

# Construction d'une carte choroplèthe

## I. Carte choroplèthe : maillage du territoire

« Le maillage 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). Ces mailles seront remplies par des aplats de couleurs dégradés représentant les données étudiées.

Le découpage irrégulier signifie que les aires géographiques ne sont pas identiques, que les frontières ne sont pas régulières. Cela correspond bien souvent aux découpages administratifs d'un pays en villes, départements et régions.

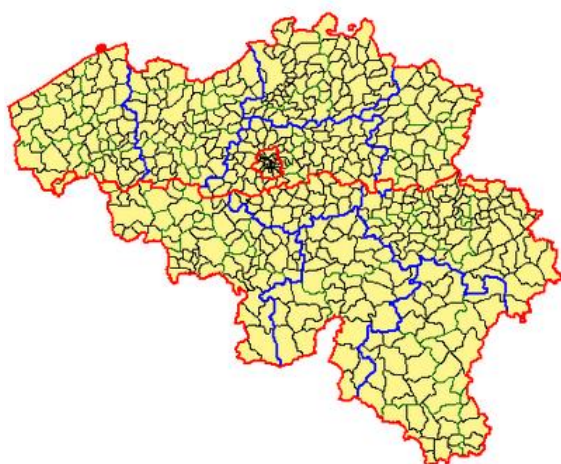


Figure 1 Découpage administratif de la Belgique

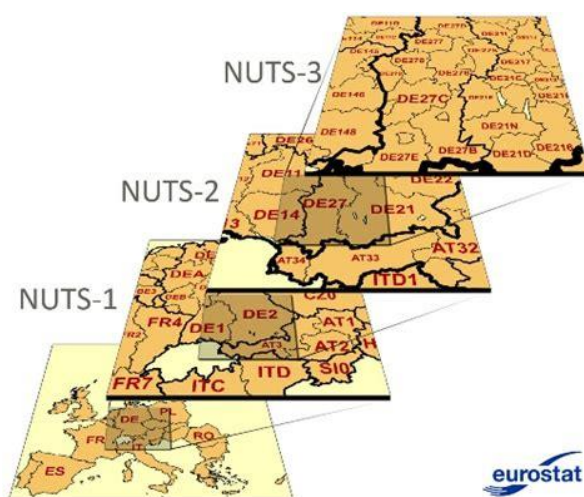


Figure 2 Découpage européen

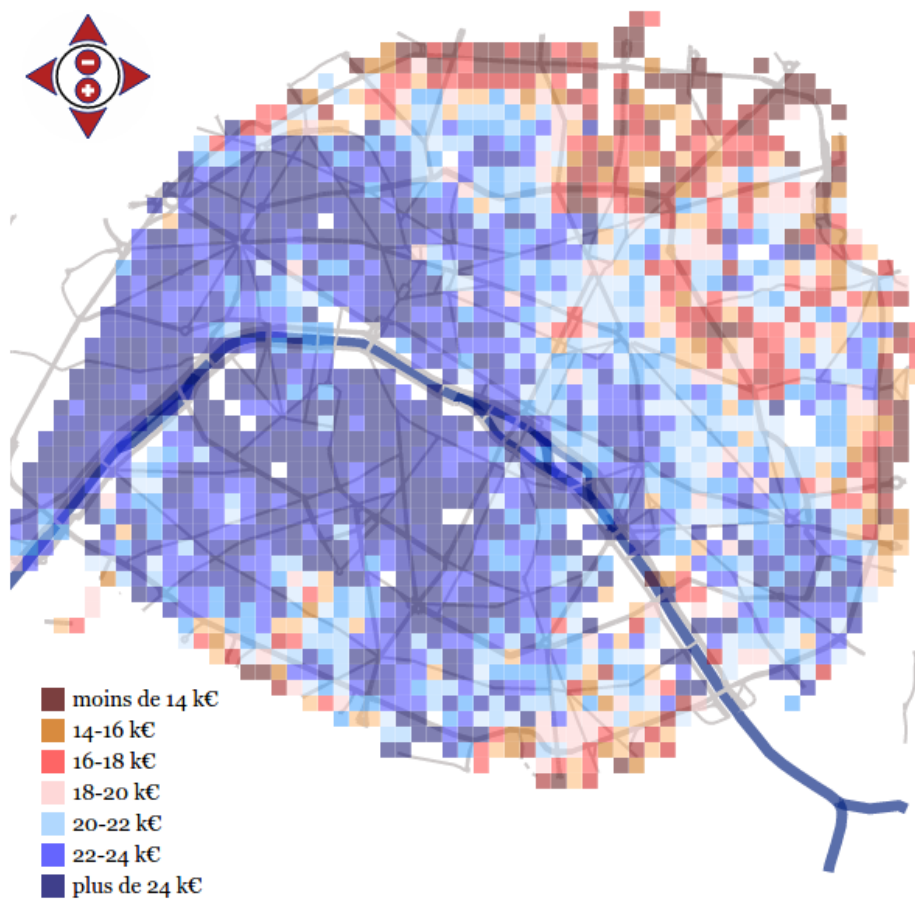
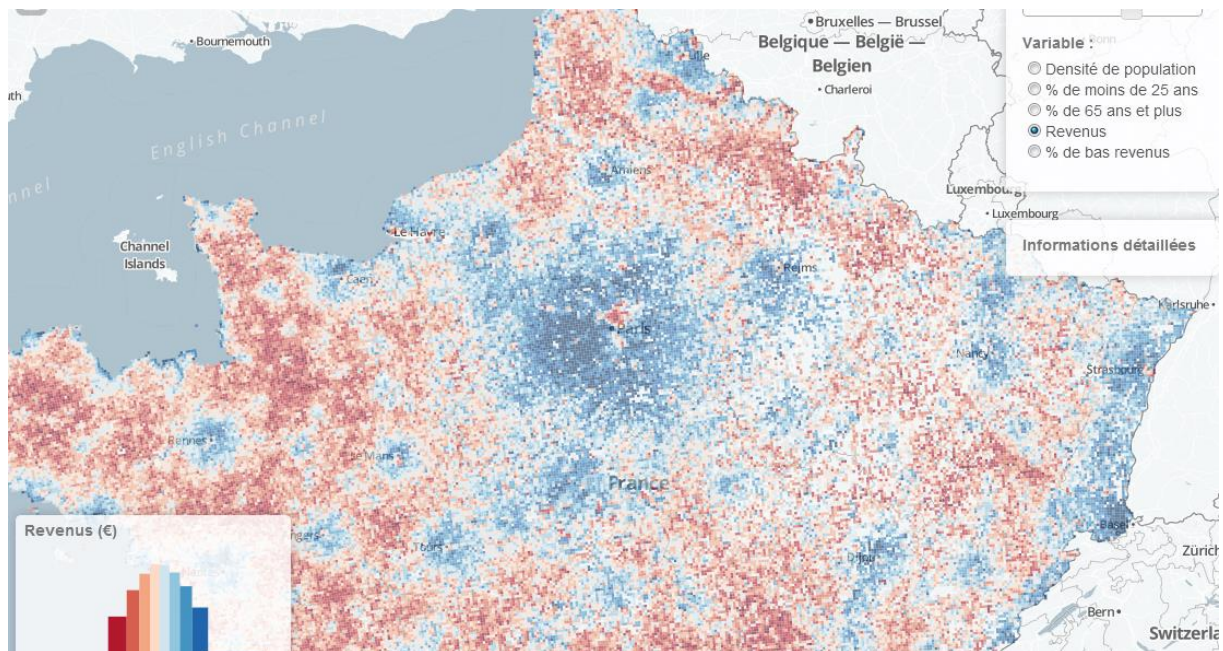
Il existe un découpage administratif Européen : NUTS pour Nomenclature des Unités Territoriales Statistiques. Ce découpage est utile pour travailler à l'échelle Européenne avec des données du site Eurostat. Je ne développerai pas plus le sujet mais voici quelques liens :

<https://ec.europa.eu/eurostat/fr/web/nuts/background>

<https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts>

« Le carroyage est une technique de quadrillage consistant à découper le territoire en carreaux pour y diffuser de l'information statistique à un niveau faiblement agrégé. Le maillage du territoire qui en résulte est plus ou moins fin selon la taille de carreau choisie. Le carroyage nécessite de disposer initialement de données pour lesquelles on connaît précisément la position géographique de chaque observation.

Les carreaux permettent de s'affranchir des limites administratives habituelles et peuvent être assemblés pour construire ou approcher n'importe quel zonage à façon. L'information statistique sur un tel zonage peut ensuite être facilement retrouvée en rassemblant les données des carreaux qui le constituent. Les carreaux ont également l'avantage d'être stables au fil du temps. » INSEE

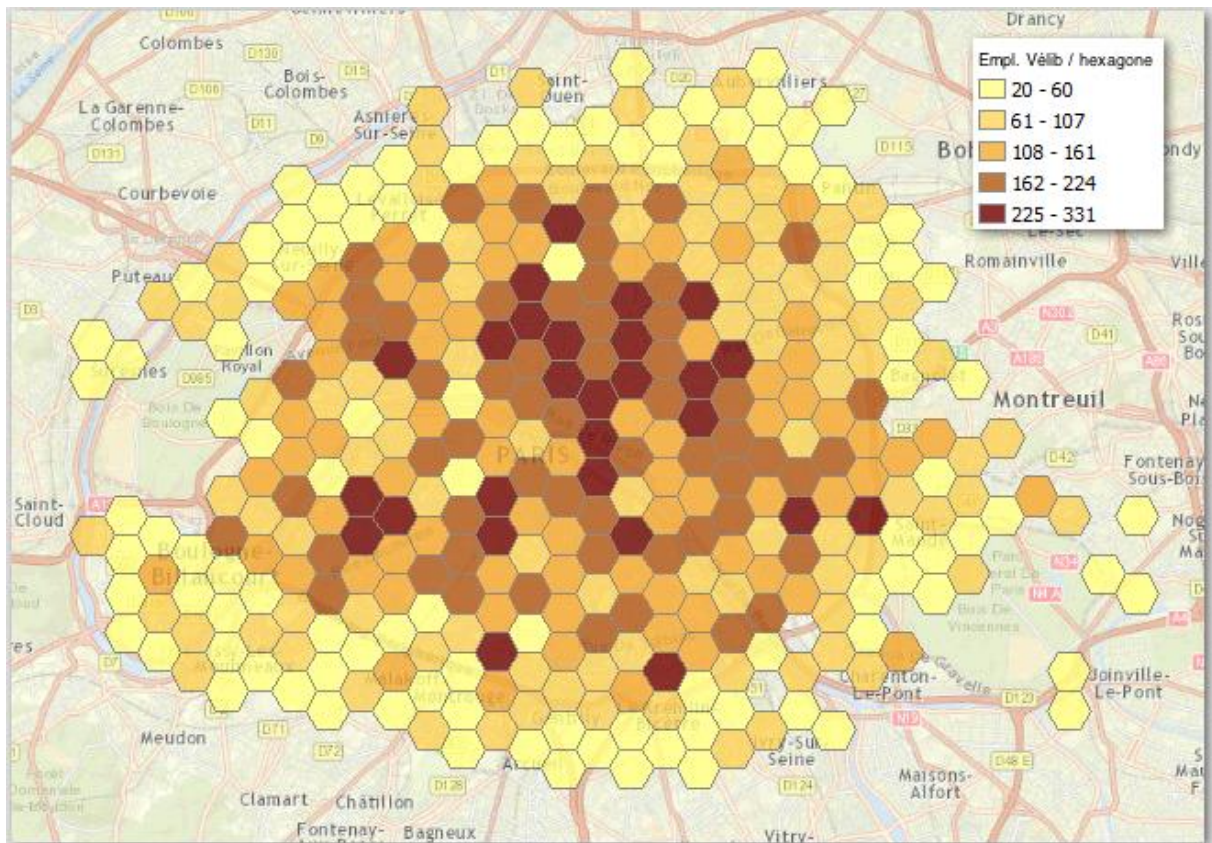


Le carroyage le plus répandu, représenté sur les deux cartes précédentes, l'est sous forme de carrés, semblables à des pixels, dont la taille est définie par l'utilisateur ou imposée par les données.

L'INSEE propose également un découpage par niveau « naturelle ». Voir script exemple en annexe.  
<https://www.insee.fr/fr/statistiques/4176281?sommaire=4176305#consulter>



Mais il existe une autre forme de découpage possible sous forme d'hexagone. Dans ce découpage, toutes les cellules voisines partagent des bords et non des points comme ça l'est dans le découpage en carré. La distance entre le centre de chaque hexagone et le centre de ses 6 hexagones adjacents est constante.



Le maillage est à choisir suivant :

- ✓ l'analyse du territoire que vous souhaitez faire : suivant la problématique que vous traitez, l'étude peut être locale, départementale, mondiale. Plus le territoire étudié est petit, plus le choix du maillage est important ;
- ✓ les données que vous avez collectées et recueillies : si les données sont locales, il est toujours possible de les agréger pour travailler à autre échelle, mais si les données sont déjà agrégées, le maillage nous a imposé par les données.

## II. Carte choroplèthe : Discrétisation des données

Maintenant que nous avons vu les fonds de carte choroplèthe, découpé en mailles, nous devons les remplir ces mailles avec nos données.

La discrétisation est **une méthode de classement** utilisée dans les métiers de la statistique, qui est destinée à **améliorer la lecture des données** recueillies en les **classifiant selon des critères justifiables**. C'est une « simplification », en partant de données complètes et précises, on obtient des valeurs condensées en classe, qui nous permet de dire que des valeurs appartenant à une même classe sont considérées identiques.

Les classes formées doivent **couvrir l'ensemble de la distribution (exhaustivité), ne doivent pas être jointives (disjonction), ni vides**. Une bonne discrétisation est obtenue par **le meilleur compromis entre la fidélité aux données initiales et une représentation visuelle satisfaisante**.

Pour choisir la bonne méthode de discrétisation, il faut avoir une excellente connaissance de ses données, avoir réalisé les calculs statistiques suivant : moyenne, écart-type, quantile, variance et connaître les valeurs minimum et maximum et connaître la forme de la distribution de la variable étudiée.

#### 1. Discrétisation par équivalence (ou amplitudes égales)

Prendre les valeurs max et min de la distribution, calculer l'étendue (max-min), puis diviser l'étendue par le nombre de classe souhaité pour obtenir l'étendue de chaque classe. Cette méthode donne des cartes faciles à lire et à comprendre mais ne permet pas de comparaison entre cartes.

#### 2. Discrétisation par effectifs égaux (ou quantiles)

Cette méthode suppose que la répartition des données soit équilibrée. On essaie de construire des classes ayant le même nombre d'individus. Si il y a 4 classes, on utilise le découpage en quartiles, on sait alors que chaque classe représente 25% des observations, 5 classes correspond à celle des quintiles et dans ce cas, chaque classe représente 20% des observations.

#### 3. Discrétisation par progression arithmétique

Avec cette méthode, l'étendue des classes augmente de façon arithmétique : la 2<sup>e</sup> classe a une étendue double de celle de la 1<sup>ère</sup> classe, et la 3<sup>e</sup> classe cumule l'étendue de la 1<sup>ère</sup> et de la 2<sup>e</sup>. Les classes sont d'amplitude croissante, c'est-à-dire que les faibles valeurs sont finement représentées tandis que les fortes valeurs sont regroupées dans la dernière classe. S'utilise essentiellement sur des séries de données dissymétriques, présentant une asymétrie à gauche.

#### 4. Discrétisation par progression géométrique

Cette méthode est appropriée aux distributions asymétriques comportant une très forte représentation des valeurs faibles et pour les distributions exponentielle et logarithmique décroissante. Les classes sont ainsi découpées suivant une progression géométrique.

#### 5. Discrétisation par déviation standard (ou écart-type)

Cette méthode consiste à utiliser l'écart-type – à la moyenne – pour calculer l'amplitude des classes. C'est donc l'écart à la moyenne qui est représenté, cela à l'avantage de mettre en avant les zones de fortes concentrations par rapport à un « bruit de fond ». La moyenne constitue une borne de classe ou le centre de la classe. Les autres bornes sont des multiples et fractions de l'écart type. Cette méthode est idéale pour les comparaisons mais le découpage des classes n'est pas simple à comprendre pour le lecteur.

## 6. Discrétisation grâce aux seuils naturels

Cette méthode consiste à observer la distribution d'une série statistique, repérer les trous dans la distribution, en représentant un histogramme, et les utiliser comme bornes. Simple à mettre en œuvre sur les données discontinues, elle est très subjective et risque de produire des classes très inégales.

## 7. Méthodes de Jenks (ou répartition automatique)

Méthode automatique et itérative, qui utilise les plus fortes discontinuités observées dans la distribution constituent des paliers marquant qu'il faut séparer en classes. La seule différence avec la méthode des seuils naturels repose sur le fait que les seuils sont calculés par un algorithme de Jenks. Cet algorithme minimise la variance au sein d'une même classe et maximise la variance entre les classes.

## 8. Choix des couleurs

A chaque classe doit correspondre une couleur.

- Si la série de données ne comporte que des valeurs positives, alors vous ne devrez utiliser une seule couleur qui sera dégradée ;
- Si la série comporte des valeurs positives et négatives, alors 2 dégradés de couleur peuvent être utilisés ;
- Il est possible d'utiliser plusieurs couleurs même si les valeurs ne sont que positives. Il s'agit d'attribuer une couleur à chaque classe mais cela est fortement déconseillé.

Voici un outil très intéressant qui montre, à partir d'un même jeu de données, les cartes obtenues en faisant varier le nombre de classes et le type de discrétisation :

<http://mappemonde.mgm.fr/geovisu/119/>

Vous trouverez un tableau récapitulatif des méthodes de discrétisation ici :

<http://soshg.free.fr/formation/discretisation.htm>

**Sous R**, la discrétisation des données se fait « simplement », grâce à la fonction **getBreaks()** du package **cartography**.

<https://www.rdocumentation.org/packages/cartography/versions/2.3.0/topics/getBreaks>

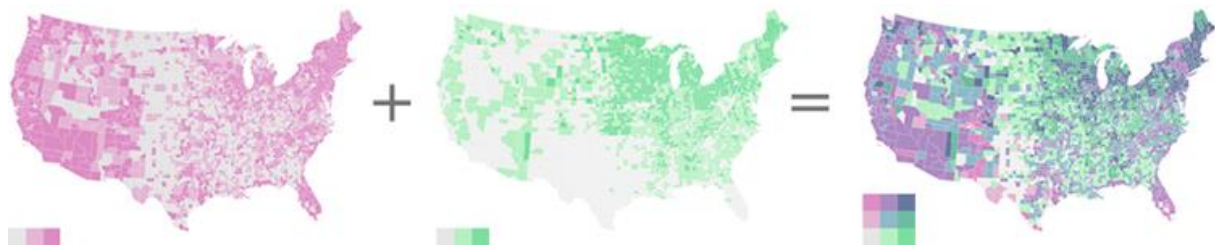
La fonction « **choroLayer** » permet de réaliser une carte choroplèthe.

<https://www.rdocumentation.org/packages/cartography/versions/2.4.1/topics/choroLayer>

Remarque : Une nouvelle version du package cartography a été mise en ligne le 2 Avril 2020

Pour aller plus loin, il est possible de représenter sur une même carte, deux variables par deux aplats de couleur. Cette carte constitue une analyse bivariable, permettant d'aller plus loin dans l'analyse que

la représentation et l'interprétation de deux cartes univariées. Une telle carte montre si deux variables sont plutôt concordantes ou discordantes.



Ci-dessus, nous avons 2 cartes choroplèthes univariées, chacune est constituée de données réparties en 3 classes. En agrégeant ces deux cartes, nous obtenons une 3<sup>e</sup> carte bivariée de 9 classes.

### III. Jointure

Maintenant que vous disposez d'un fond de carte vectoriel et d'au moins un fichier de données, vous devez créer une jointure entre les deux [si nos données sont locales (sous forme de points) cela n'est pas indispensable]. La jointure a pour fonction de lier les deux fichiers en faisant correspondre les données géographiques de chacun d'eux. Cela se fait sous R, avec des fonctions tel que :

- **merge()** <https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/merge>
- **join()** : cette fonction est plus complexe que merge <https://www.rdocumentation.org/packages/dplyr/versions/0.7.8/topics/join>

Merge() devra être utilisée pour joindre 2 fichiers par une colonne identique. Exemple :

1<sup>er</sup> fichier

Ville	Nombre habitant	Nombre crime
Paris	1000000	129
Marseille	800000	86
Lille	4500	12

2<sup>e</sup> fichier

Département	Ville	Longitude	Latitude
59	Lille	3.05	50.62
75	Paris	2.36	48.85
13	Marseille	5.36	43.29

Après avoir exécuté la fonction merge(), avec pour jointure la colonne « ville », on obtient :

Département	Ville	Longitude	Latitude	Nombre habitant	Nombre crime
59	Lille	3.05	50.62	4500	12
75	Paris	2.36	48.85	1000000	129
13	Marseille	5.36	43.29	800000	86

Dans notre exemple, partir d'un fichier contenant le nombre de crime par ville, et un autre les villes géocodées, nous obtenons un tableau avec des crimes localisés. En pratique, notre premier fichier correspond au fond de carte vectoriel et le second fichier, à vos données. Merge permet de créer un lien entre ces deux entités distinctes et de construire la carte.

Dans votre étude, le fichier 1 est le fond de carte, le fichier 2 celui contenant les données à représenter.

**Important :** Vos fichiers doivent avoir le même nombre de lignes pour que la fonction merge fonctionne !

#### IV. Exercice pratique

Créer une carte choroplèthe sous R à partir de vos données ou de données socio-économique trouvées sur le site de l'INSEE, à l'aide du package cartography. Quelle méthode de discrétisation avez-vous utilisé ? Pourquoi ?

Vous pouvez poser vos questions ici : <http://socrates.io/#CRM205Coursdu21Avril>

## ANNEXE

### CartoINSEE.R

Lefebvre Stéphanie

2020-04-21

2020-04-21

```
library(cartography)

library(rgdal)

## Loading required package: sp

## rgdal: version: 1.4-4, (SVN revision 833)
## Geospatial Data Abstraction Library extensions to R successfully loaded
## Loaded GDAL runtime: GDAL 2.2.3, released 2017/11/20
## Path to GDAL shared files: C:/Users/Utilisateur/Documents/R/win-library
/3.6/rgdal/gdal
## GDAL binary built with GEOS: TRUE
## Loaded PROJ.4 runtime: Rel. 4.9.3, 15 August 2016, [PJ_VERSION: 493]
## Path to PROJ.4 shared files: C:/Users/Utilisateur/Documents/R/win-libra
ry/3.6/rgdal/proj
## Linking to sp version: 1.3-1

library(cartography)

## Warning: package 'cartography' was built under R version 3.6.3

setwd("C:/Users/Utilisateur/Desktop/Carto_Drame_Juillet_2019/France")

zone4 <- readOGR("Filosofi2015_carreaux_niveau_naturel_reg04.shp",
                 layer="Filosofi2015_carreaux_niveau_naturel_reg04")

## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\Utilisateur\Desktop\Carto_Drame_Juillet_2019\France\Fi
losofi2015_carreaux_niveau_naturel_reg04.shp", layer: "Filosofi2015_carreau
x_niveau_naturel_reg04"
## with 886 features
## It has 31 fields

par(mar = c(1, 1, 1, 1))
plot(zone4)

cols <- carto.pal(pal1 = "red.pal",
                  n1 = 5)

Classe <- getBreaks(zone4$Men_pauv, nclass = 5, method = "quantile",
                    k = 1, middle = FALSE)

choroLayer(spdf = zone4,
```



```

var = "Men_pauv",
breaks = Classe,
col = cols,
border = "grey40",
lwd = 0.5,
legend.pos = "right",
legend.title.txt = "Nombre de\nménages\nnpauvres",
legend.values.rnd = 2,
add = TRUE)

```

```

layoutLayer(title = "Représentation du nombre de ménages pauvres à la réunion",

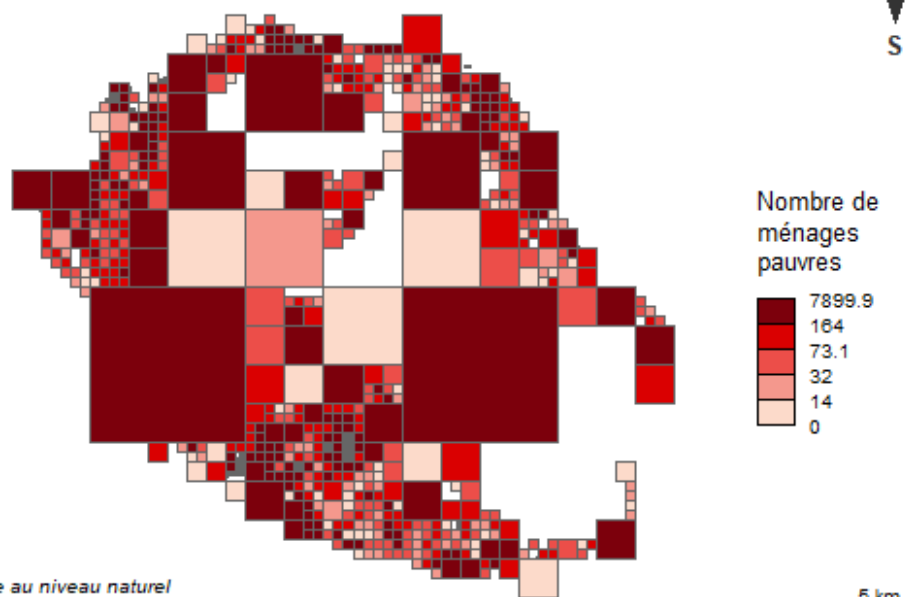
```

```

sources = "INSEE 2019/ Carroyage au niveau naturel ",
author = "S. Lefebvre, 2020", scale = 5, south = TRUE,
frame = FALSE, col = "#18423b", coltitle = "#a9f7ea")

```

### Représentation du nombre de ménages pauvres à la réunion



INSEE 2019/ Carroyage au niveau naturel  
S. Lefebvre, 2020

## ANNEXE

### CartoINSEE.R

LEFEBVRE Stéphanie

2020-04-13

```
library(rgdal)

## Loading required package: sp

## rgdal: version: 1.4-4, (SVN revision 833)
##  Geospatial Data Abstraction Library extensions to R successfully loaded
##  Loaded GDAL runtime: GDAL 2.2.3, released 2017/11/20
##  Path to GDAL shared files: C:/Users/Utilisateur/Documents/R/win-library
/3.6/rgdal/gdal
##  GDAL binary built with GEOS: TRUE
##  Loaded PROJ.4 runtime: Rel. 4.9.3, 15 August 2016, [PJ_VERSION: 493]
##  Path to PROJ.4 shared files: C:/Users/Utilisateur/Documents/R/win-libra
ry/3.6/rgdal/proj
##  Linking to sp version: 1.3-1

library(cartography)

## Warning: package 'cartography' was built under R version 3.6.3

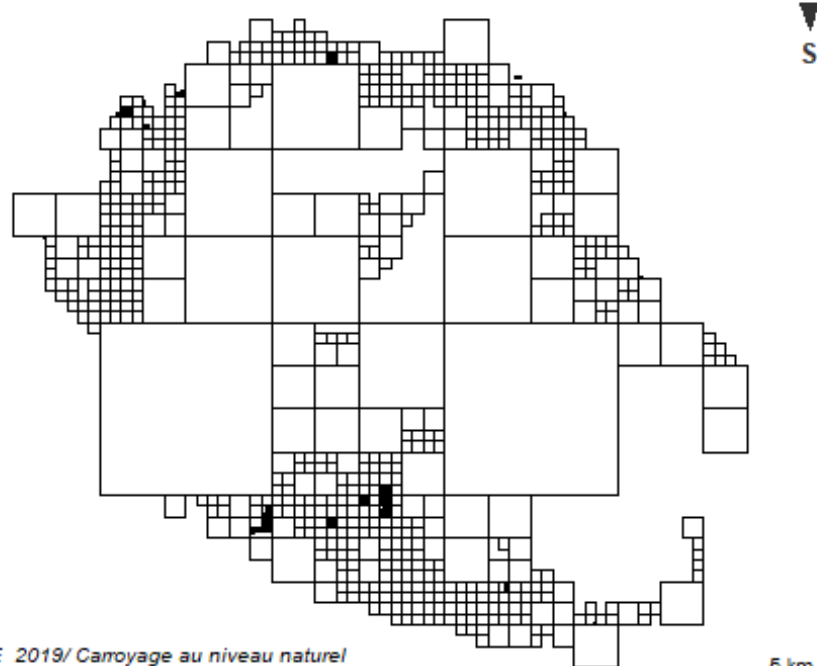
setwd("C:/Users/Utilisateur/Desktop/Carto_Drame_Juillet_2019/France")

zone4 <- readOGR("Filosofi2015_carreaux_niveau_naturel_reg04.shp",
                 layer="Filosofi2015_carreaux_niveau_naturel_reg04")

## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\Utilisateur\Desktop\Carto_Drame_Juillet_2019\France\Fi
losofi2015_carreaux_niveau_naturel_reg04.shp", layer: "Filosofi2015_carreau
x_niveau_naturel_reg04"
## with 886 features
## It has 31 fields

par(mar = c(1, 1, 1, 1))
plot(zone4)
layoutLayer(title = "Carte de la réunion",
            sources = "INSEE 2019/ Carroyage au niveau naturel ",
            author = "S. Lefebvre, 2020", scale = 5, south = TRUE,
            frame = FALSE, col = "#18423b", coltitle = "#a9f7ea")
```

## Carte de la réunion



INSEE 2019/ Carroyage au niveau naturel  
S. Lefebvre, 2020

5 km