transparence peut apporter une aide et permettre une meilleure lisibilité. Mais dans ce genre de cas, la seule solution satisfaisante est la décomposition de la carte en deux vues, l'une qui propose une représentation générale de la région et l'autre qui propose un zoom sur la zone pour laquelle la première vue est quasi illisible et n'apporte pas d'informations.

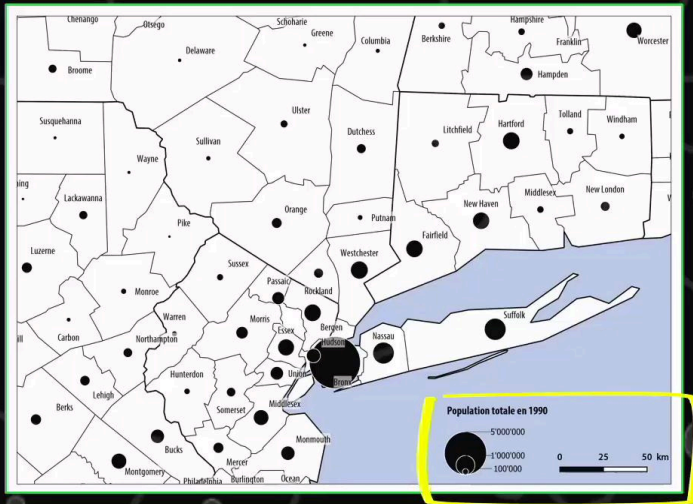
Notes

Summary



8m 25s

Symboles proportionnels – contrainte spatiale



géographique

Il est alors nécessaire d'adapter la règle de proportionnalité pour représenter l'information dans la zone sur laquelle le zoom est effectué.

Notes

Summary



9m 37s

Symboles proportionnels – le ratio valeur-surface

• Variable peu dispersée

id	var1	var2	var3	var4	var5	Var6	var7
1	23	bleu	5.5	11.5	1	ouv	pet
2	21	vert	9.7	12.2	2	ferm	grd
3	16	vert	2.3	11.0	4	ouv	moy
4	24	gris	5.9	10.8	2	ferm	grd
5	18	rose	6.2	9.9	1	ouv	moy
6	22	gris	2.3	9.8	3	ferm	grd
7	17	vert	4.4	11.5	3	ferm	grd
8	24	bleu	4.8	10.4	4	ouv	grd
9	23	gris	5.2	11.3	1	ouv	moy
10	20	rose	5.6	10.8	2	ouv	grd
11	19	vert	4.9	11.2	4	ferm	grd
12	24	bleu	1.5	10.4	3	ouv	grd

• Identifier la valeur maximum = 24

• Un rayon de 8mm = ~200 pixels²

- $S = \pi \cdot r^2$ et $r = \sqrt{\frac{S}{\pi}}$
- 24 → S=200 et r=8;
- 23 → S=190 et r=7.7;
- 21 → S=174 et r=7.4;
- 20 → S=166 et r=7.2, etc.

Introduction aux systèmes d'information géographique

Si la variable à représenter n'est pas trop dispersée comme c'est le cas ici de la variable 1, on va identifier la valeur maximum à représenter pour déterminer la taille du plus gros symbole et déterminer sur cette base le ratio qui permettra de représenter toutes les valeurs en cercles proportionnels. Si l'on veut représenter un rayon de 8 millimètres pour la valeur maximum, on sait que la surface correspondante en pixels carrés est de 200 environ. La surface est en effet égale à π multiplié par le rayon au carré. On utilise donc le ratio de 24 à 200, puis une règle de trois pour calculer la surface et le rayon des autres symboles en proportionnalité simple.

Notes

Summary



Symboles proportionnels – le ratio valeur-surface (2)

• Variable très dispersée

id	var1	var2	var3	var4	var5	var6	var7
1	23	bleu	5.5	11.5	1	ouv	pet
2	21	vert	9.7	12.2	2	ferm	grd
3	16	vert	2.3	11.0	4	ouv	moy
4	24	gris	5.9	10.8	2	ferm	grd
5	18	rose	6.2	9.9	1	ouv	moy
6	22	gris	2.3	9.8	3	ferm	grd
7	17	vert	4.4	11.5	3	ferm	grd
8	24	bleu	4.8	10.4	4	ouv	grd
9	23	gris	5.2	11.3	1	ouv	moy
10	20	rose	5.6	10.8	2	ouv	grd
11	19	vert	4.9	11.2	4	ferm	grd
12	24	bleu	1.5	10.4	3	ouv	grd

- Ratio basé sur valeur la plus forte inadapté
- Transformer le ratio
- Logarithme ou racine
- Perte du lien simple entre les valeurs et les surfaces des symboles
- Mentionner le type de transformation dans la légende

Introduction aux systèmes d'information géographique

Maintenant, si la variable à représenter est trop dispersée, la surface du plus petit symbole découlant du choix du ratio déterminé à partir de la plus forte valeur risque d'être inadaptée. Dans ce cas, il faut transformer le ratio pour le faire passer d'un ratio simple et linéaire du type "valeur x ratio = surface", à un ratio plus complexe, logarithmique ou racine, de type "log valeur x ratio = surface" ou "racine valeur x ratio = surface". On perd alors la liaison simple entre la valeur et les surfaces du symbole avec le risque que le lecteur de la carte ait de la difficulté à comprendre la transformation alors qu'il s'attend à une proportionnalité simple. C'est pourquoi on doit toujours indiquer le type de transformation dans la légende quand on y recourt.

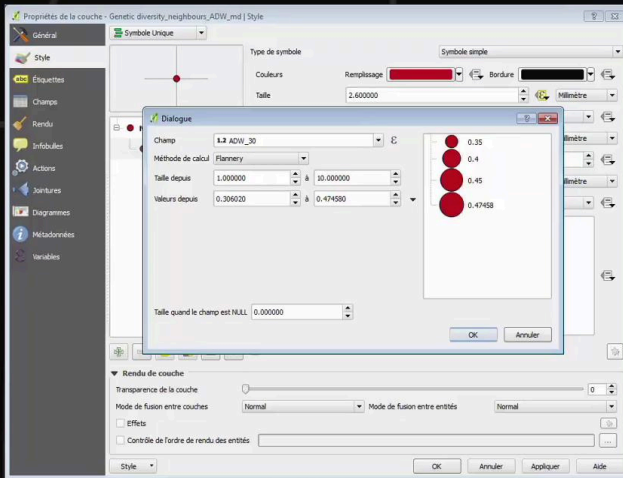
Notes

Summary



10m 28s

Symboles proportionnels dans QGIS



- «Assistant taille» dès la version 2.10

- Extension «Analyse en ronds/secteurs proportionnels»



Introduction aux systèmes d'information géographique

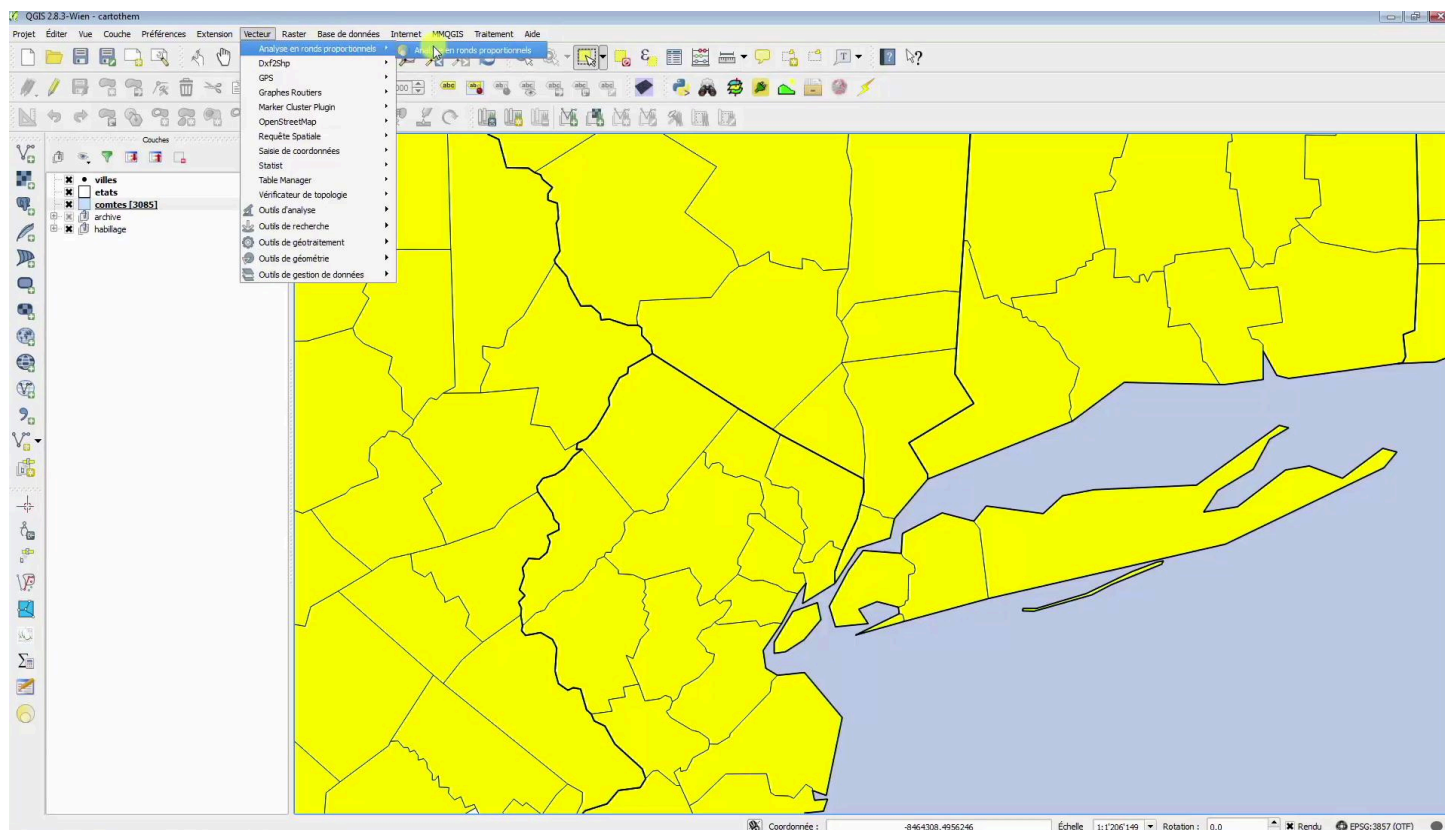
Dans la version 2.8 de QGIS que nous utilisons dans ce MOOC et qui est une version maintenue à long terme, c'est l'extension "analyse en ronds et secteurs proportionnels" qui permet de fabriquer des cartes en cercles proportionnels. Nous allons utiliser cette dernière dans l'exemple qui suit, mais pour information, sachez qu'à partir de la version 2.10 de QGIS, une fonction appelée "assistant taille" est intégrée et accessible via "les propriétés de la couche" dans l'onglet "style", sous "symbole unique", puis en cliquant sur l'icône "expression du paramètre taille". Alors regardons maintenant dans QGIS comment fonctionne l'analyse en cercles proportionnels.

Notes

Summary



11m 15s



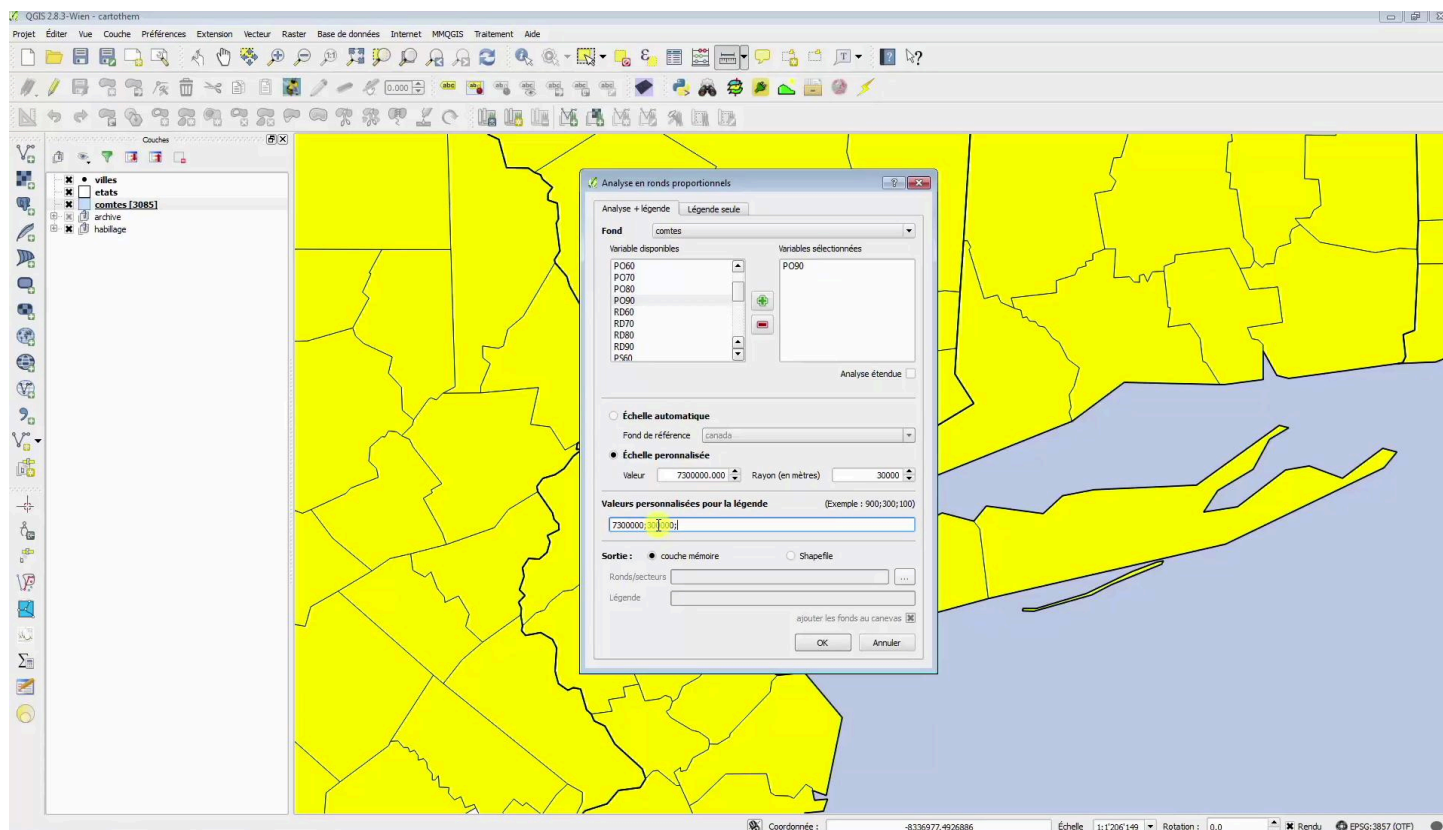
Via le menu "extensions", "installez/gérer les extensions", nous avons préalablement installé le plugin "analyse en ronds et secteurs proportionnels". Sur la carte des comtés américains, nous allons zoomer sur la région de New York pour représenter la population totale par comté, exprimée en nombre d'habitants. Comme nous l'avons vu précédemment, il est important de connaître la distribution des valeurs de variables à représenter, en particulier la valeur maximum et la valeur minimum. Il est possible d'obtenir ces informations en sélectionnant les comtés en question à l'écran, puis en allant dans le menu "vecteur", sous "outils d'analyse", puis "statistiques basiques". On sélectionne la couche des comtés en spécifiant que l'on désire utiliser uniquement les valeurs du champ désiré pour les unités géographiques sélectionnées, c'est-à-dire la population totale en 1990, puis on clique sur OK. La valeur maximale est de 7'300'000 et la valeur minimale d'environ 28'000. On peut maintenant lancer l'extension en cliquant sur l'icône représentant des cercles proportionnels ou via le menu "vecteur", puis "analyse en ronds proportionnels".

Notes

Summary



11m 55s



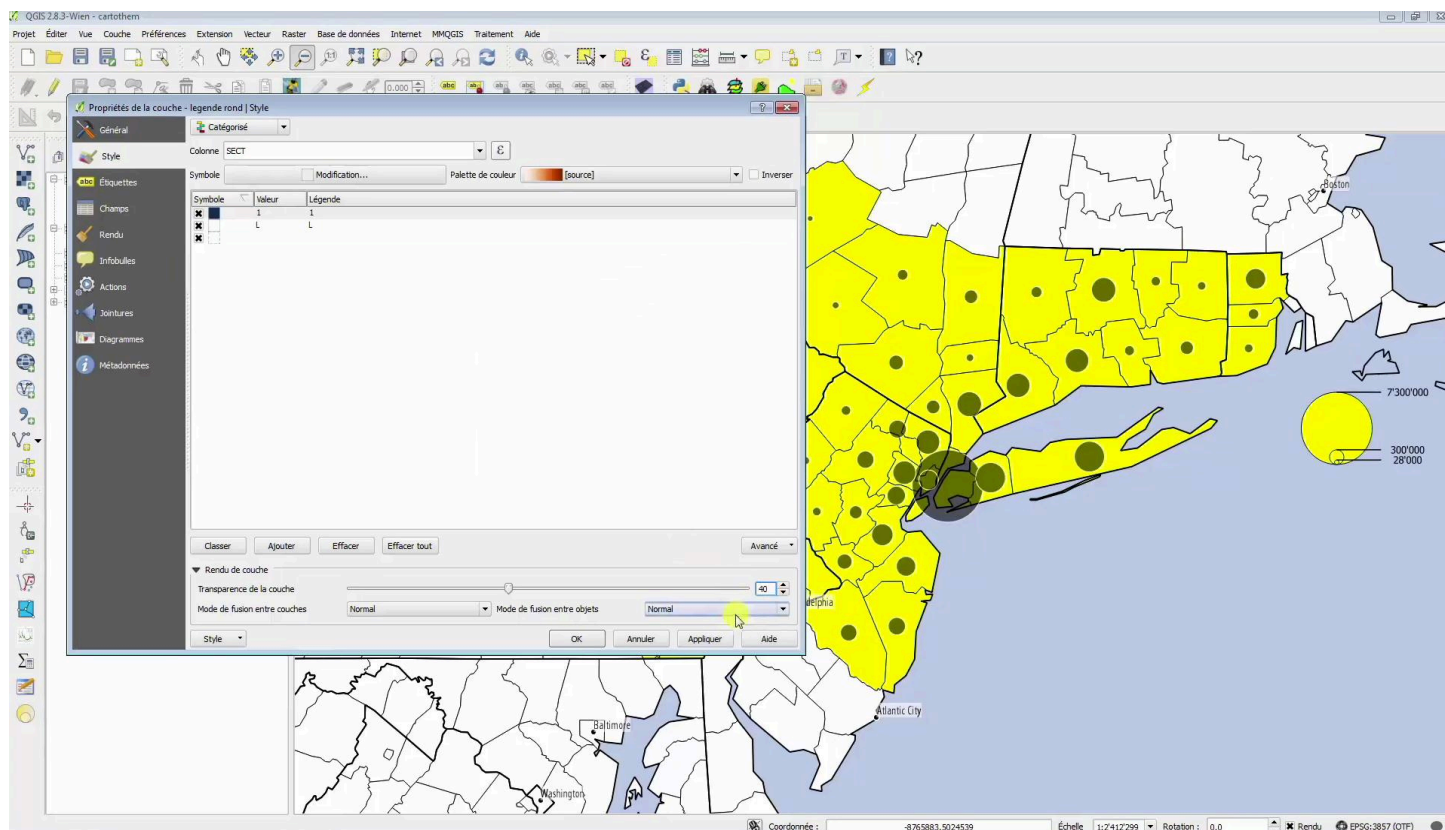
Dans l'onglet "analyse + légende" qui nous intéresse, il faut dans un premier temps sélectionner la couche des comtés, puis la variable de la population en 1990 et l'ajouter dans la liste des variables sélectionnées. Ensuite, on clique sur "échelle personnalisée", option qui demande de fournir d'une part la valeur maximum à représenter et qui est de 7'300'000, et d'autre part le rayon correspondant à cette valeur maximum et qu'il faut représenter sur la carte. Cette valeur doit être fournie en mètre. On peut utiliser l'outil "mesurer une longueur" qui va aider à déterminer ce rayon. Dans notre cas, 30'000 mètres semble une valeur adaptée et on peut la reporter dans le champ correspondant. Ensuite, dans le champ "valeurs personnalisées" pour la légende, il est possible de définir une liste de valeurs que l'on souhaite voir apparaître dans la légende. Les valeurs doivent être séparées par un point virgule ou un espace. Si la case est laissée vide, trois valeurs seront calculées par défaut : le maximum, le maximum divisé par trois et le maximum divisé par neuf. Nous allons demander d'afficher le maximum, soit 7'300'000, la médiane qui est de 300'000 et le minimum.

Notes

Summary

13m 14s





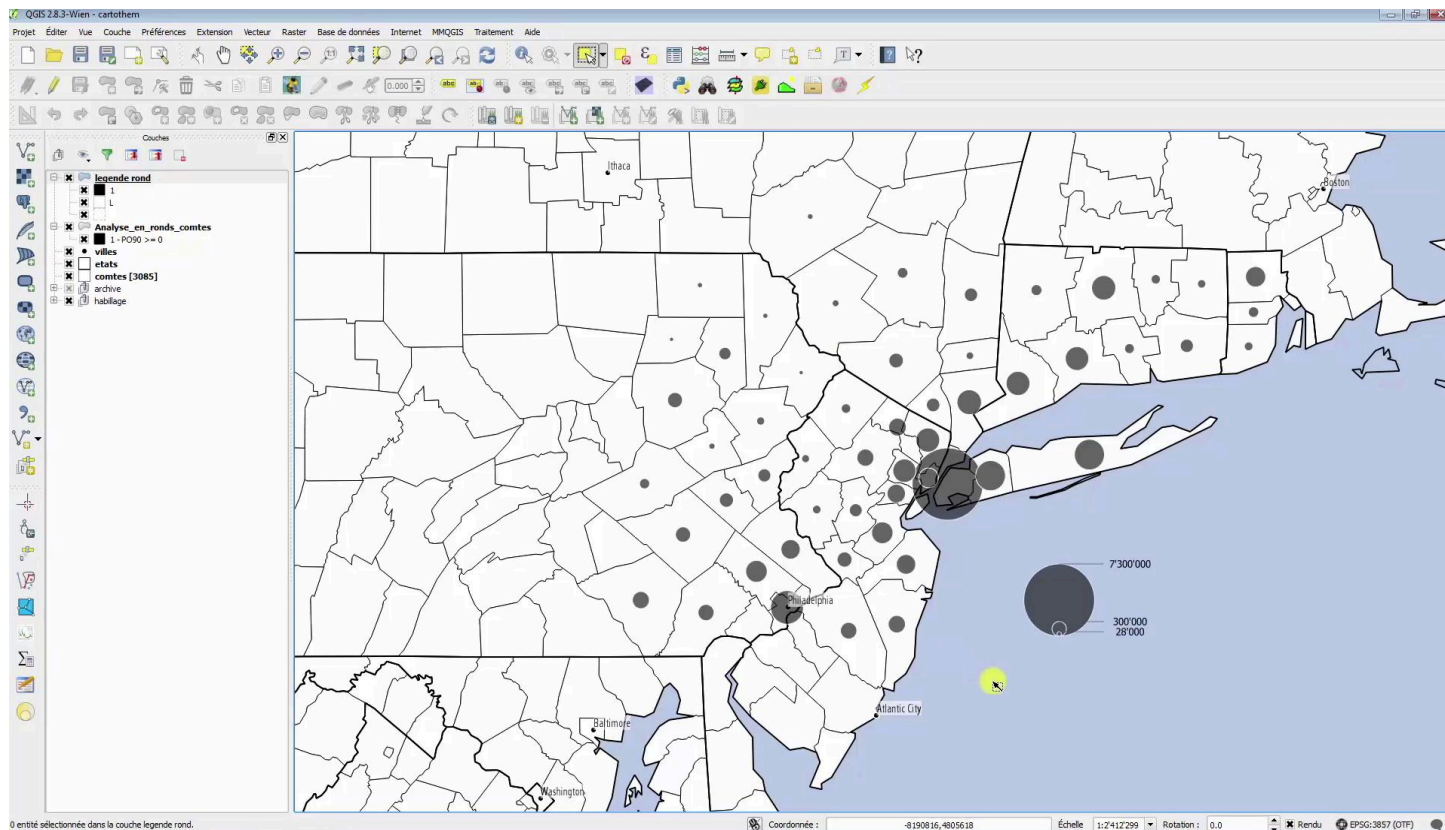
Enfin, nous pouvons choisir si la couche des cercles proportionnels et la légende correspondante doivent être stockées uniquement dans la mémoire de QGIS ou si nous voulons créer deux fichiers au format shape pour pouvoir les réutiliser ultérieurement dans une autre session. Dans ce dernier cas, il faudra alors indiquer le répertoire dans lequel enregistrer les fichiers shape. Ici, nous choisissons l'option "couche mémoire" qui nécessite l'extension "memory layer saver" pour que les couches d'information puissent être enregistrées. On clique finalement sur OK pour générer les cercles proportionnels et la légende. On va maintenant déplacer la légende et les cercles proportionnels au-dessus des autres couches, puis adapter leurs couleurs. Une suggestion est d'utiliser le noir pour le remplissage, le blanc pour le contour, puis d'utiliser la transparence à 40% dans le but de repérer les limites des couches sous-jacentes, et une largeur de bordure de 0,3. Les mêmes paramètres sont ensuite appliqués pour la légende, remplissage en noir, contour en blanc, bordure de 0,3 et transparence de 40%. Finalement, il est possible de déplacer la légende.

Notes

Summary

14m 36s





Il faut sélectionner la couche correspondante, ici "légende ronds", ensuite sélectionner la légende sur la carte, puis la rendre éditable et utiliser l'outil "déplacer l'entité" pour la mettre à l'endroit désiré. Il faut pour terminer, enregistrer la modification. Voilà ! Vous savez maintenant comment produire une carte en cercles proportionnels. Vous apprendrez dans la leçon suivante comment récupérer des éléments de légende pour la mise en page d'un composeur d'impression.

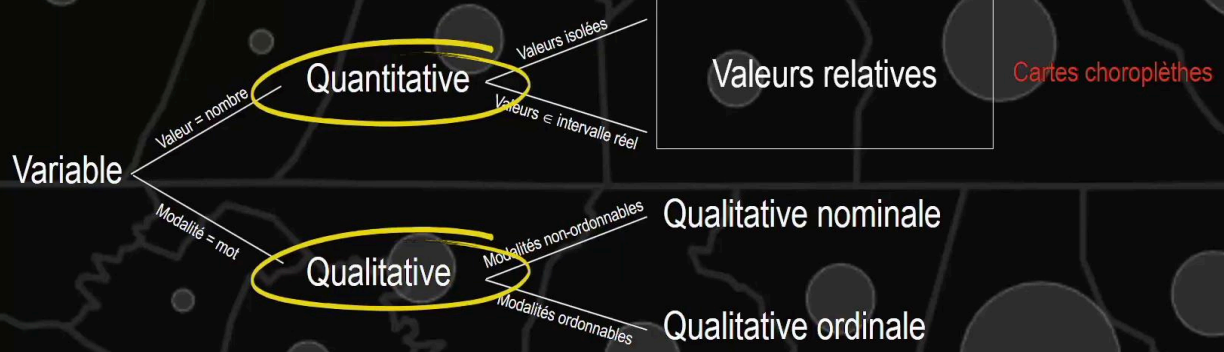
Notes

Summary

15m 57s



Cartes thématiques en plages de couleurs



- Fabrication d'une carte en plages de couleurs
- Etape numéro 1: **la discrétisation**

Introduction aux systèmes d'information géographique

Nous allons passer maintenant à l'élaboration des cartes thématiques en plages de couleurs. Cette catégorie est divisée en deux groupes, d'une part celui qui concerne la représentation des données qualitatives nominales et ordinales et celui qui traite des données quantitatives relatives. Dans ce dernier cas, on parle de cartes choroplèthes, terme que nous expliquerons un peu plus tard. Pour fabriquer une carte en plages de couleurs avec un jeu de données composé d'unités spatiales et de leurs attributs thématiques, il est nécessaire de procéder, dans un premier temps, à ce que l'on appelle la discrétisation.

Notes

Summary



Discrétisation ou mise en classes

- Découper en classes
- Simplifier en regroupant
- Classes homogènes et distinctes

Introduction aux systèmes d'information géographique

La discrétisation est l'opération qui permet de découper en classes la distribution statistique des valeurs d'une variable qualitative ou de variable quantitative, dans le but de simplifier l'information en regroupant les objets géographiques présentant les mêmes caractéristiques en classes distinctes. La discrétisation doit permettre de créer des classes homogènes et distinctes entre elles. Les objets géographiques d'une même classe doivent se ressembler plus entre eux qu'ils ne ressemblent aux objets des autres classes.

Notes

Summary



17m 09s

Discrétisation ou mise en classes



- Découper en classes
- Simplifier en regroupant
- Classes homogènes et distinctes
- Finalisation de la réduction, de l'organisation et de la hiérarchisation de l'information
- Satisfaire les règles de la sémiologie graphique et de la statistique
- Perdre le moins d'information possible
- Applicable aux variables quantitatives, qualitatives nominales et ordinales

Introduction aux systèmes d'information géographique

La discrétisation constitue la dernière étape de la réduction, de l'organisation et de la hiérarchisation de l'information et elle permet de construire une carte qui rend compte de la répartition spatiale d'une distribution statistique. L'opération de mise en classes doit satisfaire aux exigences de la représentation cartographique et à celles des principes statistiques. Elle doit conserver les caractéristiques essentielles présentées par les données et perdre le moins d'information possible, mais aussi respecter les règles de la sémiologie graphique qui permettent d'optimiser la perception visuelle et de transmettre efficacement une information géographique de qualité. Elle est applicable aux variables quantitatives et qualitatives.

Notes

Summary



17m 40s

Discrétisation de variables qualitatives



- Nominales: généralisation de l'information en fonction de la hiérarchie

Introduction aux systèmes d'information géographique

Dans le cas des variables qualitatives nominales, toutes les modalités peuvent être attribuées à des classes différentes dans le but de généraliser l'information si une structure hiérarchique existe.

Notes

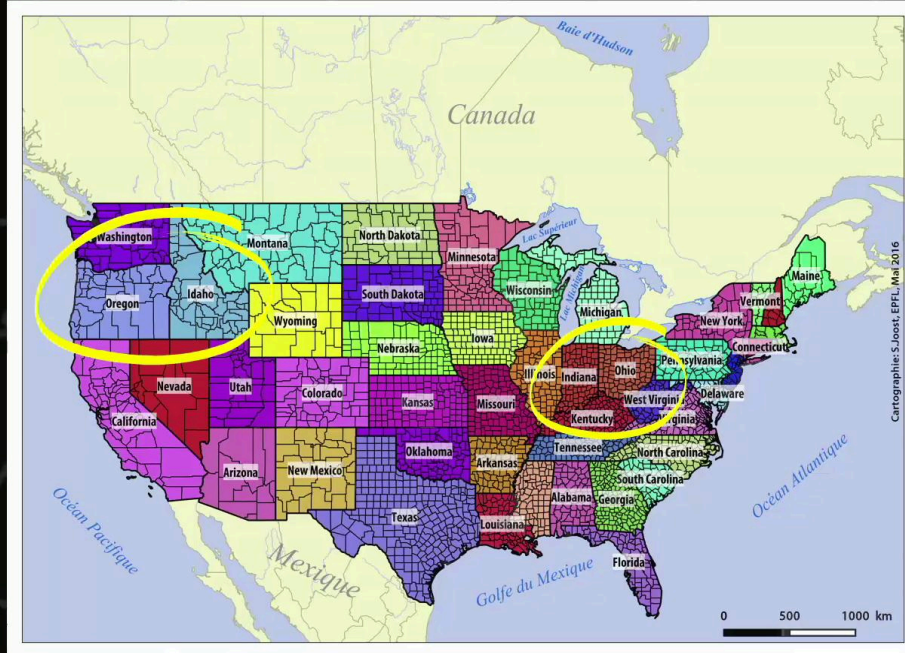
Summary



18m 20s

Discrétisation de variables qualitatives

- Nominales: généralisation de l'information en fonction de la hiérarchie
- Maximiser la discrimination entre unités spatiales



Introduction aux systèmes d'information géographique

Dans l'exemple à l'écran, on a classé les comtés américains selon leur appartenance aux états, mais les comtés auraient pu aussi être répartis entre quatre classes que sont les états du nord-ouest, du nord-est, du sud-ouest et du sud-est. Les cartes qualitatives nominales ne requièrent pas de traitement particulier, si ce n'est l'attribution de couleurs discrètes permettant de différencier au mieux les unités spatiales affichées. La qualité principale de cette carte est de permettre la distinction entre les états et il serait faux dans ce cas de recourir à une représentation faisant appel à la variation de la valeur d'une seule teinte. Ici, c'est une fonction d'attribution aléatoire de couleurs proposée par QGIS qui a été utilisée compte tenu du grand nombre d'unités. Idéalement, il aurait fallu reprendre cette représentation en changeant les couleurs des comtés dans les régions où les différences sont ténues, comme entre l'Indiana, l'Ohio et le Kentucky, ou entre l'Oregon et l'Idaho.

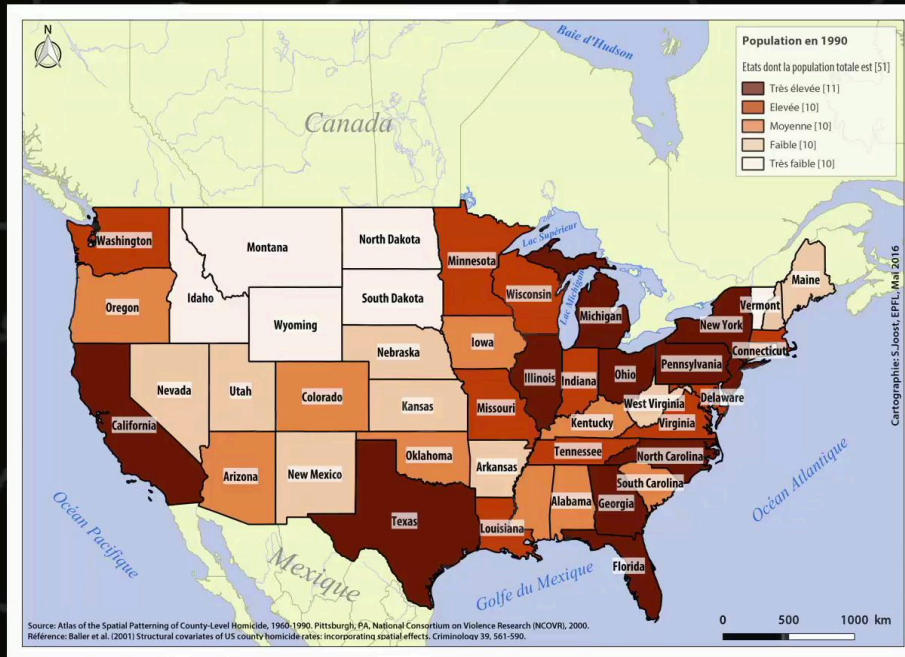
Notes

Summary



18m 31s

Discrétisation de variables qualitatives



- Nominale: généralisation de l'information en fonction de la hiérarchie
- Maximiser la discrimination entre unités spatiales
- Ordinales: utilisation du rang



Introduction aux systèmes d'information géographique

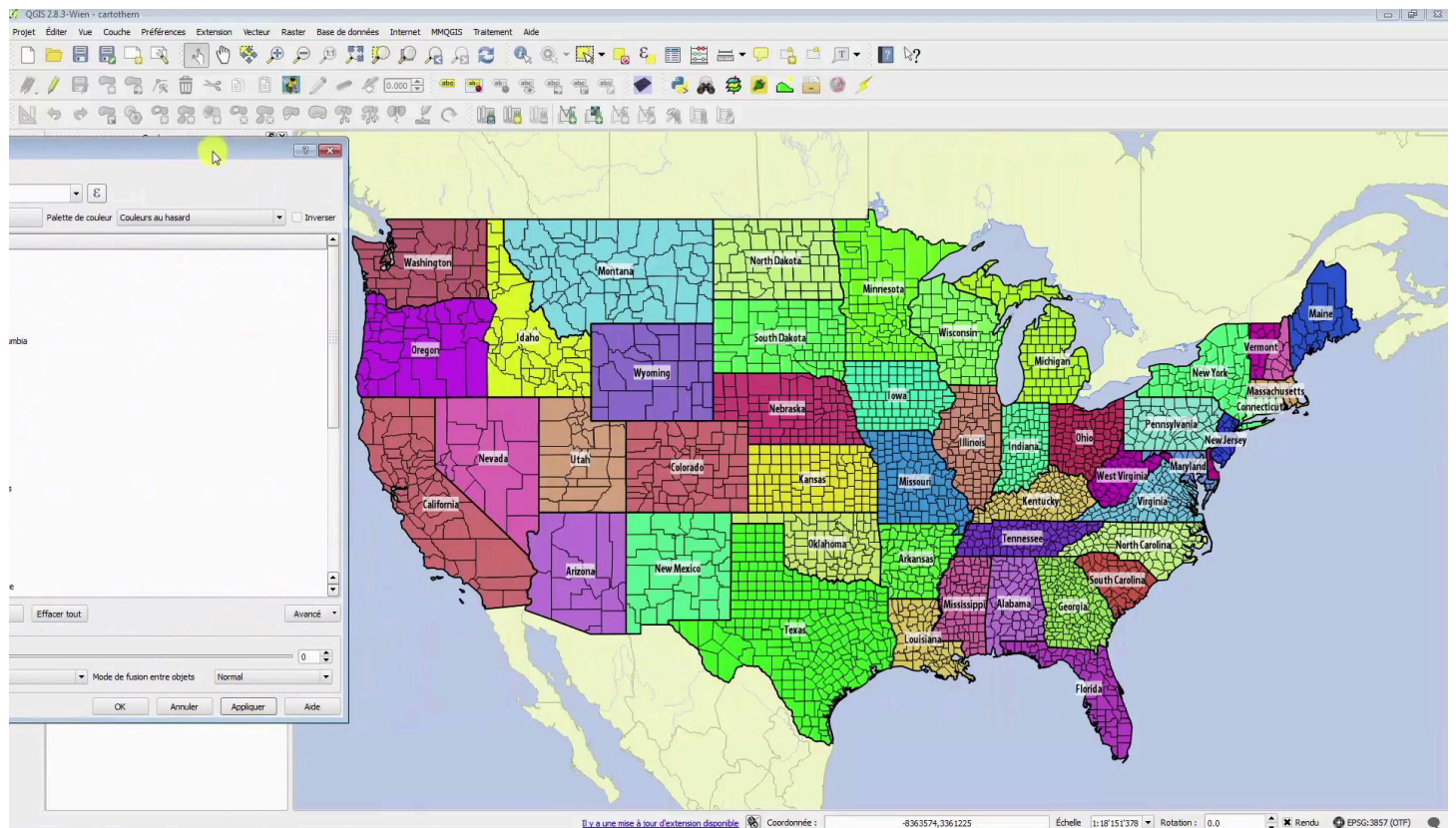
Dans le cas des variables qualitatives ordinales, comme ici des indications sur la taille de la population totale, la différence entre deux modalités successives n'est pas quantifiable et elle correspond toujours à l'unité. Mais il est possible d'attribuer un rang aux différentes valeurs et de discrétiser la distribution en question sur cette base. Nous allons maintenant passer dans QGIS pour vous expliquer comment produire une carte qualitative nominale.

Notes

Summary

19m 31s





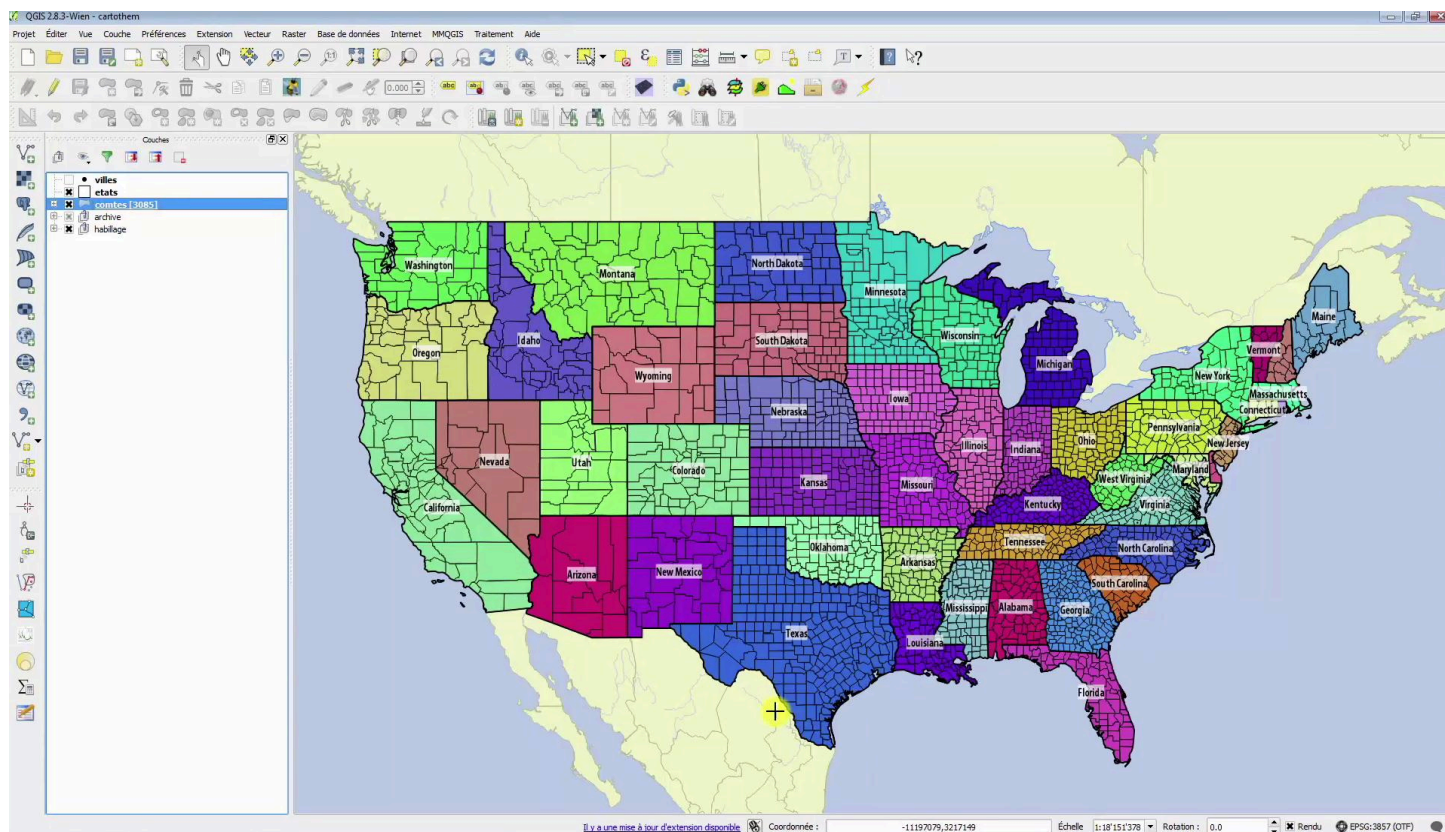
Dans QGIS, on a préalablement ouvert un projet contenant notamment la couche des comtés américains que l'on sélectionne ici. Un double-clic sur la couche permet d'atteindre les propriétés de la couche et l'onglet "style". Dans le menu déroulant, situé en haut de la fenêtre, choisir "catégoriser", puis sélectionner la colonne "state name" et enfin cliquer sur le bouton "classer" sans modifier la palette de couleurs courante. QGIS va alors lister tous les états contenus dans la couche "comté" et attribuer à chacun d'entre eux une teinte en fonction de la palette présélectionnée, ici des dégradés de bleu. Notre souci est d'attribuer des couleurs distinctes à tous les comtés de manière à pouvoir différencier au mieux les états qu'ils constituent. D'autre part, il est de pouvoir compter sur une fonction de remplissage automatique puisque le nombre d'unités considéré est important et qu'il serait fastidieux d'entreprendre cette tâche manuellement. On va donc choisir une palette de couleurs appropriée, soit "couleurs au hasard". En cliquant sur le bouton "appliquer", on peut se faire une idée de la manière dont les différentes couleurs ont été réparties sur les comtés.

Notes

Summary



19m 57s



Si la répartition des couleurs n'est pas bonne, donc s'il est difficile de différencier certains états en raison de couleurs trop proches, il faut relancer le processus en choisissant tout d'abord n'importe quelle autre palette de couleurs en guise de réinitialisation, puis d'appliquer à nouveau la palette "couleurs au hasard" et ainsi de suite jusqu'à ce qu'une configuration correcte apparaisse, ceci avant de valider définitivement le choix de cette palette. On a la possibilité de l'enregistrer ensuite pour une utilisation ultérieure en cliquant sur le menu déroulant "style", puis "enregistrer le style" et enfin, "fichier de style de couche QGIS". Il faut déterminer un répertoire où sauver ce fichier.QML et lui donner un nom, par exemple "comtés_nominal.qml". Ce fichier sera utilisable via le menu déroulant "style", puis "charger le style". Finalement, pour valider le choix des couleurs, on clique sur OK.

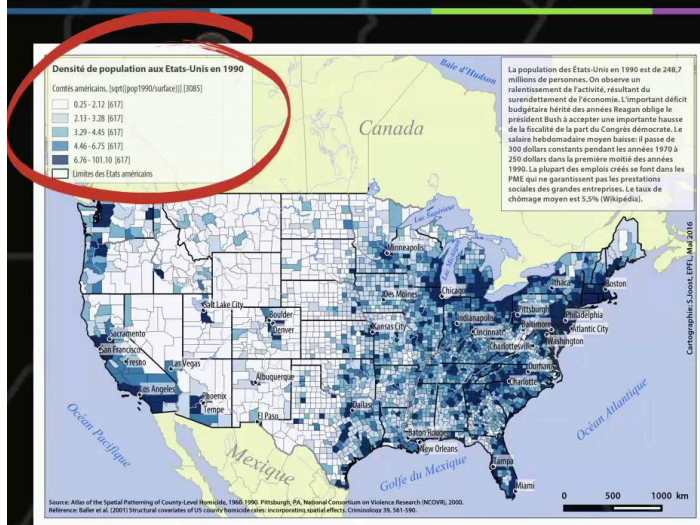
Notes

Summary

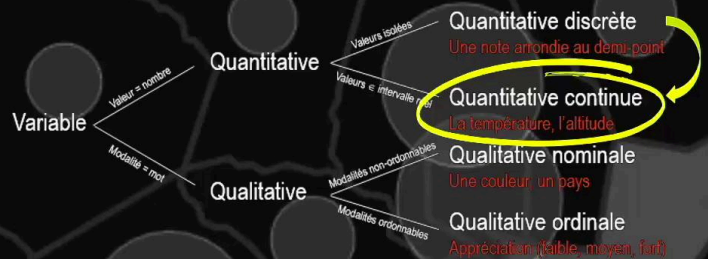
21m 14s



Cartes choroplèthes



- **Plethos** = quantité
- **Khorê** = ~aire géographique
- Comptages, mesures
- Variables transformées
- Distribution quantitatives continues



Introduction aux systèmes d'information géographique

Autre catégorie de représentation cartographique en plages de couleurs, la carte choroplèthe est le type le plus fréquent de cartes statistiques adaptées à la représentation de quantités relatives caractérisant des aires géographiques au moyen d'une échelle de valeurs graduée de teintes. Par quantité relative, il faut comprendre qu'il s'agit de variables comme des comptages ou des mesures qui ont été transformés par un calcul pour donner des rapports ou des pourcentages, distributions qui deviennent alors continues. Par exemple, pour la carte affichée ici à l'écran, la densité de population par comté, qui est une variable quantitative discrète, a été transformée. La valeur originale est devenue la racine carrée du rapport entre le nombre d'habitants par comté en 1990 et la surface du comté correspondant, soit une variable quantitative continue. La mise en classe et le choix des teintes et des valeurs sont deux opérations cruciales pour élaborer ce type de cartes.

Notes

Summary

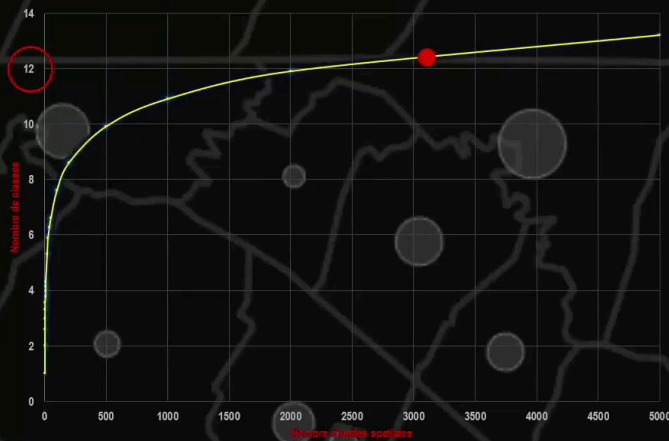


22m 21s

Discrétisation de variables quantitatives relatives

$$N_{cl} = 1 + 3,3 \log_{10}(N_{obs})$$

● Indice de Huntsberger



Introduction aux systèmes d'information géographique

Le nombre de classes optimal à déterminer est souvent présenté comme étant une fonction du nombre d'unités spatiales dans un jeu de données. Dans cet ordre d'idées, il existe des indices permettant de calculer le nombre de classes idéal pour une distribution donnée. Un exemple est l'indice de Huntsberger pour lequel le nombre de classes idéal, N_{cl} est égal à 1 plus 3.3 fois le logarithme base 10 du nombre d'unités spatiales.

Notes

Summary

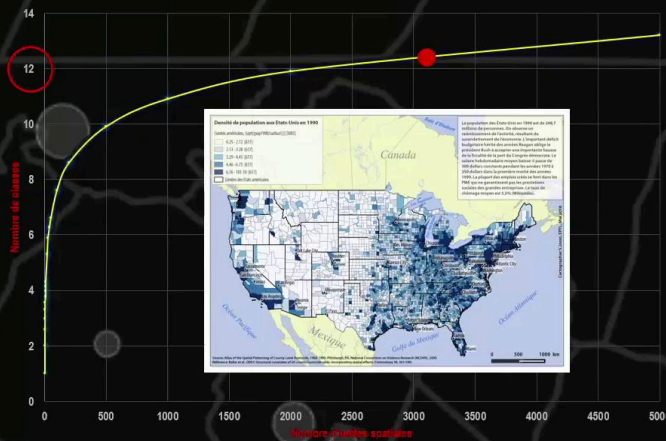


23m 22s

Discrétisation de variables quantitatives relatives

$$N_{cl} = 1 + 3,3 \log_{10}(N_{obs})$$

● Indice de Huntsberger



Introduction aux systèmes d'information géographique

Si l'on considère un nombre d'unités spatiales comparable à l'exemple des 3'080 comtés américains, cette formule recommande une discrétisation en 12 classes, ce qui pose un problème de perception car notre système visuel n'est pas capable de distinguer autant de seuils perceptifs, soit 11 dans ce cas, dans la distribution de la valeur d'une seule teinte. On se souvient que dans l'idéal, il ne faudrait pas dépasser six seuils perceptifs, soit sept classes, de manière à garantir l'efficacité maximale de la carte. Dans ce cas, on s'aperçoit sur le graphe de droite qu'on ne pourrait théoriquement pas créer de cartes choroplèthes représentant plus de 33 unités spatiales. Il faut donc considérer ce type d'indice comme une indication lorsque l'on travaille avec un petit nombre de polygones, mais il est clair que le nombre de classes idéal se situe autour de 5 dans la plupart des cas et à partir de là, c'est le thème représenté, les caractéristiques de la variable et sa distribution spatiale qui vont guider les choix de l'auteur de la carte. Par exemple, faut-il utiliser un nombre pair ou impair de classes ? Le nombre impair sera choisi lorsque l'on veut faire référence à une classe dont le comportement est moyen.

Notes

Summary

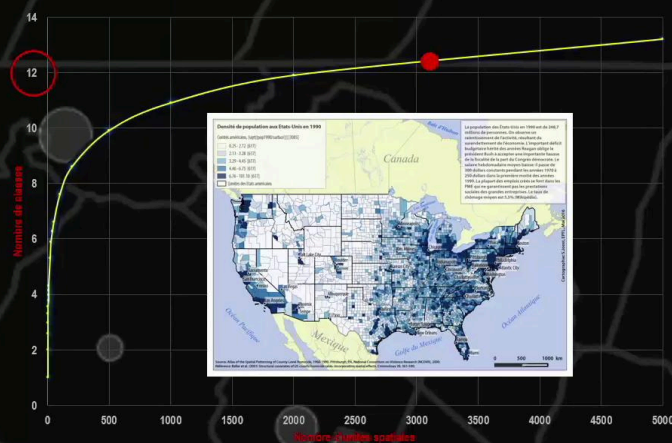


23m 49s

Discrétisation de variables quantitatives relatives

$$N_{cl} = 1 + 3,3 \log_{10}(N_{obs})$$

● Indice de Huntsberger



Introduction aux systèmes d'information géographique

Pour les jeux de données comprenant un grand nombre d'unités spatiales, il est possible d'adopter une approche pragmatique qui permet de prévoir un plus grand nombre de classes et à la fois plus de seuils perceptifs.

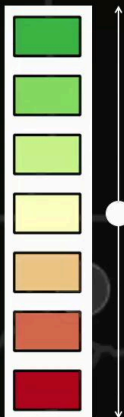
Notes

Summary



24m 56s

Nombre de classes – grand nombre d'unités spatiales



- Augmenter le nombre de seuils perceptifs
- Deux teintes
- Une classe intermédiaire
- Nombre impaire de classes

Introduction aux systèmes d'information géographique

L'idée est d'augmenter le nombre de seuils perceptifs exploitables en utilisant un gradient de valeurs pour deux teintes. Ces dernières sont séparées par une classe intermédiaire, représentée par une teinte neutre comme le jaune pâle par exemple. On utilisera une des teintes pour représenter les valeurs les plus basses et l'autre teinte pour les valeurs les plus élevées, les deux s'articulant autour de la teinte neutre.

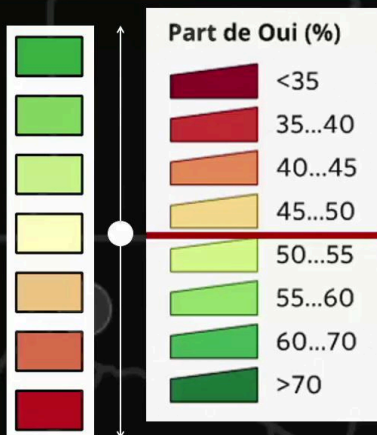
Notes

Summary



25m 09s

Nombre de classes – grand nombre d'unités spatiales



- Augmenter le nombre de seuils perceptifs
- Deux teintes
- Une classe intermédiaire
- Nombre impair de classes

Introduction aux systèmes d'information géographique

On va donc de préférence utiliser un nombre impair de classes, sauf dans les cas pour lesquels la variable traitée représente deux comportements opposés, comme l'acceptation ou le refus d'un objet de votation. Ici, un nombre pair de classes convient tout à fait.

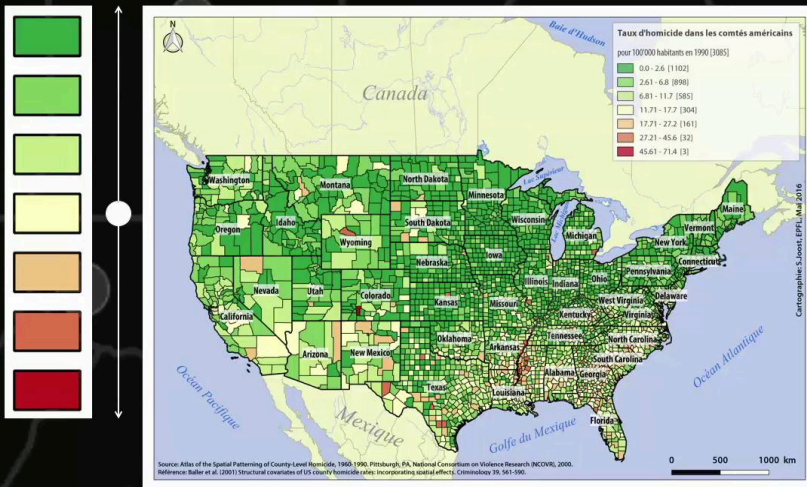
Notes

Summary



25m 38s

Nombre de classes – grand nombre d'unités spatiales



- Augmenter le nombre de seuils perceptifs
- Deux teintes
- Une classe intermédiaire
- Nombre impair de classes
- Maximum théorique: 12-1 classes
- Choix des couleurs cohérent avec le thème traité
- Explications nécessaires
- Carte fausse

Introduction aux systèmes d'information géographique

On peut alors augmenter le nombre de classes jusqu'à un maximum théorique de 12 si l'on se base sur les indications de la formule de Huntsberger. On utilisera donc 11 classes pour obtenir un nombre de classes impair avec une classe intermédiaire et deux gradations de valeur en cinq classes, voire même neuf ou sept classes, ceci dans le but de produire la carte la plus efficace possible. Ce procédé est délicat dans la mesure où il est nécessaire de bien expliquer le système de la double gradation au lecteur et où le choix des couleurs doit être cohérent avec le thème présenté. Sur cette carte des comtés américains en sept classes et deux teintes, le vert foncé représente un faible taux, soit une situation favorable. La valeur du vert diminue vers la classe intermédiaire, en jaune pâle, puis la valeur du rouge augmente vers les taux les plus élevés d'homicides dans les comtés où la situation est négative. On exploite ici le sens associé aux couleurs verte et rouge par analogie avec le code de la route et les feux de circulation. Les explications sont d'autant plus importantes que ce subterfuge induit la création d'une carte théoriquement fausse puisque la distribution statistique d'une variable comme le taux d'homicides devrait être représentée par un simple gradient de valeurs d'une seule teinte.

Notes

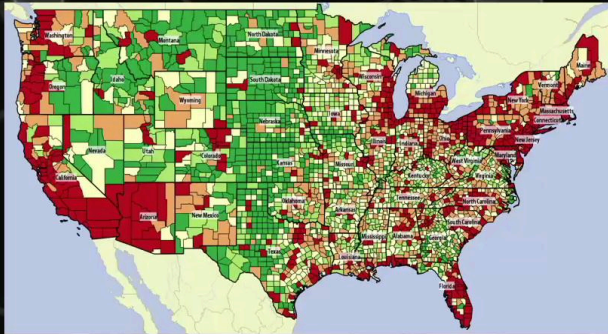
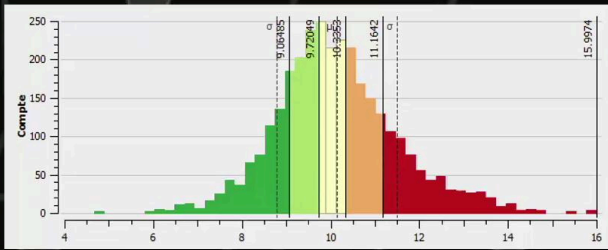
Summary

25m 51s



Choix d'une méthode de discrétisation

- Répartir les données dans les classes



Introduction aux systèmes d'information géographique

Une fois que nous avons une idée du nombre de classes à utiliser, il faut y répartir les données de la distribution statistique à représenter. Les principes de la discrétisation diffèrent selon la nature de l'information.

Notes

Summary



27m 11s

Choix d'une méthode de discrétisation



- Répartir les données dans les classes
- Variable qualitative nominale: pas de méthode spécifique, définir un critère commun de regroupement
- Variable qualitative ordinale: conserver la hiérarchie de la distribution
- Variables quantitatives, 3 types d'information à considérer:
 - Ordre de grandeur, valeurs centrales
 - Dispersion: écart-type, intervalle inter-quantiles

Introduction aux systèmes d'information géographique

Si l'information est qualitative nominale, comme nous l'avons vu, il n'existe pas de méthode spécifique, on va chercher à définir un critère commun de regroupement pour aboutir à la construction d'une typologie. Chaque regroupement d'objets géographiques est propre aux objectifs de simplification des données choisis. C'est le cas de la carte des comtés américains regroupés en fonction de leur appartenance aux états que nous avons élaboré dans QGIS il y a un instant. Si la variable est qualitative ordinale, on va chercher à conserver la hiérarchie des informations et constituer les classes sur la base des rangs, comme nous l'avons également évoqué plus tôt. Dans le cas de l'information quantitative, trois types d'information sur la distribution peuvent être pris en compte. Ces informations permettront de déterminer quelle méthode de discrétisation utiliser. Le premier type est l'ordre de grandeur qui est révélé par les valeurs centrales de la distribution. Il s'agit du mode, de la moyenne ou de la médiane qui peuvent être choisis comme limites de classe. Ensuite, la mesure de la dispersion est une mesure de l'inégalité des valeurs, elle est caractérisée par l'écart-type ou par un intervalle inter-quantiles, valeurs qui peuvent être choisies comme amplitude des classes.

Notes

Summary



27m 24s

Choix d'une méthode de discrétisation



- Répartir les données dans les classes
- Variable qualitative nominale: pas de méthode spécifique, définir un critère commun de regroupement
- Variable qualitative ordinale: conserver la hiérarchie de la distribution
- Variables quantitatives, 3 types d'information à considérer:
 - Ordre de grandeur, valeurs centrales
 - Dispersion: écart-type, intervalle inter-quantiles
 - Forme de la distribution: normale, asymétrique, exponentielle, unimodale, bimodale, plurimodale, etc.

Introduction aux systèmes d'information géographique

Une mesure de dispersion prend en compte la variance et elle permet de minimiser les différences entre objets d'une même classe et de maximiser les différences entre les classes. Finalement, la forme de la distribution renseigne sur le comportement d'un phénomène et elle doit être prise en compte pour déterminer la méthode de mise en classe la plus adaptée.

Notes

Summary



28m 36s

Méthodes de discrétisation (1)

Méthode	Définition	Calcul	Remarques	Type de distribution statistique
Ecart à la moyenne	Toutes les classes ont une même étendue égale à l'écart type, sauf les classes extrêmes	Moyenne et écart type	Si le nombre de classes est impair, la classe centrale est à cheval sur la moyenne. Intérêt: se repérer par rapport à la moyenne, mettre en évidence les extrêmes, comparer plusieurs cartes	Normale (Gauss) avec concentration des données autour de la moyenne, éventuellement peu asymétrique
Égale amplitude	Les intervalles de classes sont égaux (intervalles constants)	(Valeur maximale - valeur minimale) / nombre de classes	Méthode simple, facile à interpréter mais peu utilisée car ne convient pas si la distribution des valeurs est asymétrique : les classes peuvent être très inégales et certaines vides. Pas de comparaison possible	Distribution uniforme (données réparties uniformément sur toute l'amplitude de la série), normale (Gauss) avec concentration des données autour de la moyenne

Source: d'après «Rappels sur les discrétisations», F. Demoraes, M. Souris, T. Serrano, Laboratoire de Cartographie Appliquée, Paris X. Nanterre, & E. Ployon, Université de Savoie, Formation SIG-Santé 2010

Introduction aux systèmes d'information géographique

Nous présentons ici cinq méthodes avec des caractéristiques différentes. La méthode de l'écart à la moyenne permet de générer des classes d'étendue identique, sauf les classes extrêmes qui sont différentes. L'étendue d'une classe est égale à l'écart-type. Dans le cas où le nombre de classes est impair, la classe centrale sera à cheval sur la moyenne. L'intérêt principal de cette méthode est qu'elle rend possible la comparaison entre plusieurs cartes, mais il faut que la distribution de la variable analysée soit normale ou alors faiblement asymétrique. La méthode des intervalles égaux ou des classes d'égale amplitude a pour but de constituer des classes dont les intervalles sont constants. On calcule ces intervalles en soustrayant la valeur minimale de la distribution à la valeur maximale et en divisant le résultat par le nombre de classes préalablement déterminé. Cette méthode est simple mais peu utilisée car le risque est de tomber sur des classes vides ou d'effectifs très inégaux. En outre, elle ne permet pas les comparaisons entre cartes. Cette méthode requiert une distribution normale et uniforme, ce qui est très exigeant.

Notes

Summary



Méthodes de discrétisation (2)

Méthode	Définition	Calcul	Remarques	Type de distribution statistique
Seuils naturels	Seuils observés	Observation de l'histogramme de fréquences cumulées triées dans l'ordre croissant	Prend en compte les discontinuités de la série Cartes non comparables	Toute distribution qui présente des discontinuités Distribution plurimodale
Quantiles	Chaque classe a le même nombre d'individus	Effectif total / nombre de classes Limites de classes= nb d'individus défini par le calcul ci-dessus, statistique d'ordre	Ne tient pas compte des valeurs exceptionnelles. Des limites de classes peuvent être discutables (valeurs très proches réparties dans des classes différentes) Permet les comparaisons	Distribution uniforme ou autre distribution sans discontinuités À éviter si présence de valeurs extrêmes ou si trop grand nombre de valeurs proches
Box map	Classes basées sur l'analyse en boxplot ou boîte à moustaches			

Source: d'après «Rappels sur les discrétisations», F. Demoraes, M. Souris, T. Serrano, Laboratoire de Cartographie Appliquée, Paris X. Nanterre, & E. Ployon, Université de Savoie, Formation SIG-Santé 2010

Introduction aux systèmes d'information géographique

La méthode des seuils naturels repose sur l'observation de l'histogramme de fréquences cumulées triées dans l'ordre croissant. Si celles-ci laissent apparaître des seuils nets entre groupes d'unités spatiales, ces derniers sont utilisés comme limites de classes. Cette méthode exploite donc les discontinuités d'une distribution, mais ne permet pas ensuite de comparer les cartes entre elles. Il s'agit d'une méthode adaptée aux distributions plurimodales. Dans le cas de la méthode des quantiles, chaque classe compte le même nombre d'individus. Il s'agit d'une statistique d'ordre et l'effectif des classes est défini par le nombre total d'individus divisé par le nombre de classes. Un défaut de cette approche est qu'elle ne peut pas tenir compte des valeurs exceptionnelles ou "outliers" en anglais et certaines limites de classes peuvent être discutables puisque des valeurs très proches peuvent tomber dans des classes distinctes. Par contre, ce type de mise en classes permet de faire des comparaisons entre cartes, à éviter si la distribution présente des valeurs extrêmes. Finalement, la méthode de la boîte à moustaches exploite également une statistique d'ordre.

Notes

Summary



Méthodes de discrétisation (2)

Méthode	Définition	Calcul	Remarques	Type de distribution statistique
Seuils naturels	Seuils observés	Observation de l'histogramme de fréquences cumulées triées dans l'ordre croissant	Prend en compte les discontinuités de la série Cartes non comparables	Toute distribution qui présente des discontinuités Distribution plurimodale
Quantiles	Chaque classe a le même nombre d'individus	Effectif total / nombre de classes Limites de classes= nb d'individus défini par le calcul ci-dessus, statistique d'ordre	Ne tient pas compte des valeurs exceptionnelles. Des limites de classes peuvent être discutables (valeurs très proches réparties dans des classes différentes) Permet les comparaisons	Distribution uniforme ou autre distribution sans discontinuités À éviter si présence de valeurs extrêmes ou si trop grand nombre de valeurs proches
Box map	Classes basées sur l'analyse en boxplot ou boîte à moustaches	6 classes: quartile 1 (Q1), Q2, Q3, Q4, outliers inférieurs ($Q1 - [1.5 \cdot (Q3 - Q1)]$), outliers sups. ($Q1 + [1.5 \cdot (Q3 - Q1)]$)	Statistique d'ordre, permet les comparaisons, implémentée dans Geoda	Toute distribution, adéquat pour traiter des valeurs exceptionnelles

Source: d'après «Rappels sur les discrétisations», F. Demoraes, M. Souris, T. Serrano, Laboratoire de Cartographie Appliquée, Paris X. Nanterre, & E. Ployon, Université de Savoie, Formation SIG-Santé 2010

Introduction aux systèmes d'information géographique

Son principe est de reprendre les classes générées par les box-plot, outil d'analyse exploratoire développé par John Tukey. Elle produit six classes dont quatre sont les quartiles, et deux sont des classes permettant de mettre en évidence les valeurs extrêmes des deux côtés de la distribution. Les cartes produites avec l'aide de cette méthode sont comparables.

Notes

Summary



Règles de base



- Les classes doivent couvrir l'ensemble de la distribution et doivent être contiguës
- Une valeur ne doit appartenir qu'à une seule classe
- Les classes ne peuvent pas être vides
- Le choix des valeurs limites doit reposer sur des critères robustes
- Des valeurs non significativement différentes ne doivent pas être placées dans des classes distinctes

Introduction aux systèmes d'information géographique

Finalement, en appliquant l'une ou l'autre de ces méthodes de mise en classes, il faut veiller autant que se peut à respecter les règles suivantes : les classes doivent couvrir l'ensemble de la distribution statistique et doivent être contiguës. Toute valeur ne peut appartenir qu'à une seule classe. Aucune classe ne peut être vide. Le choix des valeurs limites doit reposer sur des critères clairs. Et les valeurs très proches ne devraient pas être placées dans des classes différentes. Nous pouvons maintenant passer à la fabrication d'une carte choroplèthe.

Notes

Summary



31m 34s

Cartographie du coefficient de Gini

- Mesurer l'inégalité de revenu au sein d'une population
- Ecart à la courbe de Lorenz



information géographique

Le but est d'analyser la distribution spatiale du coefficient de Gini dans les comtés américains en 1989. Ce coefficient est utilisé pour mesurer l'inégalité des revenus au sein des populations des différents comtés. Ce niveau d'inégalité peut ensuite être comparé d'une unité spatiale à l'autre. Le coefficient de Gini est un nombre qui varie entre 0 et 1 où 0 signifie une égalité parfaite et 1 une inégalité totale. Le coefficient de Gini se calcule par rapport à la fonction qui associe à chaque fraction de la population touchant un revenu classé par ordre croissant, la part que représente ces revenus.

Notes

Summary

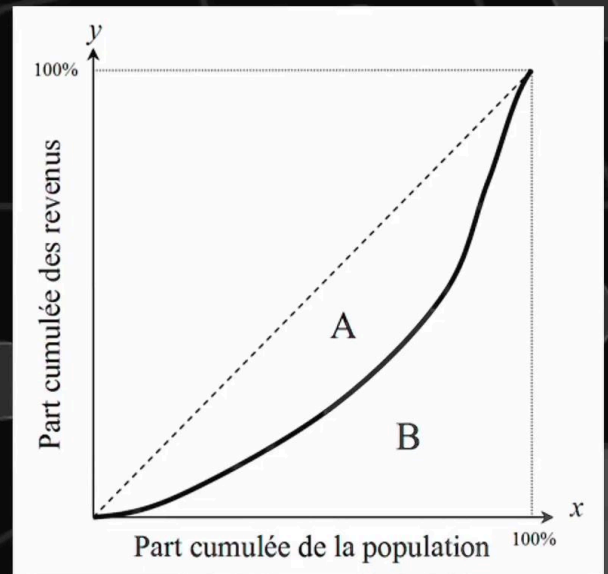
32m 10s



Cartographie du coefficient de Gini

- Mesurer l'inégalité de revenu au sein d'une population
- Ecart à la courbe de Lorenz
- Plus cette courbe est éloignée de la bissectrice, plus les inégalités sont fortes
- Coefficient = $A/(A+B) = A/(1/2) = 2A$
 $= 2(A+B-B) = 2(A+B) - 2B = 1 - 2B$

C. Gini (1921) *Measurement of inequality of income*, *Economic Journal* 31, 22-43.



Introduction aux systèmes d'information géographique

Le coefficient de Gini estime l'inégalité par l'écart à la courbe de Lorenz qui illustre la répartition de la richesse dans une société. Plus cette courbe est éloignée de la bissectrice, plus les inégalités sont fortes. Le coefficient de Gini est la différence d'aire entre le triangle formé par la bissectrice et la zone délimitée par la courbe de revenus. Pour plus de détails, nous vous renvoyons à l'article original de Corrado Gini.

Notes

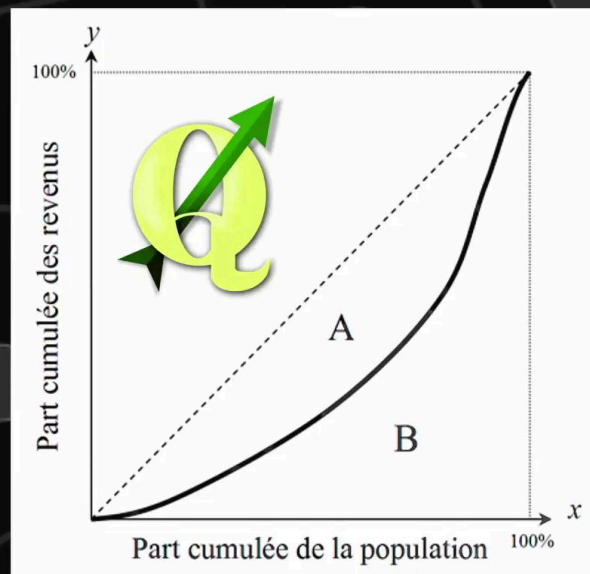
Summary



Cartographie du coefficient de Gini

- Mesurer l'inégalité de revenu au sein d'une population
- Ecart à la courbe de Lorenz
- Plus cette courbe est éloignée de la bissectrice, plus les inégalités sont fortes
- Coefficient = $A/(A+B) = A/(1/2) = 2A$
 $= 2(A+B-B) = 2(A+B) - 2B = 1 - 2B$

C. Gini (1921) *Measurement of inequality of income*, *Economic Journal* 31, 22-43.



Introduction aux systèmes d'information géographique

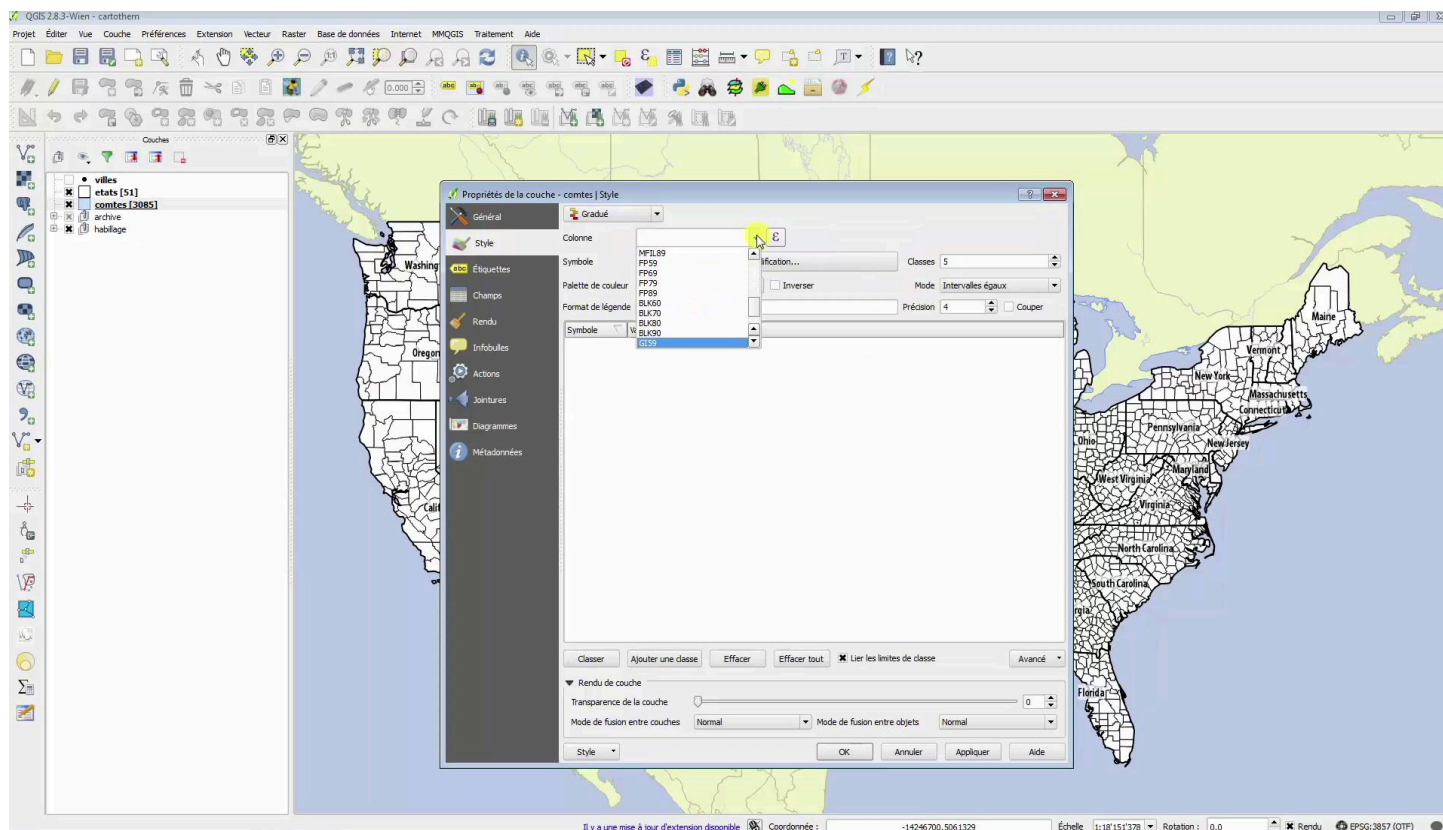
Nous pouvons maintenant passer dans QGIS.

Notes

Summary



33m 16s



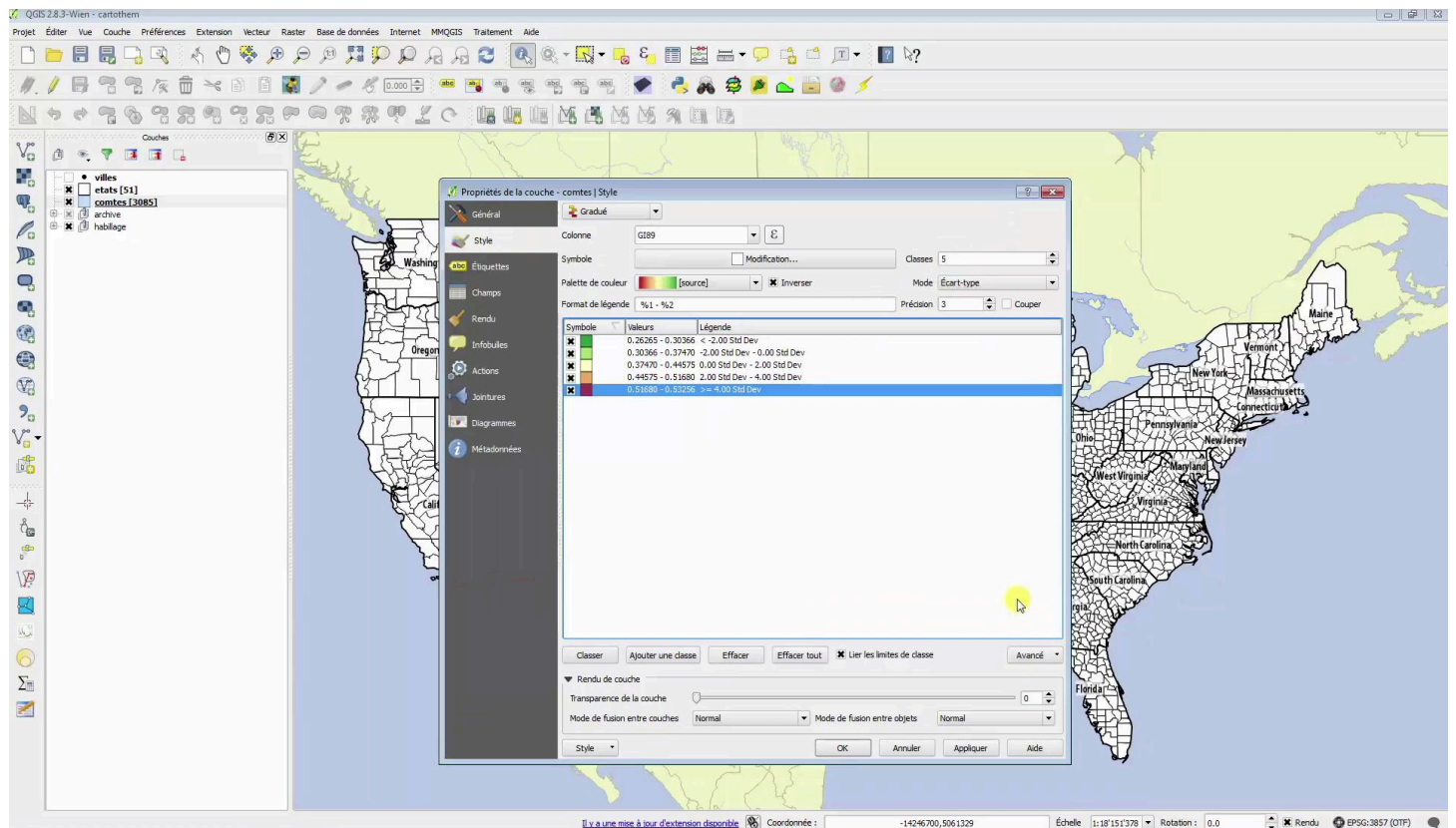
Dans un premier temps, nous allons jeter un coup d'œil à la distribution statistique de la variable du coefficient de Gini des comtés américains pour l'année 1989 et appelée GI 89. La forme de la distribution nous aidera à déterminer la méthode de discrétisation adéquate. Dans QGIS, on a préalablement ouvert un projet contenant la couche des comtés américains. Dans l'extension "Statist", que vous aurez préalablement installée, on sélectionne la couche "comté", puis la variable "GI 89" avant de cliquer sur OK. La distribution est presque normale, légèrement asymétrique, avec une moyenne un tout petit peu supérieure à la médiane. On peut dans ce cas-là utiliser la méthode de l'écart à la moyenne qui permet de générer des classes d'étendue identique à l'exception des classes extrêmes. L'étendue d'une classe sera égale à l'écart-type. Pour ce faire, un double-clic sur la couche "comté" permet d'atteindre les propriétés de la couche et l'onglet "style". Dans le menu déroulant situé en haut de la fenêtre, choisir "grader", puis sélectionner la colonne "GI 89".

Notes

Summary



33m 19s



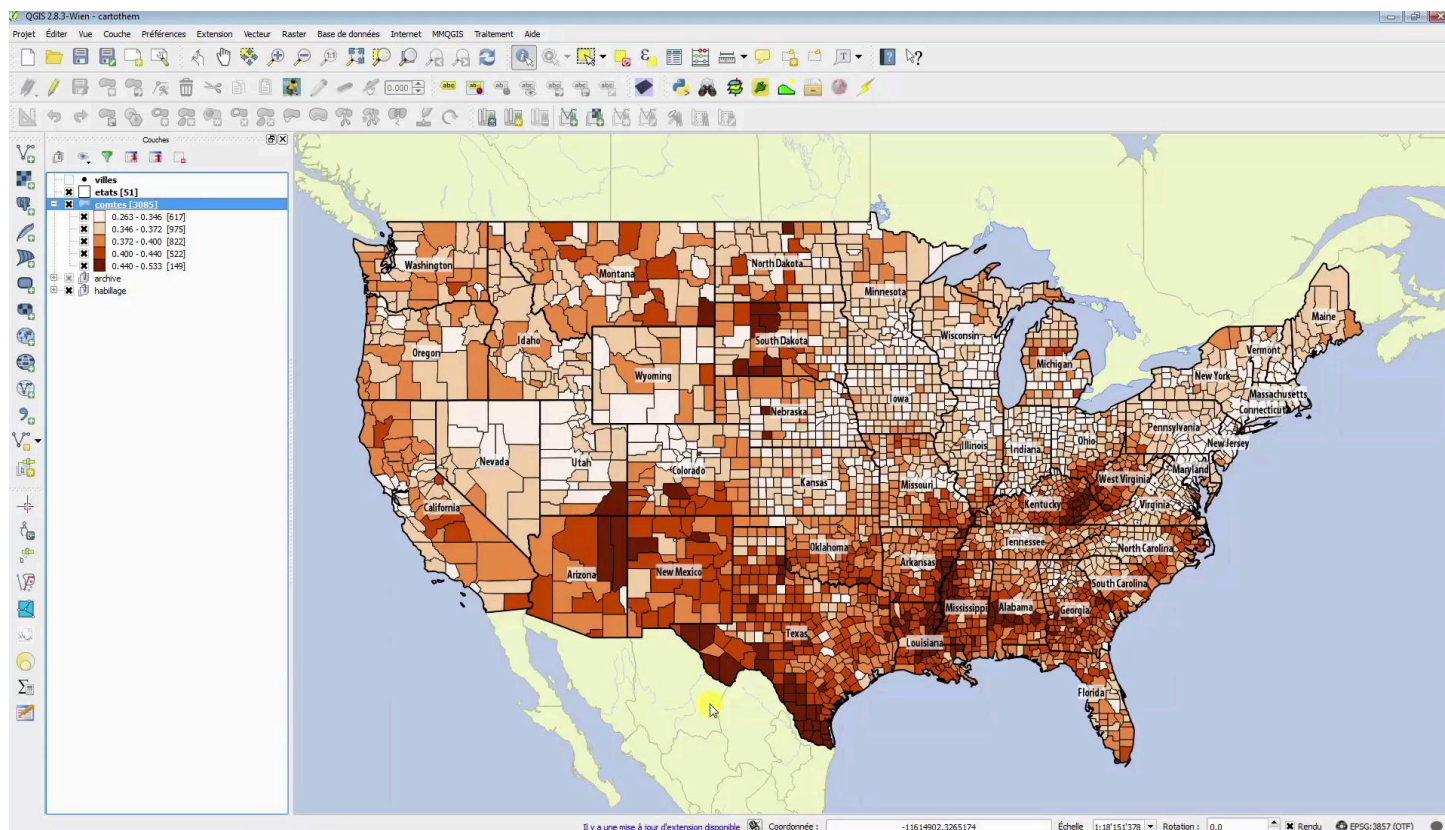
Choisir cinq classes et la palette de couleurs prédéfinies "red, yellow, green" qui est divergente avec une classe centrale en jaune clair et ensuite, on va inverser la palette de manière à ce que des indices d'inégalité élevés soient affichés en rouge puis, on choisit le mode de discrétisation "écart-type" et enfin, on clique sur le bouton "classer". A ce stade, il est possible de recourir à l'outil "ColorBrewer" décrit dans la leçon précédente. On peut en effet utiliser une palette soigneusement mise au point dans le but d'optimiser la perception des seuils entre les teintes, bien que les palettes de bases fournies par QGIS fournissent le plus souvent une solution acceptable. Sur l'interface de ColorBrewer, il faut choisir les paramètres désirés, soit le nombre de classes, ici cinq, le type de données, ici divergeant, puis la palette "red, yellow, green" correspondante. Suivant le but poursuivi, on peut également préciser certaines options dont la génération de palette utilisable par des personnes ayant des problèmes d'anomalie de perception. Ensuite, il faut reporter les valeurs des classes au format hexadécimal par exemple, par copier-coller entre ColorBrewer et QGIS.

Notes

Summary

34m 30s





Une fois que les teintes ont toutes été attribuées, on peut cliquer sur OK et on obtient le résultat suivant. Vous remarquerez que pour ce mode de discrétisation basé sur l'écart à la moyenne, QGIS n'autorise que cinq ou neuf classes, ce qui implique que la classe entre 0 et 2 écart-type prend la teinte neutre jaune pâle, alors que la classe symétrique entre 0 et moins 2 écart-type prend une teinte vert clair non neutre. Il faudrait pouvoir dans ce cas-là recourir à un nombre pair de classes et une palette en deux teintes, rouge et vert, de manière à pouvoir représenter l'opposition de comportement entre ces deux classes sur la carte. Néanmoins, la carte produite fait clairement ressortir les comtés du sud du Texas à la frontière avec le Mexique, et le long du Mississippi dans le sud des États-Unis. On peut également utiliser la distribution des valeurs de Gini séquentiellement avec un dégradé de la teinte orange en appliquant la méthode des seuils naturels en cinq classes. Cette représentation met mieux en évidence les comtés où les inégalités sont les plus fortes.

Notes

Summary

35m 50s

