**Création des cartes**

Comme énoncé lors de l’introduction, ce projet vise à créer une application de visualisation de données sur la position géographique des banques en fonction de données socio-économiques à l’échelle des zones d’emploi en France métropolitaine.

Pour cela, nous avons besoin de tracer des cartes, et pour tracer des cartes, l’un des moyens les plus utilisés est d’utiliser des *shapefiles.* Un shapefile (ou « fichier de forme ») est une combinaison de fichiers permettant de stocker les informations relatives aux tracés des cartes telle que l’emplacement, la forme ou les attributs des entités géographiques. Étant donné que nous voulons étudier l’emplacement géographique des banques au niveau des zones d’emplois, nous avons récupéré sur le site de l’**Insee** le shapefile relatif aux tracés des zones d’emploi pour l’année 2020.

Une fois cet ensemble de fichiers récupéré, nous avons dû le formater afin de récupérer seulement les données dont nous avions besoin. En effet, le découpage en zone d’emploi se fait en France métropolitaine et dans les DROM-COM. Comme nous considérons seulement la France métropolitaine, nous avons enlevé du jeu de données toutes les données faisant référence à l’outre-mer.

Après cela, nous avons pu nous atteler au tracé de nos premières cartes. Pour ce faire, nous avons utilisé le package ***sf***de ***R***. Ce package combine les fonctionnalités des packages ***sp***, ***rgdal***,***rgeos*** permettant d’importer, de manipuler et de transformer les données spatiales. Il propose des objets plus simples dont la manipulation est plus aisée. L’une de ses grandes forces est qu’il est compatible avec tous les opérateurs du ***tidyverse****.*

Pour mieux comprendre, comme vous pouvez vous référer à la Figure X, les objets de classe ***sf*** sont des data frames dont la dernière colonne est composée de *géométries*. Cette colonne est de classe *sfc* (*simple feature column*) et chaque la ligne de la colonne est de classe *sfg* (*simple feature geometry*). Plus généralement, cela correspond à des listes de coordonnées (*longitude / latitude)*. Ce format est très pratique car il permet de lier dans un même objet les données ainsi que les coordonnées géographiques associées (*géométries)*.

Une image contenant texte

Description générée automatiquement

Figure X : Représentation de l’objet de classe***sf*** faisant référence aux zones d’emplois 2020

Dans le shapefile que nous avons récupéré, nous voyons sur la Figure X que nous avons six colonnes.

Nous y retrouvons :

* Le codes des communes
* Le nom des communes,
* Le code de la zone d’emploi associée,
* Le nom de la zone d’emploi,
* La partie régionale de la zone d’emploi trans-régionale
* Les *géométries*,correspondant au tracé des frontières de chaque commune

Ainsi lorsque nous traçons la carte en fonction de ces données, vous pouvez constater sur la Figure X1 que nous avons le tracé des frontières de chaque commune. Toutefois, nous pouvons colorer la carte par zone d’emploi.

Notre objectif est donc le suivant : nous devons avoir une carte où les frontières seraient le contour de chaque bloc de couleur que nous pouvons voir sur la carte ci-dessous.

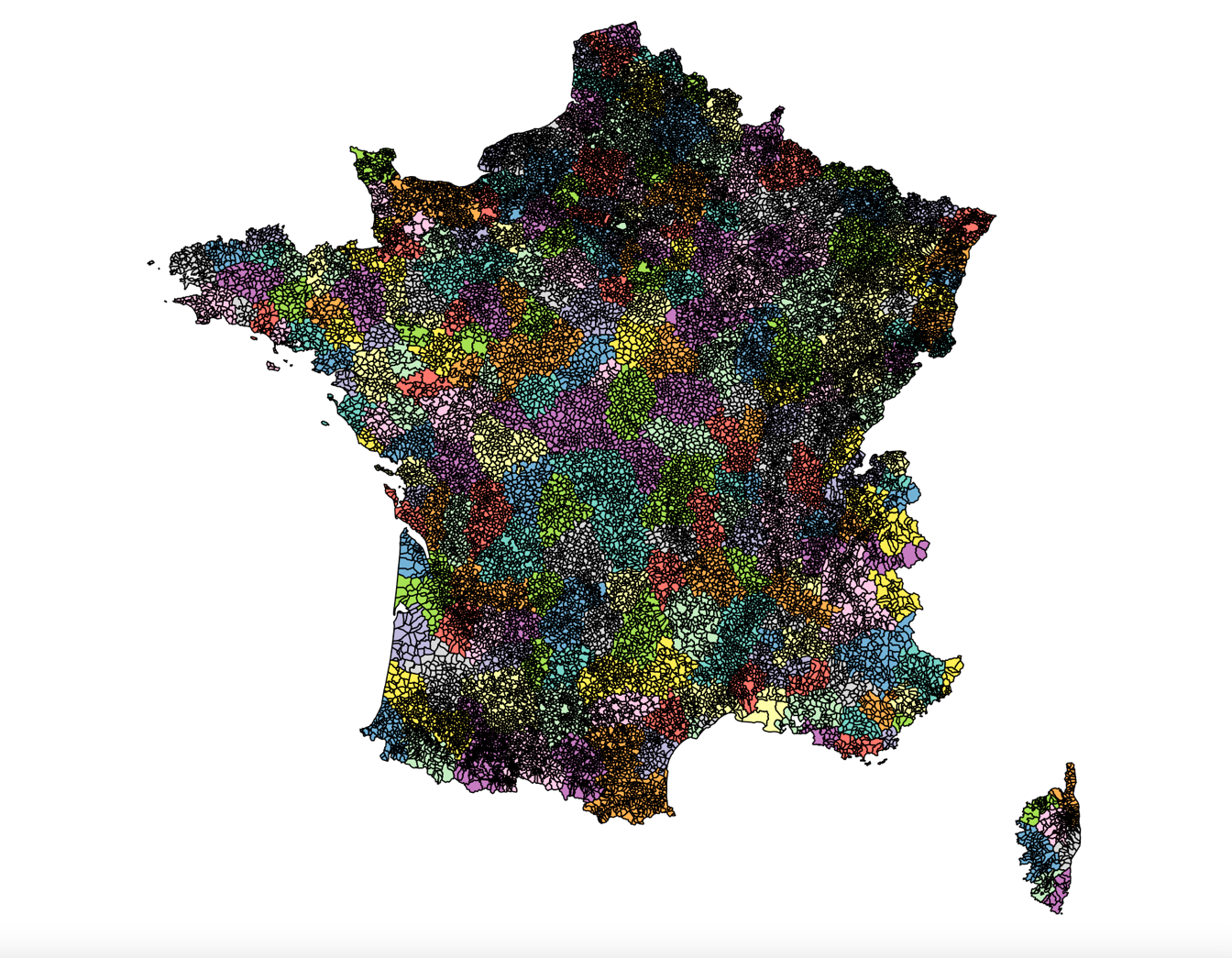


Figure X1 : Tracé des communes colorées par zone d’emplois

Nous avons donc dû retravailler sur ce fichier de données de sorte à obtenir seulement une data frame composée de deux colonnes. La première composée de la liste des codes de zones d’emplois et la deuxième composée de *géométries* répertoriant les coordonnées géographiques des contours de chaque zone d’emploi.

Ce qu’il faut comprendre ici, est que le logiciel lit les *géométries* comme des polygones. C’est-à-dire que chaque commune est considérée comme un polygone. Ainsi, au sein de chaque bloc de couleur de la carte, nous devons combiner tous les polygones de sorte à n’en obtenir qu’un seul qui serait celui de la zone d’emploi associée.

Comme expliqué un peu plus haut, les objets ***sf*** sont compatibles avec les opérateurs du ***tidyverse***. Nous avons donc utilisé des fonctions de ***dplyr***, un package de base du ***tidyverse***. Ce package permet d’effectuer de nombreuses manipulations sur les data frames. Nous avons tout d’abord utilisé la fonction *group\_by*(), qui nous a permis de regrouper toutes les communes ayant la même zone d’emplois. Et nous avons fini par appliquer sur ce regroupement de variables la fonction *summarize*() qui permet de créer une nouvelle data frame où chaque ligne sera composée du numéro de la zone d’emplois et les *géométries* associées.

Après ces opérations réalisées, nous avons tracé la carte présente dans la Figure X2.

Une image contenant texte, carte

Description générée automatiquement

Figure X2 : Tracé des zones d’emplois

(Nous constatons cette fois-ci que le tracé des zones d’emplois de la Figure X2 correspond aux blocs de couleurs que nous pouvons voir dans la Figure X1)

