

1. Analyse spatiale descriptive

SOPHIE AUDRIC, MARIE-PIERRE DE BELLEFON, ERIC DURIEUX

Insee

1.1	Différents types de données spatiales	4
1.1.1	Données ponctuelles	4
1.1.2	Données continues	5
1.1.3	Données surfaciques	6
1.2	Notions de sémiologie cartographique	7
1.2.1	Qu'entend-on par sémiologie cartographique?	7
1.2.2	Objectifs d'une carte	7
1.2.3	À chaque type de données, sa variable visuelle	7
1.2.4	Quelques conseils	8
1.3	Éléments de cartographie avec R	12
1.3.1	Manipulation d'objets spatiaux	13
1.3.2	Réalisation de cartes statistiques	18
1.3.3	<i>sf</i> : l'avenir du traitement des données spatiales sous R	21
1.3.4	De la surface au point, et réciproquement	24
1.4	Exemples d'études utilisant des données spatiales agrégées	27
1.4.1	Accès aux espaces verts - Statistique Suède	27
1.4.2	Taux de pauvreté régionaux - programme européen ESPON	27
1.4.3	Localisation optimale des éoliennes - Institut de cartographie de Grande Bretagne	28

Résumé

L'objectif de l'analyse spatiale est de comprendre et d'explorer l'intrication entre le positionnement spatial des objets et des phénomènes, et leurs caractéristiques. La littérature distingue traditionnellement trois types de données spatiales : données ponctuelles, données continues et données surfaciques. À chaque type de données correspondent des méthodes d'analyse spécifiques. Cependant, quelle que soit la nature des données spatiales, la première étape est de les prendre en main et de les agréger à une échelle géographique adaptée au processus spatial sous-jacent. Cartographier les données permet de synthétiser une information, de la rendre accessible à un public élargi et de réfléchir aux outils statistiques adaptés à la poursuite de l'étude. Cette première analyse descriptive peut également, dans la démarche d'une étude, être l'occasion de constater des problèmes particuliers dans les données (collecte, données manquantes, valeurs aberrantes, etc.) ou d'invalider certaines hypothèses nécessaires au développement de méthodes économétriques. Nous introduisons dans ce chapitre les notions de sémiologie cartographique utiles pour réaliser une carte de qualité.¹

1. Ces éléments de sémiologie sont extraits d'un ouvrage de l'Insee "Guide de sémiologie cartographique" publié en 2017 et auquel ont contribué un grand nombre de personnes que nous remercions.

Ce chapitre décrit la prise en main des données spatiales avec le logiciel R et la production de premières cartes descriptives. Des études réalisées dans divers instituts de statistique européens illustrent ces notions.

1.1 Différents types de données spatiales

Une donnée spatiale est une observation dont on connaît non seulement la valeur, mais aussi la localisation. Le support des observations, défini comme l'ensemble des coordonnées spatiales des objets à traiter, constitue une information potentiellement riche pour l'analyse.

Certaines propriétés des données spatiales contredisent les hypothèses nécessaires à l'utilisation des méthodes statistiques habituelles. Ainsi, l'hypothèse d'indépendance des observations, requise dans la plupart des modèles économétriques, n'est pas vérifiée en présence de *dépendance spatiale* : lorsque la valeur de l'observation i influence la valeur de l'observation j voisine. Les données spatiales peuvent aussi se caractériser par de l'*hétérogénéité spatiale* : l'influence des variables explicatives sur la variable dépendante dépend de la localisation dans l'espace ; une variable peut être influente sur une autre dans un voisinage donné, mais ne pas l'être dans un autre endroit. Pour analyser les données spatiales, de nombreuses méthodes spécifiques ont donc été développées.

Les méthodes et leurs objectifs dépendent de la nature des données spatiales. D'après la classification proposée par CRESSIE 1993b, on distingue trois types de données spatiales :

- données ponctuelles ;
- données continues ;
- données surfaciques.

La différence fondamentale entre ces données n'est pas la taille de l'unité géographique considérée mais le processus générateur des données.

1.1.1 Données ponctuelles

Les données spatiales ponctuelles se caractérisent par la **distribution dans l'espace** des observations. Le processus générateur des données génère les coordonnées géographiques associées à l'apparition d'une observation. On n'étudie pas de valeur associée à l'observation ; seule compte la localisation. Il s'agit, par exemple, du lieu d'apparition d'une maladie lors d'une épidémie, ou de la répartition dans l'espace de certaines espèces d'arbres. L'analyse spatiale des données ponctuelles a pour objectif de **quantifier l'écart entre la distribution spatiale des observations et une distribution complètement aléatoire dans l'espace**. Si les données sont plus regroupées que si elles étaient distribuées aléatoirement sur le territoire, on peut identifier des clusters et mesurer leur significativité.

R Les principales méthodes permettant d'analyser les données ponctuelles sont décrites dans le chapitre 4 : "Configurations de points".

■ **Exemple 1.1 — Détection de clusters.** FOTHERINGHAM *et al.* 1996 cherchent à détecter la présence de clusters significatifs de maisons inconfortables. Ils comparent la répartition spatiale de ces maisons avec la répartition qu'elles auraient si elles étaient distribuées aléatoirement parmi l'ensemble de toutes les maisons. Les hypothèses sur la distribution aléatoire dans l'espace permettent d'évaluer la significativité des regroupements de maisons (figure 1.1).

■

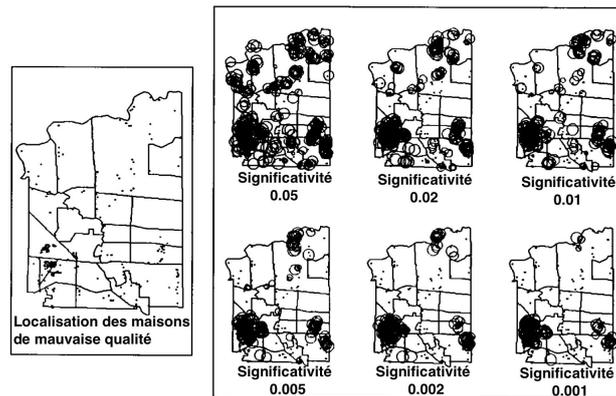


FIGURE 1.1 – Détection de clusters significatifs

Source : FOTHERINGHAM *et al.* 1996

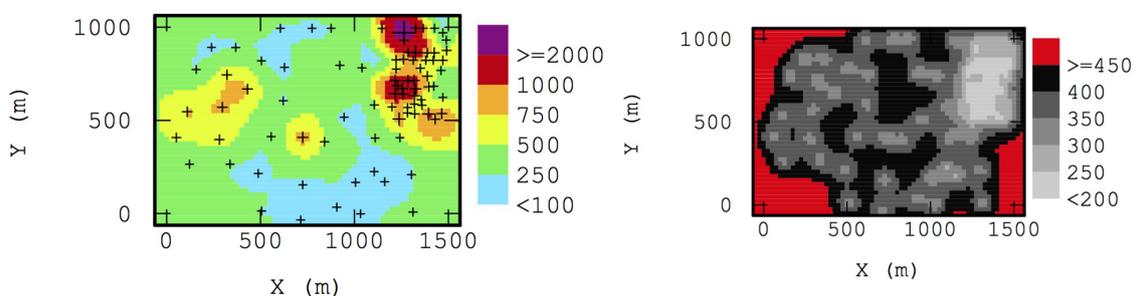
1.1.2 Données continues

En présence de données continues, il existe une valeur pour la variable d'intérêt en tout point du territoire étudié. Les données sont générées de façon continue sur un sous ensemble de \mathbf{R}^2 . En revanche, ces données sont mesurées uniquement en un nombre discret de points. Il s'agit, par exemple, de la composition chimique du sol (utile à l'industrie minière), de la qualité de l'eau ou de l'air (pour des études sur la pollution), ou encore de diverses variables météorologiques. L'analyse spatiale des données continues, appelée aussi géostatistique, cherche à prédire la valeur d'une variable en un point où elle n'a pas été échantillonnée, ainsi que la fiabilité de cette prédiction. La géostatistique aide également à optimiser le plan d'échantillonnage des données.

R Les principales méthodes permettant d'analyser les données continues sont décrites dans le chapitre 5 : "Géostatistique".

■ Exemple 1.2 — Prédiction de la pollution. CHILES *et al.* 2005

Les chercheurs du groupe de travail GeoSiPol (Les pratiques de la géostatistique dans le domaine des sites et sols pollués) prennent en compte la structure de dépendance spatiale entre les données grâce à la technique du *krigeage*. Ils prédisent la quantité de polluant en des lieux où le sol n'a pas été échantillonné et quantifient l'incertitude d'estimation (figure 1.2).

FIGURE 1.2 – Prédiction de la teneur en polluant d'un sol (mg/kg/m²) (à gauche) et écart-type de la prédiction (à droite)Source : Manuel GéoSiPol - Mines de Paris : CHILES *et al.* 2005

1.1.3 Données surfaciques

Pour des données surfaciques, la localisation des observations est considérée comme fixe, mais les valeurs associées sont générées suivant un processus aléatoire. Ces données caractérisent le plus souvent une partition du territoire en zones contiguës, mais elles peuvent également être des points fixes du territoire. Il s'agit, par exemple, du PIB par région, ou du nombre de mariages par mairie. Le terme "surfaccique" est donc trompeur, car ces données ne sont pas nécessairement représentées sur une surface. On s'intéresse aux relations entre les **valeurs des observations voisines**. L'analyse spatiale des données surfacciques commence par **définir la structure de voisinage des observations** puis elle **quantifie l'influence qu'exercent les observations sur leurs voisines**, et enfin, elle **évalue la significativité de cette influence**.

R Les principales techniques d'analyse des données surfacciques sont décrites dans les chapitres 2 : "Codifier la structure de voisinage", 3 : "Indices d'autocorrélation spatiale", ainsi que dans la partie 3.

■ **Exemple 1.3 — Dépendance spatiale locale.** GIVORD *et al.* 2016 cherchent à répondre à la question : "Les collèges favorisés sont-ils toujours situés dans un environnement favorisé?". Les auteurs utilisent pour cela des *Indices locaux d'autocorrélation spatiale*². Ces indices comparent la similarité entre le niveau social d'un collège et celui de son environnement à la similarité qu'ils auraient si les niveaux sociaux des collèges étaient répartis aléatoirement parmi l'ensemble des collèges. Les indices locaux d'autocorrélation spatiale permettent d'identifier les collèges pour lesquels l'influence du milieu social environnant est significative (figure 1.3). ■

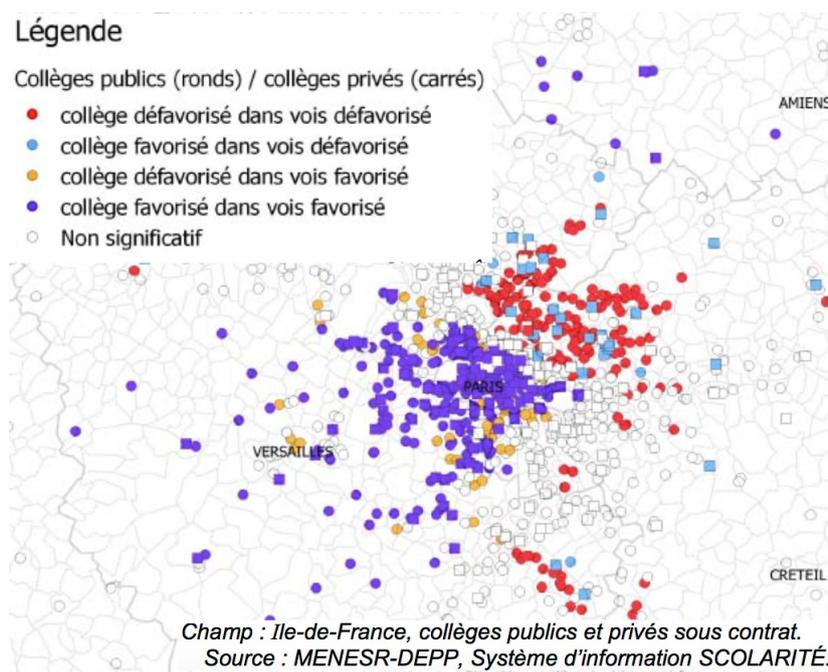


FIGURE 1.3 – Influence du niveau social du voisinage d'un collège sur le niveau social du collège lui-même

Source : GIVORD *et al.* 2016

2. On pourra se reporter au chapitre 3 : "Indices d'autocorrélation spatiale"

Encadré 1.1.1 — Une donnée spatiale peut faire partie de plusieurs catégories. La distinction en trois catégories de données spatiales permet d’orienter l’analyste vers telle ou telle méthode. Il faut néanmoins garder en tête que ces catégories sont perméables et que le choix d’analyser suivant un point de vue un phénomène est lié à l’échelle d’analyse et à l’objectif même de l’étude. Par exemple, une maison est considérée comme un objet ponctuel si on étudie les regroupements dans l’espace significatifs, mais ce peut-être aussi une donnée surfacique si on cherche à connaître la corrélation spatiale entre l’âge des habitants des maisons.

1.2 Notions de sémiologie cartographique

1.2.1 Qu’entend-on par sémiologie cartographique ?

La sémiologie cartographique est l’ensemble des règles qui permettent de transmettre le plus clairement possible une information correcte grâce à une image cartographique. Il est bon d’avoir ces règles en tête avant de passer à la réalisation pratique d’une carte avec le logiciel R. La sémiologie cartographique est un véritable langage destiné à faciliter la communication à l’aide d’outils graphiques appelés variables visuelles. La bonne utilisation de ces variables renforce le message tout en le rendant plus lisible.

Parmi les variables visuelles, on distingue la forme, la texture, la taille de l’objet à représenter, son orientation et sa couleur. Cette dernière peut être associée à des effets de transparence ou afficher un dégradé selon une échelle de valeurs donnée. La dynamique est une variable visuelle plus récente, avec l’apparition notamment des cartes animées.

Les variables visuelles se caractérisent par leur aptitude à mettre en évidence :

- des quantités, souvent représentées par des cercles proportionnels ;
- une hiérarchie, en représentant une série ordonnée de valeurs relatives, par exemple des densités de population ;
- des différences entre entités représentées, par exemple l’industrie et le tourisme ;
- des similitudes, en regroupant en un seul ensemble les différents objets d’un même thème.

Par ailleurs, une combinaison bien appropriée de plusieurs variables visuelles peut renforcer le message.

1.2.2 Objectifs d’une carte

Un graphique permet une appréhension directe et globale d’une information et remplace avantageusement un long tableau. C’est encore plus vrai pour une carte. Son principal intérêt est d’intégrer la dimension spatiale, surtout quand le nombre de territoires est relativement élevé. Ainsi, la carte permet d’un seul coup d’œil de percevoir une information. Grâce à la dimension spatiale, sont prises en compte la situation géographique, la proximité au littoral, à la montagne, aux grandes villes, aux pays voisins, etc. D’où l’importance de rajouter des repères géographiques : régions et pays voisins, noms de ville, fleuves, axes de communication, etc. De plus, la carte est un bon outil de communication. Elle est en effet de compréhension facile : on reconnaît généralement son territoire, et elle est une illustration plaisante. L’évolution technologique des outils de cartographie, gratuits et simples d’accès, permet de réaliser facilement des cartes esthétiques. Cependant, l’esthétisme ne doit pas primer sur la pertinence et encore moins déformer l’information apportée par la carte.

1.2.3 À chaque type de données, sa variable visuelle

La première question à se poser est de savoir ce que l’on veut représenter. En effet, pour représenter une variable en volume ou un chiffre absolu, on utilise des ronds proportionnels ; pour

des ratios, densités, évolutions, parts et typologie, on utilise une carte en aplats de couleurs ; les données bilocalisées ou les flux sont illustrés par des oursins, flèches proportionnelles ou résultantes vectorielles. Enfin, la localisation, par exemple d'équipements, se fait par des cartes à symboles.

Dans le cas d'une carte en aplats de couleurs (appelée aussi analyse en classes ou carte choroplète), les valeurs positives sont dans des teintes chaudes (rouge, orange) alors que les valeurs négatives sont généralement dans des teintes froides (bleu, vert). Par ailleurs, à une hiérarchie de valeurs correspond un dégradé de couleurs dont les couleurs les plus foncées (ou les plus claires) correspondent aux valeurs extrêmes.

Des règles existent également pour la discrétisation des données, c'est-à-dire la manière dont les observations sont regroupées en classes. Le nombre de classes se calcule en fonction du nombre d'observations. Différentes théories existent pour déterminer le nombre optimal. Selon la règle de Sturges par exemple, il est égal à $1 + 3,3 * \log_{10}(N)$, où N est le nombre d'observations.

En pratique :

- pour moins de 50 observations : 3 classes ;
- pour 50 à 150 observations : 4 classes ;
- pour plus de 150 observations : 5 classes.

La forme de la distribution des données nous aide aussi dans ce choix. Ainsi, on rajoute une classe en cas de présence de valeurs négatives et positives. Une fois le nombre de classes déterminé, une méthode de regroupement doit être choisie. Plusieurs méthodes existent, chacune d'elles présentant des avantages et des inconvénients.

- **La méthode des quantiles** : elle consiste à utiliser le même nombre de valeurs par classe. Elle produit une carte harmonieuse et facile à lire, les couleurs de la légende se répartissant à parts égales. Cependant, elle ne s'adapte pas toujours à la distribution des données.
- **La méthode des classes de même amplitude** : elle consiste à découper l'intervalle de valeurs en plages de même longueur. Cette méthode est simple à comprendre mais s'adapte très rarement à la distribution ; certaines classes peuvent ne contenir aucune valeur.
- **Les méthodes de Jenks et k-means** : elles visent à créer des classes homogènes en maximisant la variance entre les classes et en minimisant la variance au sein de chacune d'entre elles. Ces méthodes, contrairement aux deux précédentes, s'adaptent parfaitement aux données en éliminant les effets de seuil. Cependant, le temps de calcul de la méthode de Jenks peut être très long si les observations sont nombreuses. Pour cette raison, on peut utiliser la méthode k-means dont le calcul est plus rapide même avec un nombre élevé d'observations. Celle-ci peut cependant être instable, en donnant des classes différentes pour un même jeu de données. On gère ce problème en répétant la k-means plusieurs fois pour garder la meilleure répartition.
- **La méthode de l'arrangement manuel** : elle consiste à fixer soi-même les bornes des classes. Elle est utile pour faire apparaître des valeurs significatives (borne à zéro ou autour de zéro, moyenne. . .) ou pour améliorer à la marge le positionnement de certains seuils en fonction de la distribution locale. Elle permet également de rendre des cartes comparables entre elles en fixant des bornes identiques de classes. Cette méthode nécessite d'analyser au préalable la distribution des données, en utilisant dans un premier temps la méthode de Jenks ou de k-means pour avoir des classes homogènes puis en ajustant les bornes des classes manuellement pour éviter les effets de seuil.

1.2.4 Quelques conseils

- **Un message simple par carte.** Une carte est souvent difficilement compréhensible quand elle comporte trop d'informations. Par exemple, aucun message ne se dégage de la carte présentée en figure 1.4 car elle est trop compliquée. D'où la règle élémentaire de faire simple

pour être efficace. Pour ce faire, le nombre de variables à représenter sur une même carte doit être limité.

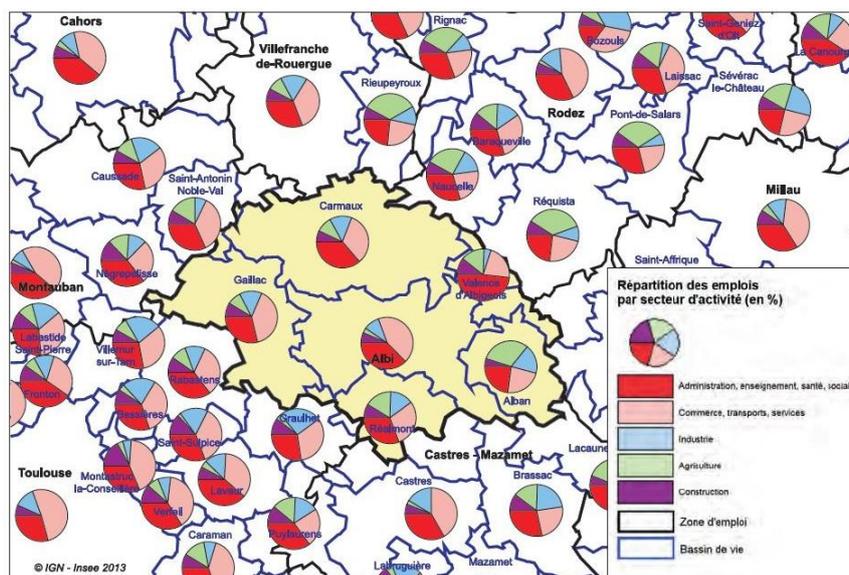


FIGURE 1.4 – Répartition des emplois par secteur d'activité dans les bassins de vie
Source : Insee, recensement de la population 2010

- **Faire figurer les informations de base.** Une carte doit impérativement comporter un titre informatif (le plus souvent associé à un sous-titre descriptif), une mention du zonage représenté, une légende, une source et un copyright. L'échelle, un logo ou la flèche Nord peuvent éventuellement y figurer.
- **Ne pas présenter le territoire comme une île.** Il est souhaitable de donner des éléments d'environnement au lecteur pour qu'il situe le territoire représenté ; par exemple, les départements ou régions limitrophes, des éléments de topographie comme la mer ou le réseau routier. Dans la figure 1.5, il aurait été judicieux de représenter les communes des départements environnants, notamment Dijon au nord ou Lyon au sud pour pouvoir illustrer le titre qui n'est pas très explicite.

Par ailleurs, il peut être intéressant d'élargir les analyses effectuées sur le territoire à l'environnement alentour, à la condition que le territoire d'intérêt ressorte bien comme dans la figure 1.6 (contour vert foncé et trame vert clair). L'analyse élargie permet ici de situer le dynamisme démographique de Toulouse par rapport à celui de Bordeaux et de mieux comprendre l'importance du système urbain languedocien, dans la continuité de celui du couloir rhodanien.

- **Des cartes comparables.** Lorsque deux cartes illustrant le même territoire avec les mêmes variables visuelles sont disposées côte à côte ou l'une en dessous de l'autre, le lecteur est incité à faire des comparaisons. Pour faciliter cette opération, les deux cartes doivent avoir une légende harmonisée (mêmes classes, cercles ou flèches) et une même échelle avec un zoom identique. Dans les cartes de la figure 1.7, les légendes harmonisées permettent de comparer l'évolution annuelle de la population sur les deux périodes 1982-2011 et 2006-2011.
- **Choisir son indicateur : parts ou effectifs ?** L'analyse en classes est utilisée pour représenter une sous-population en valeur relative (ou part) ou une évolution. Elle est prohibée pour la représentation d'effectifs ou de volumes car elle pourrait induire le lecteur à interpréter la carte de manière erronée. L'œil établirait en effet une correspondance entre le volume repré-

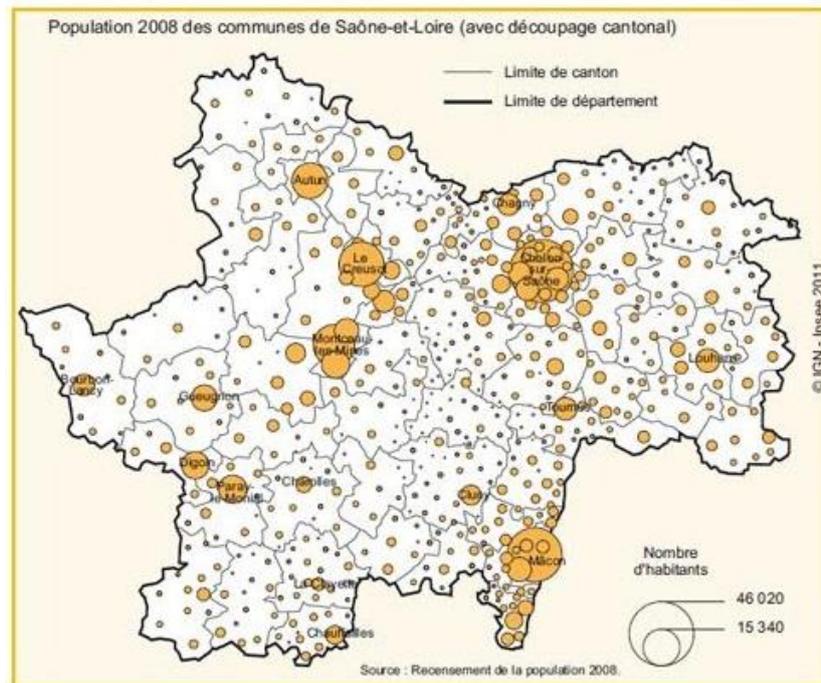


FIGURE 1.5 – Plusieurs villes moyennes
 Source : Insee, Recensement de la Population 2008

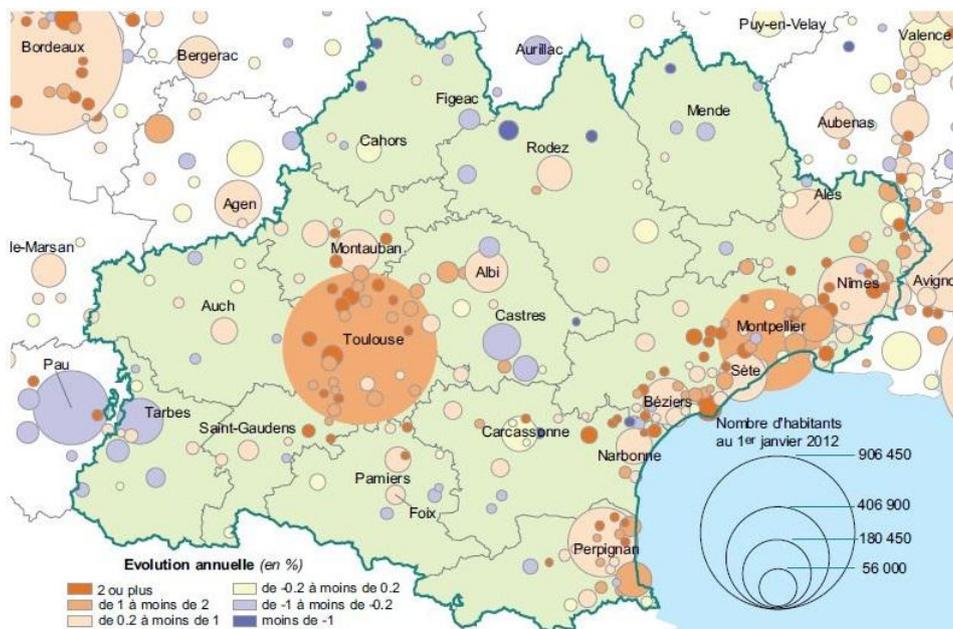
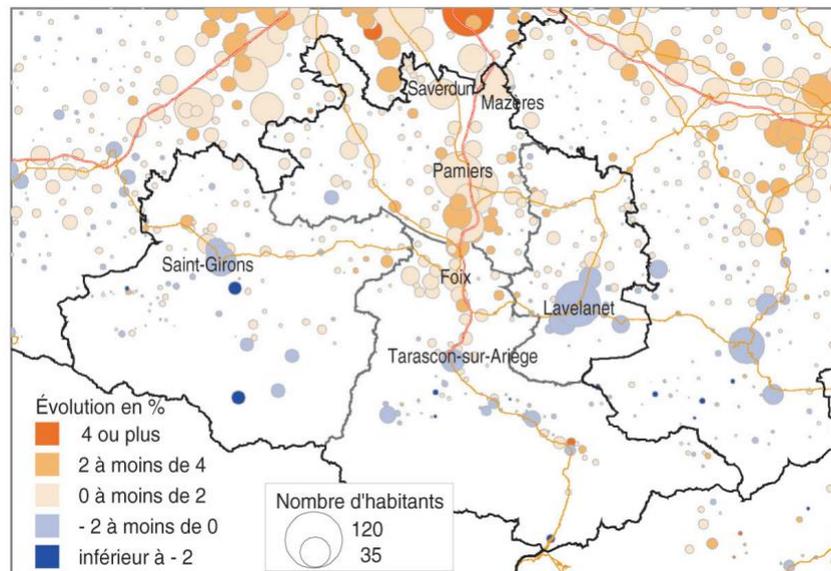
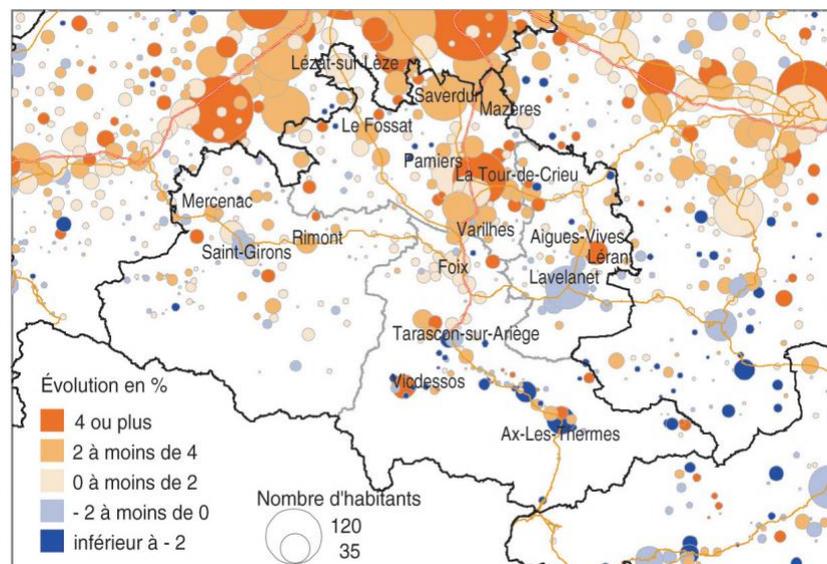


FIGURE 1.6 – Un système urbain monocentré autour de Toulouse et polycentré sur le littoral
 Source : Insee, recensements de la population 2007 et 2012



(a) Évolution annuelle de la population des communes entre 2006 et 2011



(b) Évolution annuelle de la population des communes entre 1982 et 2011

FIGURE 1.7 – Évolution annuelle moyenne de la population des communes de Basse-Ariège
 Source : Insee, Recensements de la Population 1982, 2006 et 2011

senté et la surface du territoire colorié. Ainsi, une analyse en classes sur le nombre d'habitants par commune induirait une surestimation visuelle de la population d'Arles, commune la plus étendue de France. Par ailleurs, une analyse en classes seule peut parfois être trompeuse car des pourcentages élevés peuvent concerner des petits effectifs. C'est pourquoi, il est parfois nécessaire de combiner ce type d'analyse avec une analyse en ronds proportionnels portant sur les effectifs. Selon le message que l'on veut faire passer, on choisira de colorier des ronds avec une analyse en classes (figure 1.7) ou de plaquer des ronds sur une analyse en classes (figure 1.8). Dans le cas de ronds coloriés, l'œil est davantage attiré par la taille des ronds et dans l'autre cas, l'œil sera d'abord attiré par les couleurs les plus foncées de l'analyse en classes.

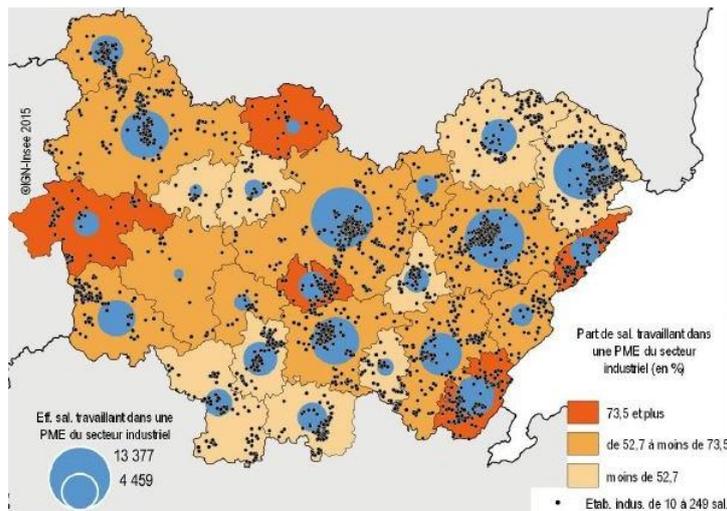


FIGURE 1.8 – Répartition des salariés travaillant dans une PME du secteur industriel
Source : Insee, *Connaissance Locale de l'Appareil Productif 2012*

1.3 Éléments de cartographie avec R

Les données géolocalisées peuvent être agrégées à une échelle géographique plus ou moins grande. On peut ensuite les cartographier de différentes façons. Nous décrivons dans cette section comment appréhender simplement la cartographie avec R, et quelques packages appropriés. De nombreux packages permettent de représenter des données spatiales ; ceux que nous mettrons en œuvre dans ce manuel sont :

- *sp* : package de base définissant les objets spatiaux ;
- *rgdal* : import/export d'objets spatiaux ;
- *rgeos* : manipulation de la géométrie ;
- *cartography* : réalisation de cartes d'analyse.

Nous présenterons également le package *sf* qui regroupe l'ensemble des fonctions des packages *sp*, *rgdal* et *rgeos*.

1.3.1 Manipulation d'objets spatiaux

Points, Polygones, Lignes

Le package *sp* permet de créer ou de convertir en objet *sp* différentes géométries : des points, des lignes, des polygones ou encore des grilles. En général, chaque objet *sp* est composé de différentes parties : les slots. Chaque slot contient une information particulière (coordonnées géographiques, table d'attributs, système de coordonnées, étendue spatiale, etc.)

L'accès à un slot d'un objet *sp* se fera à l'aide de l'opérateur @ (objet@slot).

Les objets spatiaux peuvent être abordés sous différentes formes. La première correspond à des **points**, c'est-à-dire un ensemble de points géoréférencés.

```
library(sp)
```

```
# contenu d'une table communale contenant les coordonnées des mairies
# en WGS84 (latitude/longitude)
```

```
head(infoCom)
```

```
##          nom_commune latitude longitude préfecture
##          <chr>      <dbl>      <dbl>      <chr>
## 1 Faches-Thumesnil 50.58333  3.066667   Lille
## 2 Lille             50.63333  3.066667   Lille
## 3 Lezennes         50.61667  3.116667   Lille
## 4 Lille             50.63333  3.066667   Lille
## 5 Ronchin          50.60000  3.100000   Lille
## 6 Villeneuve-d'Ascq 50.68333  3.141667   Lille
```

```
# Transformation en objet spatial
```

```
communes <- SpatialPoints(coords=infoCom[,c(2,3)])
```

```
#Visualisation des slots disponibles
```

```
slotNames(communes)
```

```
##[1] "coords"      "bbox"         "proj4string"
```

```
# Connaitre l'étendue spatiale
```

```
communes@bbox # ou bbox(communes)
```

```
##          min      max
##latitude 50.000000 51.083333
##longitude 2.108333  4.183333
```

On peut aussi représenter graphiquement cet objet *via* l'instruction graphique classique `plot` (illustration en figure 1.9).

```
plot(communes)
```

Notre objet spatial peut également posséder une table d'attributs décrivant les objets géographiques qu'il contient. L'objet appartient alors à la classe des `SpatialPointsDataFrame` :

```
#Ajout de la table d'attributs
```

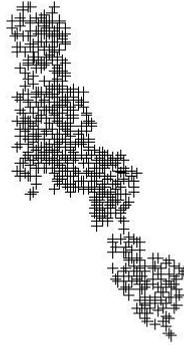


FIGURE 1.9 – Communes du Nord de la France

Source : Insee

```
nord<- SpatialPointsDataFrame(coords=infoCom[,c(2,3)],data =infoCom[,c
(1,4)])
```

On accède à cette table d'attributs *via* le nouveau slot créé @data :

```
nord@data
##          nom_commune préfecture
##          <chr>      <chr>
## 1  Faches-Thumesnil  Lille
## 2          Lille     Lille
## 3      Lezennes     Lille
## 4          Lille     Lille
## 5      Ronchin     Lille
## 6 Villeneuve-d'Ascq  Lille
## 7    La Madeleine   Lille
## 8          Lille     Lille
## 9      Comines     Lille
## 10      Deulemont   Lille
## # ... with 611 more rows
```

La création de **polygones géoréférencés**, bien qu'un peu plus complexe, suit la même logique.

En premier lieu nous allons créer des polygones simples à l'aide des coordonnées des sommets :

```
# Création des séries de coordonnées
x1 <- c(439518.5, 433091.8, 455774.1, 476566.1, 476944.2, 459554.4,
439518.5)
y1 <- c(8045280, 8031293, 8018439, 8026756, 8044902, 8054731, 8045280)
c1 <- data.frame(x1, y1)
x2 <- c(444929.2, 417667.9, 501837.1, 499792.5, 444929.2)
y2 <- c(8121306, 8078029, 8067465, 8109039, 8121306)
```

```

c2 <- data.frame(x2, y2)
x3 <- c(456530.1, 450481.5, 472785.8, 476566.1, 456530.1)
y3 <- c(8101608, 8089510, 8087620, 8099717, 8101608)
c3 <- data.frame(x3, y3)

```

```
# création des polygones
```

```

p1 <- Polygon(coords = c1, hole = F)
p2 <- Polygon(coords = c2, hole = F)
p3 <- Polygon(coords = c3, hole = T)

```

Le paramètre `hole` sert à identifier les polygones représentant des trous à l'intérieur d'autres polygones.

Ces objets possèdent 5 slots, dont :

- `@labpt` qui donne les coordonnées du centre;
- `@hole` qui dit s'il s'agit d'un trou;
- `@coords` qui permet de récupérer les coordonnées des sommets.

Ils peuvent ensuite être assemblés en polygones multiples :

```

P1 <- Polygons(srl = list(p1), ID = "PolygA")
P2 <- Polygons(srl = list(p2, p3), ID = "PolygB")

```

Ainsi le polygone *P1* sera composé de *p1* et *P2* sera *p2* avec un trou au centre défini par *p3*

Ils possèdent encore 5 slots différents, dont :

- `@Polygons` qui donne la liste des polygones ayant servi à sa création;
- `@ID` qui donne les identifiants donné au polygone.

On spatialise ensuite cet ensemble de polygones pour en faire un unique objet spatial :

```
SP <- SpatialPolygons(Sr1 = list(P1, P2))
```

Notre objet spatial se structure donc de la manière suivante : le `SpatialPolygons` contient une liste de deux polygones (polygones multiples) contenant chacun une liste de `Polygons` (polygones simples), lesquels contiennent les coordonnées qui les délimitent. Ainsi, pour accéder aux coordonnées du premier polygone simple contenu dans le second polygone multiple, nous devons écrire :

```
SP@polygons[[2]]@Polygons[[1]]@coords
```

```

##           x2           y2
## [1,] 444929 8121306
## [2,] 499793 8109039
## [3,] 501837 8067465
## [4,] 417668 8078029
## [5,] 444929 8121306

```

Pour ajouter une table d'attributs à notre objet géographique, il suffit de créer un dataframe contenant autant de lignes que de polygones multiples dans notre objet. Les lignes doivent être triées dans le même ordre que les polygones et chaque ligne identifiée par le même identifiant.

```

Info <- c("Simple", "Hole")
Value <- c(342, 123)
mat <- data.frame(Info, Value)
rownames(mat) <- c("PolygA", "PolygB")

```

```
SPDF<- SpatialPolygonsDataFrame(Sr = SP, data= mat )
```

Un nouveau slot est rajouté, @data, pour récupérer la table des attributs. On peut représenter graphiquement cet objet, ce qui donne la figure 1.10.

```
plot(SPDF,col =c("lightgrey","black"))
```

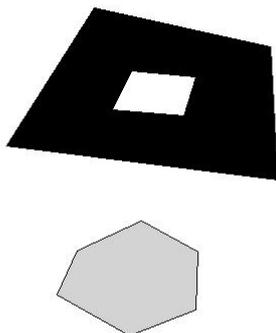


FIGURE 1.10 – Polygones créés

On peut construire des objets de type **lignes géoréférencées** de la même manière que celle présentée précédemment pour les polygones. Cette fois-ci, ce seront les fonctions `SpatialLines` et `SpatialLinesDataFrame` qui serviront.

Ainsi introduits, nos fonds de cartes communaux, départementaux, etc, seront des objets de type `SpatialPolygons(DataFrame)`, nos fonds routiers ou cours d'eau de type `SpatialLines(DataFrame)` et nos fonds d'aéroports ou des mairies de type `SpatialPoints(DataFrame)`.

Travail sur une couche vectorielle

La grande majorité du temps, nous ne créons pas d'objets géographiques de toute pièce mais manipulons des objets déjà existants. Plusieurs packages permettent d'importer ou exporter des objets géographiques. Le plus simple et complet demeure `rgdal` qui permet de lire et manipuler un très grand nombre de formats. Le format vectoriel le plus répandu est l'"ESRI ShapeFile", qui fournit un fond de carte à travers 5 fichiers devant être présents côte à côte dans un même dossier (.shp, .shx, .dbf, .prj, .cpg). Tous ces fichiers portent le même nom, seule l'extension diffère.

Pour importer le fond de carte, on utilise la fonction `readOGR` :

```
library(rgdal)
comr59<- readOGR(dsn = "Mes Documents\\fonds", layer = "comr59", verbose =
  FALSE)
```

Les paramètres de la fonction `readOGR` sont :

- `dsn` : chemin du dossier où sont stockés les fichiers ;
- `layer` : nom du fichier (sans extension).

On obtient alors un objet R de type `SpatialPolygonsDataFrame` (exemple en figure 1.11).

```
class(comr59)
## [1] "SpatialPolygonsDataFrame"
## attr(,"package")
## [1] "sp"
plot(comr59)
```

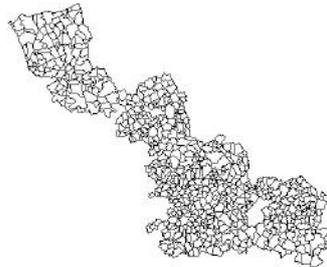


FIGURE 1.11 – Fond communal Nord

Source : *Insee*

`readOGR` permet d'importer une large palette de formats cartographiques. Pour un fond issu de MapInfo, la syntaxe change peu :

```
comr59MI<- readOGR(dsn= "Mes documents\\fonds\\comr59.tab", layer="comr59",
  verbose=FALSE)
```

Comme pour les ShapeFiles, le format MapInfo se compose d'un certain nombre de fichiers qui doivent être tous présents dans le même dossier, avec un nom identique mais des extensions différentes. Dans ce cas, le `dsn` pointe jusqu'au fichier `.tab`, et le `layer` prend le nom des fichiers du fond.

Pour sélectionner un sous-ensemble de notre carte, on se réfère au dataframe associée *via* le slot `@data`. Ainsi, pour sélectionner les communes de superficie supérieure à 200 km² :

```
comr59_etendue<- comr59[comr59@data$surf_m2>20000000, ]
```

Pour visualiser cette sélection, on superpose les 2 objets en colorant la sélection en gris (résultat visible en figure 1.12) :

```
plot(comr59)
plot(comr59_etendue,col ="darkgrey",add =TRUE)
```

Le paramètre `add=TRUE` permet de superposer les 2 fonds.

Pour sauvegarder notre nouveau fond cartographique, on utilise la fonction `writeOGR` qui prend comme paramètres :

- `obj` : objet R à exporter ;
- `dsn` : chemin du dossier de sauvegarde ;
- `layer` : nom commun des fichiers (sans extension) ;
- `driver` : format d'exportation de l'objet.

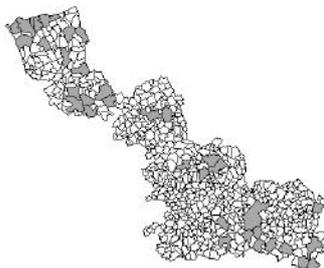


FIGURE 1.12 – Fond communal Nord étendu

Source : Insee

L'ensemble des formats possibles est fourni par la fonction `ogrDrivers()`.

Ainsi pour exporter notre sélection au format ShapeFile :

```
writeOGR(comr59_etendue, dsn="Mes documents\\fonds", layer="comr59_etendue",
driver="ESRI Shapefile")
```

Au format MapInfo :

```
writeOGR(comr59_etendue, dsn="Mes documents\\fonds\\comr59_etendue_MI.tab",
layer="comr59_etendue_MI", driver="MapInfo File")
```

1.3.2 Réalisation de cartes statistiques

Le système de projection

Les données spatiales sont toujours associées à un système de projection. Celui-ci est identifiable par le slot `@proj4string`.

```
comr59@proj4string
## CRS arguments:
## +proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0
```

On peut affecter un système de projection à un objet créé. Si on reprend nos polygones précédents, ils n'ont par défaut aucun système de projection associé. Pour signifier qu'ils sont en WGS84 :

```
SPDF@proj4string<- CRS( "+proj=longlat +datum=WGS84 +ellps=WGS84")
```

Le standard EPSG permet également d'identifier les système de projection par un code unique. Il est de 4326 pour le WGS84. Dans ce cadre, l'affectation précédente pourrait se coder :

```
SPDF@proj4string<- CRS( "+init=epsg:4326")
```

L'ensemble des correspondances des codes EPSG peut être obtenu en exécutant `make_EPSG()`. Ainsi pour le Lambert 93 (utilisé entre autres par l'IGN) le code EPSG est le 2154.

Si maintenant on souhaite reprojeter un objet géographique dans un nouveau système de coordonnées, on utilise la fonction `spTransform()`.

```
comr59_193<- spTransform(comr59, CRSobj=CRS("+init=epsg:2154"))
```

Cette reprojection est nécessaire notamment pour pouvoir superposer deux fonds qui ne possèdent pas le même système de coordonnées.

Pour réaliser très simplement des cartes, nous présentons le package *cartography* qui en plus de sa facilité de prise en main est relativement complet dans ses possibilités.

Cartes en symboles proportionnels :

La cartographie de données de stocks (comme une population, un nombre d'équipements...) se fait à l'aide de symboles proportionnels à la grandeur représentée. Le plus commun est le rond, mais on peut imaginer tout autre symbole. Le code ci-dessous permet d'obtenir la figure 1.13.

```
library(rgdal)
library(cartography)
metr_nice <- readOGR(dsn=~\\fonds",layer="metr_nice",verbose=F)

# Table des données de population
head(donnees_communes)
##   CODGEO          LIBGEO REG DEP P13_POP
## 1  01001  L'Abergement-Clémenciat  84  01    767
## 2  01002    L'Abergement-de-Varey  84  01    236
## 3  01004      Ambérieu-en-Bugey  84  01  14359
## 4  01005      Ambérieux-en-Dombes  84  01   1635
## 5  01006              Ambléon  84  01    108
## 6  01007              Ambronay  84  01   2503

#tracé du fond de carte
plot(metr_nice)

#ajout de l'analyse
propSymbolsLayer(spdf=metr_nice, df=donnees_communes, spdfid = "Codgeo",
                 dfid = "CODGEO", var= "P13_POP", col="salmon",
                 symbols="circle", legend.pos="right")

#habillage de la carte
layoutLayer(title = "Population de Nice métropole",
            author = "INSEE", sources = "Recensement 2013",
            scale = NULL, north = TRUE)
```

Les différents paramètres de la fonction sont :

- `spdf` : le `SpatialPolygonsDataFrame` ;
- `df` : le dataframe contenant les données à analyser ;
- `spdfid` : identifiant de la maille cartographique (dans le slot `@data`) ;
- `dfid` : identifiant de ligne dans le dataframe. Doit correspondre avec le précédent ;
- `var` : variable du dataframe à analyser.

D'autres paramètres existent et peuvent être listés dans l'aide de la fonction.

Cartes choroplèthes :

Pour la représentation des taux, on utilise des cartes en aplats de couleur ou choroplèthes. La variable est répartie dans des classes et un dégradé de couleurs schématise la croissance des valeurs (voir figure 1.14).

```
plot(metr_nice)
```

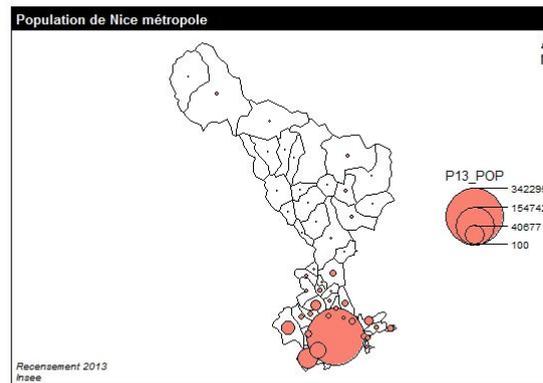


FIGURE 1.13 – Symboles proportionnels

Source : Insee, Recensement de la population 2013

```
choroLayer(spdf=metr_nice,df =donnees_communes4, spdfid = "Codgeo",
  dfid = "CODGEO",var = "TCHOM", nclass=4, method="fisher-jenks",
  legend.pos="right")
```

```
layoutLayer(title= "Taux de chômage des communes de Nice metropole",
  author = "Insee", sources = "Recensement 2013",
  scale = NULL, north = TRUE)
```

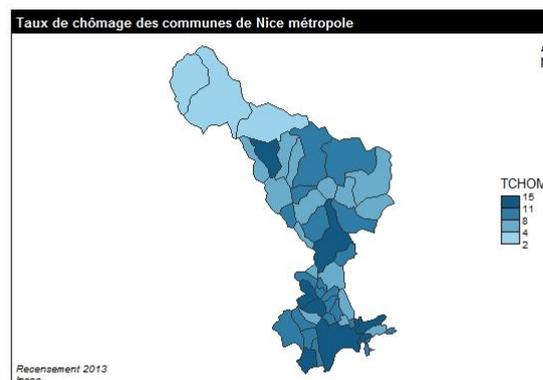


FIGURE 1.14 – Carte choroplèthe

Source : Insee, Recensement de la population 2013

La classification se réalise soit en spécifiant le nombre de classes (`nclass`) et la méthode de découpage (`method` qui permet de choisir parmi les méthodes présentées en section 1.2) ou en donnant un vecteur de bornes (`breaks`).

Autres fonctions cartographiques :

- *propSymbolsChoroLayer* : il s'agit d'un mélange entre les symboles proportionnels et les cartes choroplèthes (pour représenter simultanément un nombre de chômeurs et un taux de chômage par exemple) ;
- *typoLayer* : pour représenter une typologie en spécifiant une variable qualitative et un vecteur de couleur de même longueur que le nombre de modalités ;
- *gradLinkLayer* : pour représenter des flux ou des liens.

D'autres packages permettent de réaliser des cartes statistiques sous R. On peut citer entre autres :

- *RgoogleMaps* : réaliser des cartes en utilisant des rasters routiers ou satellites GoogleMaps ;
- *leaflet* : réaliser des cartes interactives avec raster OpenStreetMap pouvant être insérées dans des pages Web voire du RShiny.

1.3.3 sf : l'avenir du traitement des données spatiales sous R

Comme nous l'avons vu précédemment, jusqu'à présent le traitement des données cartographiques se faisait à travers trois packages principaux sous R :

- *sp* pour l'implémentation des classes de type spatiales ;
- *rgdal* pour les bibliothèques d'entrée/sortie ;
- *rgeos* pour les opérations sur les objets géométriques.

Depuis peu, il existe un package unique, nommé *sf*, qui regroupe l'ensemble des fonctionnalités de ces 3 packages réunis. Il fournit aux utilisateurs une classe unique pour la manipulation de l'ensemble des objets spatiaux. Dans ce chapitre, nous présentons rapidement les principales fonctionnalités du package *sf*. Pour aller plus loin dans la compréhension de ce package, la manipulation des géométries riches ou encore la gestion des entrées/sorties, nous invitons le lecteur à consulter les différentes vignettes mises à disposition avec le package sur le site du CRAN. En effet, ce package n'est pas encore compatible avec l'ensemble des packages d'analyse spatiale présentés dans ce manuel qui sont le plus souvent construits à partir de *sp*, *rgdal* et *rgeos*.

Il est notable que les entrées/sorties sont bien plus rapides avec *sf* qu'avec *rgdal*.

Les objets de classe *sf* se définissant comme des `data.frame` agrémentés d'attributs géométriques, la manipulation des objets géographiques en est simplifiée et se fait nativement comme une table quelconque sous R. Concrètement le package définit trois classes d'objets différents :

- *sf* : un `data.frame` avec des attributs spatiaux ;
- *sfc* : la colonne du `data.frame` stockant les données géométriques ;
- *sfg* : la géométrie de chaque enregistrement.

Ainsi un objet spatial sera représenté comme en figure 1.15.

L'importation de fond cartographique existant est simplifiée sous *sf*. Elle se fait ainsi :

```
library(sf)
depf <- st_read("J:/CARTES/METRO/An15/Shape/Depf_region.shp")
```

Il est à noter qu'il n'est pas nécessaire de spécifier le driver d'importation : `st_read` s'adapte automatiquement au format du fichier en entrée. La fonction est compatible avec la grande majorité des formats cartographiques courants (ESRI-Shapefile, MapInfo, PostGIS, etc.). On peut cartographier facilement les données spatiales avec la fonction `plot` (figures 1.16 et 1.17).

L'exportation de fond est tout aussi simple :

```
st_write(depf, "U:/fond_dep.shp")
```

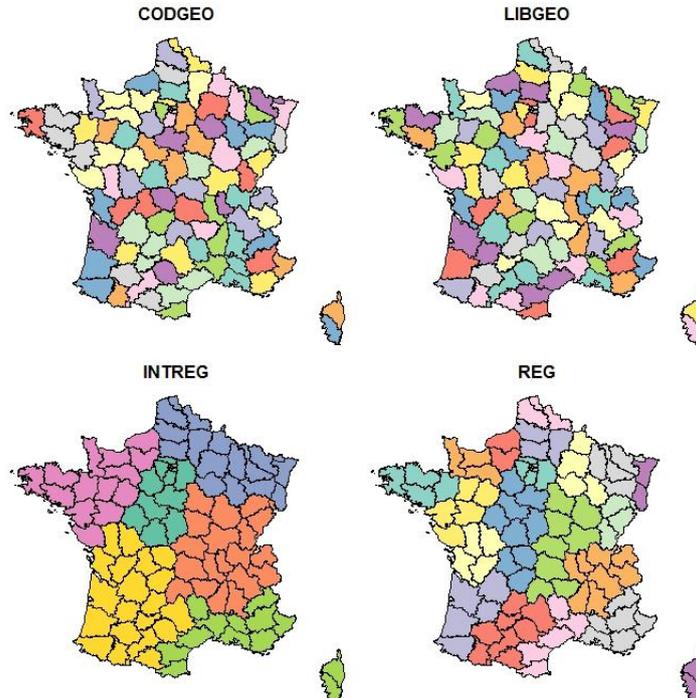
```

simple feature collection with 96 features and 4 fields
geometry type: MULTIPOLYGON
dimension: XYZ
bbox: xmin: 99225.97 ymin: 6049647 xmax: 1242375 ymax: 7110480
epsg (SRID): NA
proj4string: +proj=lcc +lat_1=44 +lat_2=49.000000000001 +lat_0=46.5 +lon_0=3
+x_0=700000 +y_0=6600000 +datum=NAD83 +units=m +no_defs
First 10 features:

```

	CODGEO	LIBGEO	INTREG	REG	geometry
1	01	Ain	ES	82	MULTIPOLYGON Z (((943513 65...
2	02	Aisne	NE	22	MULTIPOLYGON Z (((790281 69...
3	03	Allier	ES	83	MULTIPOLYGON Z (((777281 65...
4	04	Alpes-de-Haute-Provence	SE	93	MULTIPOLYGON Z (((1016633 6...
5	05	Hautes-Alpes	SE	93	MULTIPOLYGON Z (((1022838 6...
6	06	Alpes-Maritimes	SE	93	MULTIPOLYGON Z (((1077507 6...
7	07	Ardeche	ES	82	MULTIPOLYGON Z (((848816 64...
8	08	Ardennes	NE	21	MULTIPOLYGON Z (((873032.1 ...
9	09	Ariège	SO	73	MULTIPOLYGON Z (((632344 61...
10	10	Aube	NE	21	MULTIPOLYGON Z (((838365 67...

↑
sfc

FIGURE 1.15 – Représentation d'un objet spatial avec le package *sf*FIGURE 1.16 – Carte obtenue avec le code : `plot(depf)`

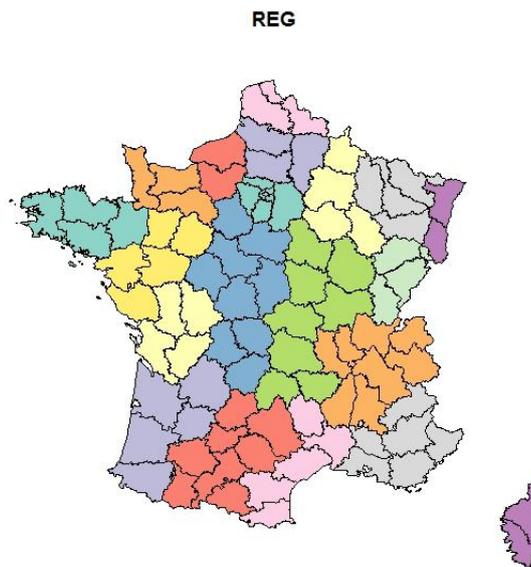


FIGURE 1.17 – Carte obtenue avec le code : `plot(depf["REG"])`

Plus généralement, le package *sf* propose un ensemble d'opérateurs de données spatiales tous préfixés par `st_` et présentés en figure 1.18 :

```
## [1] "st_agr<-.sf"      "st_agr.sf"        "st_as_sf.sf"
## [4] "st_bbox.sf"       "st_boundary.sf"   "st_buffer.sf"
## [7] "st_cast.sf"       "st_centroid.sf"   "st_convex_hull.sf"
## [10] "st_coordinates.sf" "st_crs<-.sf"     "st_crs.sf"
## [13] "st_difference.sf" "st_geometry<-.sf" "st_geometry.sf"
## [16] "st_intersection.sf" "st_is.sf"         "st_line_merge.sf"
## [19] "st_make_valid.sf" "st_polygonize.sf" "st_precision.sf"
## [22] "st_segmentize.sf" "st_set_precision.sf" "st_simplify.sf"
## [25] "st_split.sf"      "st_sym_difference.sf" "st_transform.sf"
## [28] "st_triangulate.sf" "st_union.sf"      "st_voronoi.sf"
## [31] "st_zm.sf"
```

FIGURE 1.18 – Ensemble des opérateurs présents dans le package *sf*

Le package *sf* est également totalement intégré dans l'environnement *tidyverse* et ainsi se conjugue parfaitement avec les fonctionnalités du package *dplyr* :

```
library(dplyr)
depf <- left_join(dep, pop_dep, by = "CODGEO")
```

ou encore le code ci-dessous, illustré dans la figure 1.19.

```
library(dplyr)
depf %>%
  mutate(
```

```

  area = st_area(.), # on crée la nouvelle variable sur la superficie
) %>%
group_by(REG) %>%
summarise(mean_area = mean(area)) %>%
plot

```

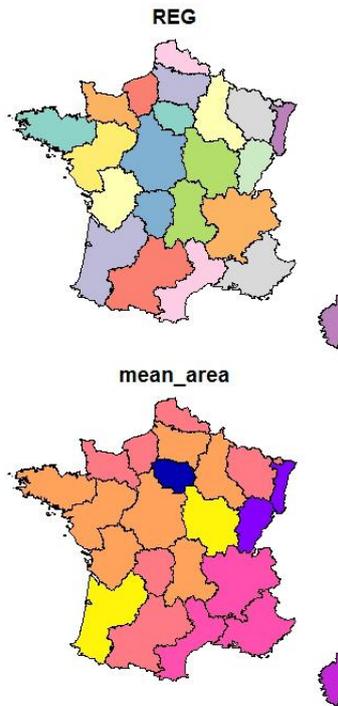


FIGURE 1.19 – Carte réalisée grâce aux packages *sf* et *dplyr*

La plupart des packages liés au traitement de données géographiques se sont adaptés à cette nouvelle classe d'objets. Certains, comme *spdep*, sont en phase de test de leur adaptation et nécessitent ainsi encore le recours au package *sp*.

Concernant *cartography*, l'adaptation est effective depuis sa version 2.0 et la syntaxe a évolué :

```
choroLayer(x, spdf, spdfid, df, dfid, var, ...)
```

où *x* est un objet de type *sf*. S'il est renseigné, les objets *spdf*, *spdfid*, *df* et *dfid* sont ignorés car l'ensemble des ces renseignements sont inclus dans l'objet *x*.

1.3.4 De la surface au point, et réciproquement

Une particularité des données surfaciques est qu'elles peuvent consister en une partition de l'ensemble du territoire ou en un ensemble de points de référence aux coordonnées géographiques distinctes. On peut cependant passer facilement d'une représentation à l'autre :

- les polygones de Voronoï permettent de créer une partition du territoire à partir de points de référence ;
- utiliser le centroïde d'une aire permet de passer d'une partition du territoire à un ensemble de points.

Définition 1.3.1 — Polygone de Voronoï associé au point x_i . Il s'agit de la région de l'espace qui est plus proche de x_i que de tout autre point de l'ensemble d'étude \mathbf{x} :

$$C(x_i|\mathbf{x}) = \left\{ u \in \mathbb{R}^2 : \|u - x_i\| = \min_j \|u - x_j\| \right\} \quad (1.1)$$

Encadré 1.3.1 — Des polygones très utilisés. Les polygones de Voronoï font partie des structures géométriques les plus utilisées par la communauté scientifique. D'après AURENHAMMER 1991, trois grandes raisons expliquent cet intérêt. La première est que les polygones de Voronoï sont directement observables dans la nature (dans les arrangements cristallins, par exemple). Deuxièmement, ils sont l'une des constructions les plus fondamentales définies par un ensemble discret de points : ils présentent de très nombreuses propriétés mathématiques, et sont reliés à plusieurs autres structures géométriques fondamentales. Enfin, les polygones de Voronoï permettent de simplifier un grand nombre de problèmes algorithmiques. Le polygone de Voronoï associé à un point est souvent considéré comme sa "zone d'influence".

Historiquement, Gauss (en 1840) puis Dirichlet (en 1850) utilisèrent les polygones de Voronoï dans leur étude des formes quadratiques. Voronoï généralisa leur travail à des dimensions supérieures en 1908. Quelques années plus tard, en 1934, Delaunay construisit une triangulation associée aux polygones de Voronoï et démontra la richesse de ses propriétés mathématiques.

Le package R *deldir* permet de calculer les polygones de Voronoï associés à un ensemble de points. La fonction *deldir* renvoie en sortie un objet représentable avec la fonction *plot*. Le package calcule également plusieurs statistiques associées aux polygones, telles que la surface de chaque polygone, ou encore le nombre de ses sommets (voir documentation détaillée :).

Il existe de nombreux algorithmes permettant de construire des polygones de Voronoï (le plus efficace est l'algorithme de Fortune (FORTUNE 1987)). L'algorithme implémenté par la fonction *deldir* commence par construire une triangulation de Delaunay à partir des points de référence. Cette triangulation maximise l'angle minimal des triangles. Les sommets du diagramme de Voronoï sont les centres des cercles circonscrits des triangles de la triangulation de Delaunay. Les arêtes du diagramme de Voronoï sont sur les médiatrices des arêtes de la triangulation de Delaunay (l'algorithme est détaillé dans LEE et al. 1980).

Application avec R

```
#Packages nécessaires
library(deldir)
library(sp)

# Génération des points aléatoires
x <- rnorm(20, 0, 1.5)
y <- rnorm(20, 0, 1)

#Fonction "deldir" permettant de calculer les polygones de Voronoï
#à partir de deux jeux de coordonnées géographiques
vtess <- deldir(x, y)

#crée une fenêtre de travail
plot(x, y, type="n", asp=1)

#représente les points
```

```
points(x, y, pch=20,col ="red", cex=1)

#représente les polygones de Voronoi associés
plot(vtess, wlines="tess", wpoints="none", number=FALSE, add=TRUE, lty=1)
```

Pour passer d'une partition du territoire à un ensemble de points, on peut calculer les centroïdes des surfaces (figure 1.20).

Définition 1.3.2 — Centroïde d'une surface S. Point qui minimise la distance quadratique moyenne à tous les points de S :

$$\min_c \frac{1}{a(S)} \int_S \|x - c\|^2 dx$$

$$c = \frac{1}{a(S)} \int_S x dx$$

Coordonnées de c : moyenne des coordonnées de **tous les points de S**

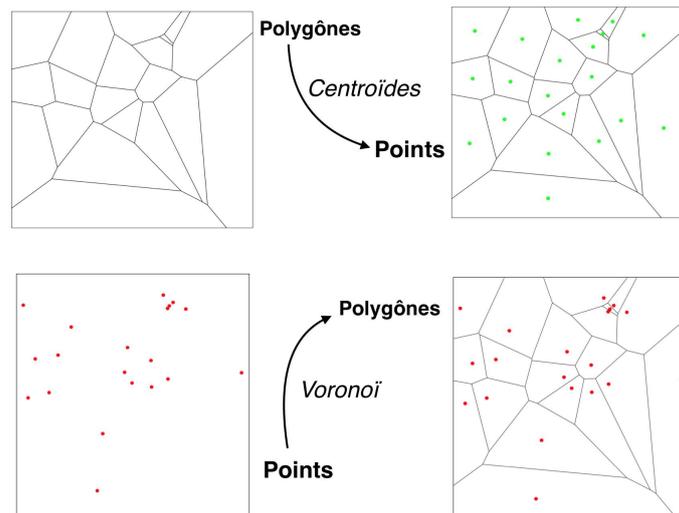


FIGURE 1.20 – Passage des points aux polygones et des polygones aux centroïdes

Application avec R

```
#Calcul des centroïdes des polygones
#A partir d'un fichier "Spatial Polygon Data Frame"

library(GISTools)
centroids<- getSpPPolygonsLabptSlots(polygone)

plot(polygone)
points(centroids, pch = 20,col = "Green", cex=0.5)
```

1.4 Exemples d'études utilisant des données spatiales agrégées

Le groupe européen pour l'intégration des données³ souligne que représenter les données sur une carte avec une bonne résolution spatiale et temporelle permet de détecter des phénomènes invisibles autrement. Une représentation adaptée permet de bien comprendre la situation économique, sociale ou environnementale et de mettre en place des politiques publiques pertinentes. À travers les travaux réalisés par trois instituts de statistique européen, cette section illustre la variété des analyses descriptives utilisant des données spatiales : projet européen d'étude des taux de pauvreté régionaux ; analyse de la distance aux espaces verts par l'institut de statistique suédois ; analyse de la localisation optimale des éoliennes par l'institut de cartographie britannique.

1.4.1 Accès aux espaces verts - Statistique Suède

Augmenter l'accès aux espaces verts publics fait partie des objectifs environnementaux des politiques publiques suédoises. Dans de nombreuses municipalités suédoises, des débats opposent les partisans de la densification de l'habitat et ceux de la préservation des espaces verts.

L'association de données cartographiques satellitaires et d'informations statistiques localisées issues du recensement aide à mieux comprendre la situation sur le terrain, et ainsi à ajuster les politiques publiques. Cette étude rentre dans le cadre du 11ème objectif de développement durable des Nations Unies : "rendre les villes et les installations humaines inclusives, sûres, résistantes et durables".

En 2013, l'Institut de Statistique Suédois s'est appuyé sur l'analyse conjointe d'images satellites et de données administratives pour caractériser les espaces verts suédois, en fonction de leur statut de propriété, et de la qualité de leur végétation. La majorité des aires urbaines suédoises voient plus de 50 % de leur territoire recouvert par des espaces verts. En moyenne, trois quarts de ces espaces sont publics. Lidingö est la ville suédoise la plus recouverte d'espaces verts, puisque ceux-ci représentent environ 72 % de sa surface totale (figure 1.21). La deuxième partie de l'étude se concentre sur l'accessibilité à ces espaces verts. Grâce aux données du recensement de la population, l'Institut Suédois étudie la proportion d'adultes et d'enfants vivant à moins d'une certaine distance d'un espace vert public. Ainsi, **dans 26 aires urbaines suédoises, moins d'1 % de la population vit à plus de 300 mètres d'un espace vert accessible**. En revanche, dans certaines villes, comme Malmö, 15 % des enfants de moins de 6 ans n'ont pas accès à un espace vert à moins de 200 mètres.

1.4.2 Taux de pauvreté régionaux - programme européen ESPON

Le projet européen ESPON a pour objectif de favoriser l'harmonisation des politiques publiques européennes en mettant à disposition des décideurs des statistiques régionales pertinentes. Les différences de richesse entre les régions peuvent exacerber les sentiments d'exclusion et les tensions au niveau national. Cartographier le taux régional de pauvreté de la population permet de distinguer les zones les plus fragiles, et ainsi de mieux cibler les politiques d'aide au développement.

Le seuil de pauvreté (*At-Risk-of-Poverty (ARoP) threshold*) est défini comme 60 % de la médiane du niveau de vie national. Le seuil de pauvreté varie donc selon les pays (de 20 362 euros en Suisse à 5 520 en Grèce). Le taux de pauvreté (*ARoP rate*) est défini comme la part d'individus dont le niveau de vie est inférieur au seuil de pauvreté national. Dans la figure 1.22, on représente le ratio entre cet indicateur calculé au niveau infra-national (NUTS3) et le taux de pauvreté national. Cela permet **d'identifier les pays présentant les disparités régionales les plus importantes**, et de **visualiser au sein de chaque pays les zones les plus extrêmes**. Les plus

3. UN-GGIM : United Nations Committee of Experts on Global Geospatial Information management - Working Group B - Europe

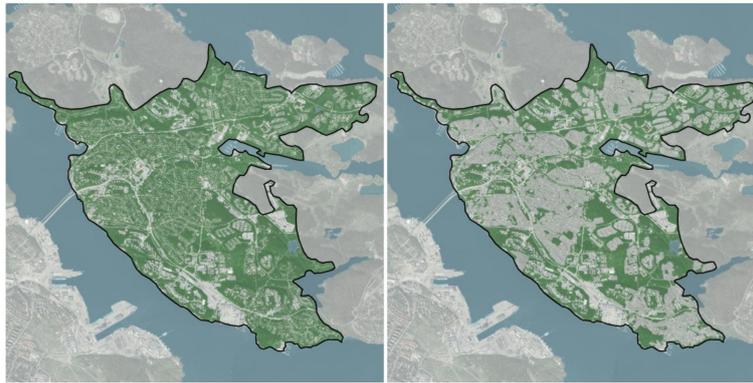


FIGURE 1.21 – Ville de Lidingö. Gauche : tous les espaces verts ; Droite : espaces verts accessibles au public (non privés)

Source : *Institut National de Statistiques Suédois*

grandes disparités inter-régionales dans les populations à risque sont observées en Turquie, Albanie, Hongrie, Allemagne, Croatie, Italie et Espagne. Les pays scandinaves, ainsi que les Pays-Bas, les États baltes, le Portugal et la Grèce ont une distribution plus uniforme des *ARoP rates*. Au sein des pays, on observe de faibles taux de pauvreté dans les alentours des capitales et des grandes villes, mais pas forcément au sein des villes elles-mêmes. Le taux de pauvreté est plus élevé dans les régions les moins accessibles, comme le sud de l'Italie, le centre de l'Espagne ou l'est de la Hongrie.

La représentation cartographique des taux de pauvreté ainsi définis permet d'aider la prise de décision publique, aussi bien au niveau national qu'europpéen. Dans ce but, le programme ESPON a mis en ligne de nombreuses analyses cartographiques de données démographiques et sociales : carte des ratios hommes-femmes par région, des différents profils d'innovation, des variations de taux d'emploi ou encore d'impact potentiel du changement climatique (<https://www.espon.eu/tools-maps>).

1.4.3 Localisation optimale des éoliennes - Institut de cartographie de Grande Bretagne

Le gouvernement écossais a pour objectif d'augmenter la production d'énergie renouvelable d'ici 2020. Le conseil régional joue un rôle clé dans la préservation de l'équilibre local, en cherchant à développer le parc éolien tout en préservant la qualité de vie des habitants. Les données spatiales fournies par l'Institut de cartographie britannique (*Ordnance Survey*) ont beaucoup aidé à la prise de décision au niveau local.

Les objectifs de l'étude sont de donner des lignes directrices claires et pratiques pour la localisation des champs d'éoliennes. L'étude doit prendre en compte de nombreux éléments environnementaux et sociaux comme les caractéristiques des paysages et le pittoresque des vues. Les données cartographiques doivent être suffisamment détaillées pour pouvoir servir de support aux planificateurs locaux, tout en restant suffisamment lisibles pour être comprises rapidement par l'ensemble des parties prenantes (figure 1.23).

Pour atteindre ces objectifs, l'*Ordnance Survey* a travaillé avec de nombreux experts locaux et utilisé de nombreuses bases géolocalisées. Les différents acteurs pouvaient suivre l'avancement de l'étude grâce à une cartographie interactive. Kevin Belton, ingénieur SIG et membre du conseil général régional souligne la valeur ajoutée de l'étude : "La communication d'informations complexes grâce à des données spatiales a permis au conseil régional de travailler avec un grand nombre de partenaires : aussi bien des membres du secteur public que des promoteurs commerciaux. Cette étude a

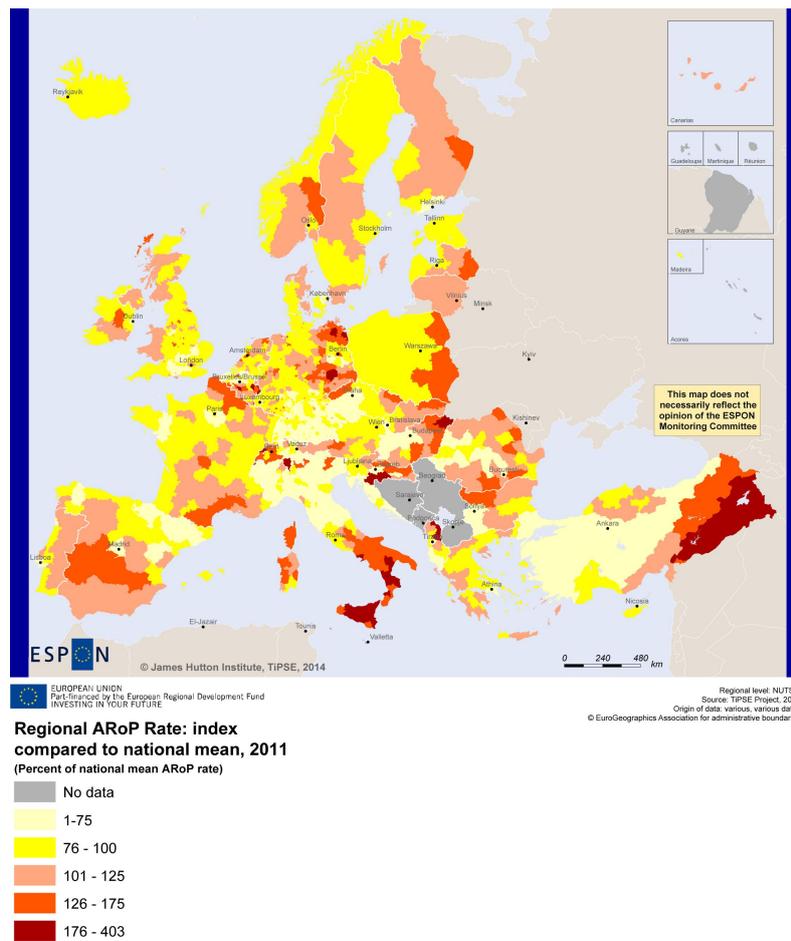


FIGURE 1.22 – Indicateur d'un risque élevé de pauvreté

Source : *Projet ESPON*

Note : pourcentage de l'indice régional rapporté à l'indice national

aidé les constructeurs à ne pas gaspiller de moyens dans des installations qui auraient été contraires aux politiques de préservation de l'environnement et de la qualité de vie des communautés locales."

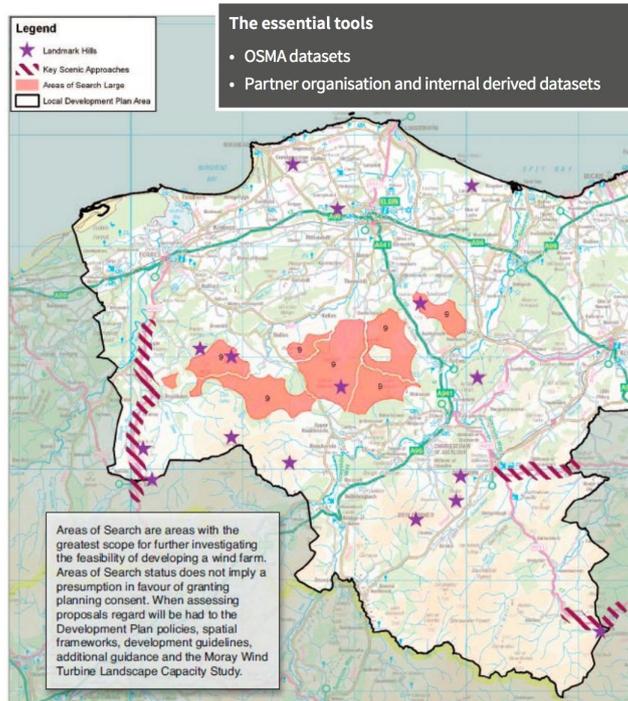


FIGURE 1.23 – Étude d'implantation d'éoliennes

Source : *Office National de Statistiques Britannique*

Note : Les aires de recherche "areas of search" sont les zones où il est le plus prometteur de lancer des investigations plus approfondies sur la possibilité d'y installer des éoliennes. Le statut d'"area of search" ne garantit pas l'octroi par les responsables de l'urbanisme local d'une autorisation de construction.

Références - Chapitre 1

- AURENHAMMER, Franz (1991). « Voronoi diagrams : a survey of a fundamental geometric data structure ». *ACM Computing Surveys (CSUR)* 23.3, p. 345–405.
- BIVAND, Roger S., Edzer PEBESMA et Virgilio GOMEZ-RUBIO (2008). *Applied spatial data analysis with R*. Springer.
- CHILES, Jean-Paul et al. (2005). *Les pratiques de la géostatistique dans le domaine des sites et sols pollués*. GeoSiPol.
- CRESSIE, Noel A.C. (1993b). « Statistics for spatial data : Wiley series in probability and statistics ». *Wiley-Interscience, New York* 15, p. 105–209.
- FORTUNE, Steven (1987). « A sweepline algorithm for Voronoi diagrams ». *Algorithmica* 2.1-4, p. 153.
- FOTHERINGHAM, A. Stewart et F. Benjamin ZHAN (1996). « A comparison of three exploratory methods for cluster detection in spatial point patterns ». *Geographical analysis* 28.3, p. 200–218.
- GIVORD, Pauline et al. (2016). « Quels outils pour mesurer la ségrégation dans le système éducatif ? Une application à la composition sociale des collèges français ». *Education et formation*.
- LEE, Der-Tsai et Bruce J SCHACHTER (1980). « Two algorithms for constructing a Delaunay triangulation ». *International Journal of Computer & Information Sciences* 9.3, p. 219–242.