Université Grenoble Alpes

RAPPORT

Implantation des banques coopératives en France métropolitaine

18 octobre 2021 - 2 mai 2022

Bastiste Amistadi, Cheikh Beye, Charline Champ, Antoine Grancher et Alicia Lorandi
Étudiants en Master Statistique et Sciences des données

Enseignants référents
Jean-François Coeurjolly et Amélie Artis

School year
2021–2022

SOMMAIRE

Remerciements

Introduction

Les banques sont aujourd'hui un élément central de l'économie. Les activités bancaires participent au développement économique des sociétés. Nous pouvons ainsi nous demander si ce développement économique, en fonction des ménages et du secteur d'activité des entreprises, a influencé la position géographique des banques.
En France, nous distinguons des grands types de banques : les banques lucratives et les banques coopératives.
Si une banque est lucrative, alors les décisions sont prises par des actionnaires sur le principe « une personne, une voix, une action ».
En revanche, si une banque est coopérative, alors il n'y a aucun actionnaire, les décisions sont prises par les clients. Ces clients ont le rôle de sociétaire et possèdent des parts sociales représentants le capital social de la banque. Ce statut leur permet de participer aux prises de décisions lors des assemblées générales.

C'est pourquoi, dans le cadre de la première année du master Statistique et Sciences de Données de l'Université Grenoble Alpes (UGA), nous avons réalisé un projet concernant la visualisation de données spatiales pour tenter d'expliquer la position géographique des banques en France métropolitaine. Ce projet visera donc à proposer une application de visualisation de données qui nous permettra de mettre en relation la position des banques en fonction de données socio-économiques à l'échelle des zones d'emploi.

Ce rapport sera ainsi divisé en cinq parties. Tout d'abord nous verrons comment nous nous sommes organisés pour mener à bien ce projet. Nous vous présenterons ensuite les données avec lesquelles nous avons construit notre outil de visualisation. Pour continuer, nous verrons comment nous l'avons réalisé. Nous vous proposerons également une petite analyse qui viendra appuyer nos visualisations, et pour finir nous ferons une rétrospective sur le travail réalisé.

Gestion de projet

Cahier des charges

Contexte et définition du problème :
Ce projet est né dans le cadre d'un projet tutoré de la première année de master Statistique et Sciences des données de l'Université Grenoble Alpes.
Le but du projet est de visualiser des données spatiales et spatio-temporelles associées à des banques lucratives et coopératives en France métropolitaine.

Objectifs du projet :
Le projet a pour but de visualiser des données spatiales pour tenter d'expliquer la position géographique de banques coopératives en France. Il faudra donc récolter, formater ces données de localisations et de les mettre en relation avec des covariances spatiales, obtenues à l'échelle des bassins d'emploi. Enfin, afin de pouvoir visualiser ces données, nous proposerons une application ***R-Shiny*** permettant de mettre en relation une ou plusieurs covariables, ainsi que la position géographique des banques.

Périmètre :
Libre.

Description fonctionnelle :
En premier lieu, il nous faudra récupérer toutes les données, c'est-à-dire les données socio-économiques et les données relatives aux banques.
Concernant les données socio-économiques, nous avons en notre disposition un fichier les répertoriant. Cependant, ces données datent de 2014. Nous tenterons de les réactualiser en les récupérant sur le site de l'Insee.
En ce qui concerne les banques, elles seront récupérées sur leur site officiel. Une fois ces données en notre possession, il faudra y ajouter les longitudes et latitudes de chaque banque. Cela donnera lieu à une unique base de données répertoriant toutes les banques, leur type (coopérative/lucrative), leur adresse, leur longitude et leur latitude.

En second lieu, nos codes s'occuperont du traitement des données et de la création d'une interface graphique. L'interface graphique sur laquelle l'utilisateur filtrera suivant différents paramètres : types de banque, banques, critères socio-économiques et zone d'emploi.
Les éléments présents sur cette interface seront des cartes de la France métropolitaine colorées en fonction d'un critère socio-économique pour chaque zone d'emploi avec la possibilité de superposer la position géographique des banques.

Délai :
Environ 7 mois (du 18/10/2021 au 02/05/2022).

Livrable :
Le code source, une interface ***R-Shiny*** affichant les cartes ainsi qu'un rapport.

Répartition des tâches

Discutons maintenant de la répartition des tâches et du travail au sein de notre groupe. Dès le début, nous avions des objectifs à atteindre comme vous avez pu le voir dans la partie précédente. Au fur et à mesure de l'avancement du projet, nous avons pu fixer de nouveaux objectifs en fonction du temps restant mis à notre disposition. Lors des premiers jours, nous avions planifié notre travail grâce à un diagramme de Gantt. Cependant nous avions sous-estimé certaines tâches, nous n'avions pas prévu certaines difficultés. Tout cela a rendu ce diagramme très vite obsolète. Nous avons donc fonctionné autrement : à chaque fin de session de travail, un bilan était réalisé afin de faire le point sur l'avancement du projet et ainsi nous permettre de se fixer les nouvelles tâches à réaliser lors de la session suivante. Nous nous répartissions ensuite ces tâches en essayant de garder la répartition initialement prévue. Comme notre diagramme de Gantt reflète peu la réalité de ce qu'il s'est passé, vous en trouverez un ci-dessous représentant les différentes tâches au moment où elles ont réellement été effectuées.

Tâches	Semaine du 18/10 au 06/12	Semaine du 13/12	Semaine du 20/12 au 07/03	Semaine du 14/02	Semaine du 28/03	Semaine du 04/04	Semaine du 11/04	Semaine du 25/04	Semaine du 02/05
Données									
Récupération des banques									
Récupération des données socio-économique									
Visualisation									
Création des cartes									
Ajout des zones d'emplois sur la carte									
Ajout des banques sur la carte									
Ajout des banques et des données socio-économique sur la carte									
Amélioration visuel des cartes									
Création de l'interface R-Shiny									
Création d'un package R									
Analyse									
Création du data frame avec le nombre de banques par zone d'emploi									
Modèle statistiques									
Présentation									
Réalisation présentation 1									
Soutenance présentation 1									
Rédaction du rapport									
Réalisation présentation 2									
Soutenance présentation 2									

Répartition	Couleur
Tous ensemble	
Antoine	
Batiste	
Charline	
Alicia & Batiste & Cheikh	
Antoine & Charline	

Figure X : Diagramme de Gantt

Notre travail s'est divisé en cinq morceaux : la récupération de données, le traitement de données, l'interface ***R-Shiny***, l'analyse et la préparation du rendu.
Notre répartition a globalement a été la suivante :

- Tout le groupe s'est occupé de la récupération de données
- Antoine et Batiste ont réalisé le traitement des données
- Alicia, Batiste et Cheikh ont géré l'interface ***R-Shiny***
- Antoine et Charline ont fait une analyse
- Et nous nous sommes tous occupés de la préparation du rapport

Présentation des données

Banques

Comme expliqué dans l'introduction, notre projet se base sur deux types de banques : les lucratives et les coopératives. En effet, rappelons que le but de ce projet est de visualiser des données spatiales pour tenter d'expliquer la position géographique de banques en France métropolitaine. De ce fait mettre en comparaisons les deux types de banques était le plus cohérent. Pour rentrer plus en détails, nous allons nous concentrer sur cinq banques. Parmi celles-ci, nous en avons trois coopératives :

- La Banque Populaire,
- Le Crédit Agricole,
- Le Crédit Mutuel.

Ainsi que deux lucratives :

- La BNP Paribas
- La Société Générale

À présent, un bref résumé de chaque banque va vous être présenté.

Banque Populaire

La Banque Populaire est un groupe mutualiste français de services bancaires et financiers. Il s'agit d'une société anonyme à directoire et conseil de surveillance dont le siège social est situé à Paris, avenue Pierre Mendès. C'est le deuxième groupe bancai re en France.

C'est il y a 100 ans, le 13 mars 1917, que la loi Clémentel donnait officiellement naissance aux Banques Populaires. En imposant une vision économique audacieuse, basée sur la coopération et la solidarité, elle a permis aux artisans, commer çants et petits industriels d'accéder au crédit bancaire.

De nombreux changements sont intervenus au sein de la Banque Populaire, notamment en 2009, avec la fin de la Banque Populaire. En effet, après la crise de 2008, la Banque Populaire et la Caisse d'Épargne ont fusionné pour former le groupe BPCE, qui est le centre d'intérêt des deux groupes. BPCE est notamment chargé d'assurer la représentation des affiliées auprès des autorités de tutelle, d'organiser la garantie des déposants, d'agréer les dirigeants et de veiller au bon fonctionnement des établissements du Groupe.

En 2018, l'organisation s'est implantée dans plusieurs régions pour se rapprocher des clients membres et conserver une certaine indépendance. Le groupe compte 12 Banques populaires régionales, CASDEN et Crédit Coopératif, et environ 2548 agences en France, au service de 8 900 000 clients, dont 3 900 000 adhérents.

Crédit Agricole

Le Crédit agricole est né d'un mouvement de solidarité à la fin du XIXème siècle. En effet, le système bancaire français répondant alors très mal aux besoins de crédit de l'agriculture, les agriculteurs s'organisent et deviennent leurs propres banquiers. Ils rassemblent ainsi les fonds rendus nécessaires pour financer le progrès technique. En 1885, dans le Jura, naît la première association coopérative de ce type, dont le siège est à Salins.

En 1920, est créé l'organe central : l'office national du Crédit Agricole qui deviendra en 1926 la Caisse Nationale du Crédit Agricole (CNCA). Puis en 1967, celle-ci obtient son autonomie financière vis à vis de l'état. Il n'est plus un organisme gestionnaire de subventions gouvernementales et doit désormais constituer lui-même sa réserve de liquidité pour couvrir ses opérations de crédit à moyen et long terme.

Le 18 janvier 1988, la loi de mutualisation a transformé la Caisse Nationale, établissement public, en Société Anonyme (SA) de droit privé au capital de 0,65 milliards d'Euros. Le 14 décembre 2001, l'introduction en bourse de Crédit Agricole SA, répond à la volo nté du groupe d'accélérer et d'amplifier ses développements dans tous ses métiers et d'élargir encore ses marges stratégiques. Représentative de l'ensemble des métiers du Groupe

Aujourd'hui, leader de la banque de proximité en France grâce à sa structure décentralisée et à la densité de son réseau de 7 200 agences, le groupe Crédit Agricole est aussi très présent des Français avec plus de 16,1 millions de clients. Le groupe est aussi très présent auprès des grandes clientèles, sur les marchés et à l'international, où il est implanté dans 60 pays. Par ailleurs, il poursuit activement son développement dans les métiers de la gestion d'actifs et de la banque privée, en France, en Europe et à l'international.

Crédit Mutuel

Créée en 1882 à la Wantzenau près de Strasbourg sur le modèle bancaire conçu par Fréderic Guillaume Raiffensen en Rhénanie, le crédit mutuel est un réseau bancaire français constitué de 5.390 caisses locales coopératives et mutualistes, regroupées en 18 fédérations régionales, elles-mêmes constituées en confédération nationale. C'est une banque coopérative et mutuelle fortement implantée en France à travers une organisation non centrale, son objectif principal est la qualité des relations et des services qu'il apporte à ses clients.

Le Crédit Mutuel s'organise selon :

- **Caisse locale** : son capital social est détenu par ses sociétaires (porteurs de parts sociales), clients ou salariés. Chaque année se tient une Assemblée générale. Les sociétaires participent à l'élection des administrateurs qui, durant 3 ans, vont les représenter au sein du conseil d'administration de la caisse locale. La caisse a une organisation autonome et gère son budget. Elle est elle-même sociétaire de la caisse régionale dont elle dépend.

- **Caisse Fédérale (au niveau régional)** : Les caisses locales sont réunies en Fédérations (elles détiennent le capital de la caisse Fédérale). La caisse Fédérale est une banque de plein droit. Elle collecte l'épargne, distribue les crédits et propose des services bancaires. Elle prend en charge les emplois réglementaires des caisses locales : réserves obligatoires, ressources affectées, dépôts reversés à la caisse Centrale. De plus, elle est actionnaire majoritaire de la caisse centrale. La Fédération détermine les grandes orientations, décide de sa stratégie et organise la représentation et le contrôle des --caisses locales. Les organes de décision de la Fédération sont la chambre syndicale, véritable parlement interne qui réunit les représentants élus par les Caisses locales, et le conseil d'administration. C'est au sein de la Fédération que s'exprime le dialogue avec les partenaires sociaux. La caisse Fédérale de Crédit Mutuel est une société coopérative de banques, affiliée à la Confédération Nationale du Crédit Mutuel.

- **Caisse centrale (au niveau national)** : son activité est diversifiée via des filiales spécialisées dans la banque d'investissement, les assurances, la gestion d'actifs ou l'immobilier.

BNP Paribas

Un Groupe financier puissant et performant, issu d'une fusion réussie, avec un solide ancrage en Europe, le Groupe BNP PARIBAS est leader en Asie et actif aux Etats-Unis. Parmi les grandes banques françaises, la Banque Nationale de Paris est la plus jeune mais néanmoins celle qui possède l'histoire la plus riche. La BNP est née en 1966 de l'union de deux banques françaises créées au siècle dernier, la BNCI (Banque Nationale pour le Commerce et l'Industrie) et le CNEP (Comptoir National d'Escompte de Paris). Cette fusion a donné le jour à l'une des plus grandes banques mondiales. Sa privatisation, en 1993, a marqué un nouveau temps fort. Enfin, la fusion BNP Paribas a donné naissance à un acteur incontournable du paysage bancaire mondial. A la suite de la fusion du 23 mai 2000 de BNP et de PARIBAS, le rapprochement de PARIBAS Luxembourg et de la Banque Nationale de Paris (Luxembourg) S.A. s'est concrétisé le 17 juillet 2000. BNP PARIBAS Luxembourg occupera ainsi la première place parmi les banques françaises à Luxembourg ; ce rapprochement a permis de dynamiser le développement des créneaux stratégiques en pleine expansion comme la banque privée, la gestion d'actifs, le métier Titres et la Banque de Financement et d'Investissement.

Cette fusion est un événement majeur dans l'histoire bancaire européenne. Grâce à sa taille critique et à son large portefeuille de métiers, BNP Paribas aborde en force la consolidation de l'industrie bancaire en Europe. Il est le premier groupe bancaire en France. Sa capitalisation boursière le place au deuxième rang parmi les banques de la zone Euro.

BNP PARIBAS dispose de l'un des premiers réseaux internationaux au monde, fort de sa présence dans plus de 80 pays, articulée autour de sept places financières de premier plan. La complémentarité de ses activités commerciales et financières permet à BNP PARIBAS de s'imposer dès à présent comme un acteur majeur de la banque de grandes clientèles et de marchés, de la banque internationale et de la gestion d'actifs.

Société Générale

La société générale est une multinationale française, une banque universelle, plus précisément l'une des plus anciennes et importantes banques de France et d'Europe dont le siège est paris. Son histoire commence en 1864 lorsque, le 04 mai Napoléon III signe le décret fondateur de la société générale.

Depuis ses débuts, la société générale s'est développée rapidement sur l'ensemble du territoire français et est rapidement devenue la première institution française. Elle a été privatisée en 1945 et a étendu ses réseaux en fondant Boursorama (avant Fimatex) et en rachetant le Crédit du Nord. Aujourd'hui elle est la 6ème banque d'Europe et la 3ème par le total des actifs. La banque ne veut pas rester à la limite du développement durable prôné par tous les pays du monde. En tant que partenaire du développement durable, elle prend en compte ses responsabilités environnementales et sociales.

A l'échelle de la France métropolitaine

BANQUE	*EFFECTIF*	*FREQUENCE*
Banque Populaire	***2548***	***17,32%***
Crédit Agricole	***5947***	***40,44%***
Crédit Mutuel	***2758***	***18,75%***
Société Générale	***1747***	***11,88%***
BNP Paribas	***1704***	***11,58%***
Total	***14704***	***100,00%***

Tableau X : Tri à plat du nombre d'agences de chaque Banque en France métropolitaine

Donnée finale

Vous trouverez ci-dessous une partie de ce jeu de données représentant les coordonnées des banques.

Banque	Type	Adresse	Longitude	Latitude
Banque Populaire	Coopérative	55-57, RUE ALEXANDRE BÉRARD01500 AMBERIEU EN BUGEY	5.356969	45.96187
Banque Populaire	Coopérative	57, RUE DE LA RÉPUBLIQUE01200 VALSERHONE	5.824057	46.106706
Banque Populaire	Coopérative	7, BD DU MAIL01300 BELLEY	5.688074	45.760875
Banque Populaire	Coopérative	BD EDOUARD HERRIOT01000 BOURG EN BRESSE	5.220434	46.210136
Banque Populaire	Coopérative	12, PL NEUVE01000 BOURG EN BRESSE	5.227292	46.205273
Banque Populaire	Coopérative	168, PL DE LA RÉPUBLIQUE01400 CHATILLON SUR CHALARONNE	4.957873	46.119841
Banque Populaire	Coopérative	865, GRANDE RUE01570 FEILLENS	4.889963	46.3361
Banque Populaire	Coopérative	20, RUE PASTEUR01150 LAGNIEU	5.348758	45.90351
Banque Populaire	Coopérative	37, RUE DE LYON01800 MEXIMIEUX	5.19013	45.90175
Banque Populaire	Coopérative	189, GRANDE RUE01120 MONTLUEL	5.056085	45.851377
Banque Populaire	Coopérative	13, AVE DE BRESSE01460 MONTREAL LA CLUSE	5.574646	46.171224
Banque Populaire	Coopérative	8, PL DU 3 SEPTEMBRE01340 MONTREVEL EN BRESSE	5.128687	46.337708
Banque Populaire	Coopérative	142, RUE ANATOLE FRANCE01100 OYONNAX	5.655023	46.256573
Banque Populaire	Coopérative	81, RUE DES TERREAUX01170 GEX	6.058042	46.332805
Banque Populaire	Coopérative	27, GRANDE RUE01220 DIVONNE LES BAINS	6.138759	46.355629
Banque Populaire	Coopérative	CENTRE D'AUMARD01210 FERNEY VOLTAIRE	6.108245	46.25431
Banque Populaire	Coopérative	PL DE LA FONTAINE01630 ST GENIS POUILLY	6.020704	46.243169
Banque Populaire	Coopérative	33, RUE DE L'EUROPEBÂT A01960 PERONNAS	5.202316	46.179288
Banque Populaire	Coopérative	51, RUE MASONOD01110 PLATEAU D'HAUTEVILLE	5.599233	45.978275
Banque Populaire	Coopérative	25, RUE DU PALAIS01600 TREVOUX	4.77862	45.940309
Banque Populaire	Coopérative	87, RUE DU COMMERCE01330 VILLARS LES DOMBES	5.028284	46.002281
Banque Populaire	Coopérative	45, RUE DU PLATEAU01440 VIRIAT	5.213611	46.214792

Figure X : Data frame caractérisant les coordonnées des différentes banques

Socio-économique

Parlons à présent des données socio-économiques. Pour expliquer brièvement, ces données sont relatives ou concernées par l'interaction des facteurs sociaux et économiques. En général, elles sont utiles afin d'examiner l'évolution économique des sociétés. Dans nos cas, nous avons mis en lien les banques coopératives avec ces données. Pour mieux les comprendre, remémorons-nous que cette étude est basée sur les zones d'emploi. Par définition, une zone d'emploi est un espace géographique à l'intérieur duquel la plupart de la population active réside et travaille. Ce découpage en zone d'emploi sert de référence pour certains critères socio-économiques (par exemple le taux de chômage). La France métropolitaine ainsi que des DROM-COM sont découpés en zone d'emploi.
Cette étude se focalisera sur le découpage le moins ancien, c'est-à-dire celui défini en 2020. Ainsi, nous avons sélectionné et récupéré par zone d'emploi différentes variables socio-économiques sur le site internet de l'Insee afin de mener à bien notre projet. Nous avons pour cela choisi 32 variables récentes comprises entre 2018 et 2020. Ces variables sont réparties selon des catégories distinctes résumant chacune le profil de la population au sein de ces territoires.

Pour commencer, nous distinguons les revenus et pauvreté des ménages en 2019 avec les variables suivantes :

- Nombre de ménages fiscaux
- Nombre de personnes dans les ménages fiscaux
- Médiane du revenu disponible par unité de consommation (en euros)
- Part des ménages fiscaux imposés (en %)
- Taux de pauvreté (en %) – Ensemble
- Taux de pauvreté (en %) – Propriétaire
- Taux de pauvreté (en %) – Locataire
- Revenus (en %) – Ensemble
- Revenus (en %) - Part des revenus d'activité

- Revenus (en %) - Dont part salaires et traitements
- Revenus (en %) - Dont part indemnités et chômage
- Revenus (en %) - Dont part revenus des activités non salariées
- Revenus (en %) - Part des pensions, retraites et rentes
- Revenus (en %) - Part des revenus du patrimoine et autres revenus
- Revenus (en %) - Part de l'ensemble des prestations sociales
- Revenus (en %) - Dont part des prestations familiales
- Revenus (en %) - Dont part des minima sociaux
- Revenus (en %) - Dont part des prestations logement
- Revenus (en %) - Part des impôts
- Distribution des revenus (en euros) - Médiane du revenu disponible par unité de consommation
- Distribution des revenus (sans unité) - Rapport inter décile (9e décile/1er décile)
- Distribution des revenus (en euros) - 1er décile
- Distribution des revenus (en euros) - 9e décile

Ensuite, nous remarquons les variables caractérisant les emplois datant de 2018 avec :

- Emploi salarié – Agriculture
- Emploi salarié – Industrie
- Emploi salarié – Construction
- Emploi salarié - Tertiaire marchand
- Emploi salarié - Tertiaire non marchand
- Emploi salarié - Total
- Emploi non salarié

Pour finir, la variable illustrant le taux de chômage en 2020.

Pour vous donner une idée de la représentation finale de ce jeu de données vous trouverez ci-dessous une partie de notre base de données représentant les données socio-économiques que nous avons utilisé tout au long de ce projet.

Zone d'emploi 2020	Libellé	Nombre de ménages fiscaux	Nombre de personnes dans les ménages fiscaux	Médiane du revenu disponible par unité de consommation (en euros)	Part des ménages fiscaux imposés (en %)	Taux de pauvreté (en %) - Ensemble	Taux de pauvreté (en %) - Propriétaire
0051	Alençon	54227	117585	20580	49,3	14,7	6,8
0052	Arles	62318	142025	20000	51	20,3	9,1
0053	Avignon	124955	283612	20380	53,2	19,7	8,1
0054	Beauvais	115005	280642	21750	59,2	12,9	5,6
0055	Bollène-Pie	33292	77648	20860	53,7	16,9	8,4
0056	Cosne-Cour	32046	64053	20590	50,9	14,8	9,5
0057	Dreux	57941	144689	21760	58,8	13,5	6,2
0058	La Vallée de	35093	76761	20050	49,7	14,5	7,5
0059	Mâcon	71748	159987	21840	57,5	11,7	5,5
0060	Nevers	74340	151259	20640	50,9	15	8,2
0061	Nogent-le-	29036	62245	20800	51,9	13,3	7,7
0062	Redon	37166	83420	20480	48,2	12,4	8,3
0063	Ussel	36621	73575	20070	46,1	15,3	11,5
0064	Valréas	28744	60135	19980	48,8	19,2	12,3
1101	Cergy-Vexir	240753	621050	24050	70,1	12,6	5
1102	Coulommi	29966	74765	22960	63,5	10,2	5,2
1103	Etampes	43559	108263	24320	69,6	9,4	5
1104	Evry	231140	604133	22510	66,7	15,6	5,5
1105	Fontainebl	82031	194807	23500	65,3	12,3	5,1
1106	Marne-la-V	182135	470897	24540	71,8	11,1	5
1107	Meaux	79055	202680	23520	67,3	11,3	5
1108	Melun	77562	192996	22800	65,1	13,5	5
1109	Paris	2869474	6490476	24280	69,3	17,2	6
1110	Provins	23517	57412	21830	59,8	12,1	6,3
1111	Rambouille	32400	79335	28400	78	5,1	5
1112	Roissy	336302	950838	19900	60	22	10,7
1113	Saclay	214741	537494	25700	73,6	11	5
1114	Seine-Yveli	317815	811660	25560	71,7	12	5
1115	Versailles-S	241381	600648	28270	77	8,3	5
2401	Blois	76154	170788	21990	58,2	12,6	5
2402	Bourges	92496	196199	21460	55,1	13,4	6,2
2403	Chartres	89358	208566	23160	64	10	5
2404	Châteaudu	25380	55446	20850	52,3	13	7,1
2405	Châteauro	93502	192187	20430	49,1	14,7	8,1

Figure X : Data frame caractérisant les données socio-économiques

Vous trouverez dans la partie *Réalisation* les descriptions de la récupération ainsi que du nettoyage de ces données. Malheureusement nous n'avons pas pu effectuer des analyses statistiques pour observer les disparités et les similitudes entre les différentes zones d'emploi sur le territoire français. Cependant, nous avons réalisé une carte superposant les données socio-économiques avec la position géographique des banques dans l'intégralité de la France

métropolitaine pour chaque zone d'emploi. Ce qui peut déjà donner des idées sur une analyse traduisant la disparité des types de territoires en y ajoutant directement la notion de banque. Nous avons tenté de mener à bien une analyse faisant entrer en jeu le nombre de banques par zone d'emploi pour y analyser les données socio-économiques plus en détail, tout cela est disponible dans la suite de ce rapport dans la partie *Analyse*.

Réalisation

Récupération des données

De nos jours, une façon possible d'automatiser la récupération de données est le web scraping. D'après Wikipédia, le web scraping est une technique d'extraction du contenu de sites web, via un script ou un programme, dans le but de le transformer pour permettre son utilisation dans un autre contexte. Grâce à cette technique, lorsque les pages ont une même structure, il suffit de disposer de tous les liens auxquels nous voulons soutirer des données. Afin de produire notre projet final, basé sur la visualisation de données spatiales associées aux banques coopératives en France, nous avons dû créer deux jeux de données à l'aide de cette méthode.

Tout d'abord nous avons récupéré les données sur toutes les banques présentées auparavant. Cette base de données a été nommée **bdd_coordonnees_banques2022**. Elle recense le nom de la banque, son type, son adresse ainsi que sa longitude et latitude. Ensuite, une deuxième récupération de données a été nécessaire. En effet, les données socio-économiques pour chaque zone d'emploi de la France métropolitaine de 2020 ont été récupérées et stockées dans un fichier nommé **bdd_social_ze2020**. Toutes les variables obtenues sont présentées dans la partie *Présentation des données.*

Ce qui suit va vous présenter étapes par étapes notre démarche lors de la création de ces deux fichiers. Nous commencerons par expliquer le processus mis en place pour la récupération des coordonnées des banques pour ensuite vous expliquer la récupération des données socio-économiques.

Banques

La première tâche à effectuer lors de notre projet était l'une des plus importante : nous devions récupérer les données, sur lesquelles nous devions travailler. Cinq banques en France métropolitaine sont à récupérer depuis leur site officiel. Cependant pour la BNP Paribas, nous n'avons pas pu récupérer les données sur leur site donc nous avons dû utiliser un autre site où les données étaient disponibles. Vous trouverez ce lien en *Référence X.* Pour être plus précis, l'objectif premier était d'uniquement récupérer l'adresse de toutes les agences, banque par banque.

Comme précisé précédemment, nous avons décidé de faire l'intégralité de notre projet en langage ***R***. Nous avons dû installer différentes librairies (nommé *package* dans ce langage, nous utiliserons cette appellation plus tard dans le rapport). La première librairie utile pour le web scraping est ***rvest***. Elle aide à l'extraction d'information sur des pages web et facilite l'expression des tâches courantes de web scraping. Une deuxième bibliothèque utile pour notre récupération de données est le package ***plyr***. Il regroupe un ensemble d'outils qui permettent de résoudre des problèmes tels que diviser une grosse structure de données en morceaux homogènes, appliquer une fonction à chaque morceau et enfin combiner tous les résultats. Troisièmement, nous avons utilisé la librairie ***dplyr***, elle fournit une grammaire de manipulation de données, fournissant un ensemble de verbes aidant à la résolution de manipulation de structure de données. Une autre librairie utile a été ***BanR***. Ce package utilise l'api BAN (Base Adresse Nationale). BAN est un jeu de données public des adresses

françaises produit par OpenStreetMap, la Poste, l'IGN et Etalab. Ce package comporte plusieurs fonctions pour trouver une longitude et latitude à partir d'une adresse et inversement. Enfin, vous comprendrez un peu plus loin mais nous avons eu besoin d'un package de manipulation de chaine de caractères, nous nous sommes servis de ***stringr*** ainsi que ***stringi***.

Nous avons présenté presque tous les outils nécessaires pour pouvoir commencer le web scraping, un dernier et pas des moindres est l'extension de Google ***selectorGadge*****t**. Il facilite la génération et la découverte de sélecteur CSS sur des sites complexes. Pour faire simple, il permet d'attribuer une espèce d'adresse à un élément remarquable d'un site internet (un titre ou un tableau). Lorsque l'on lance l'extension, une boîte s'ouvre en bas à droite de l'écran, il suffit de cliquer sur l'élément de la page que vous voulez récupérer. Si un sélecteur correspond, un cadre vert apparaitra et nous pouvons récupérer l'adresse de l'élément. Ci-dessous, vous trouverez un exemple pour sélectionner les adresses du Crédit Agricole :

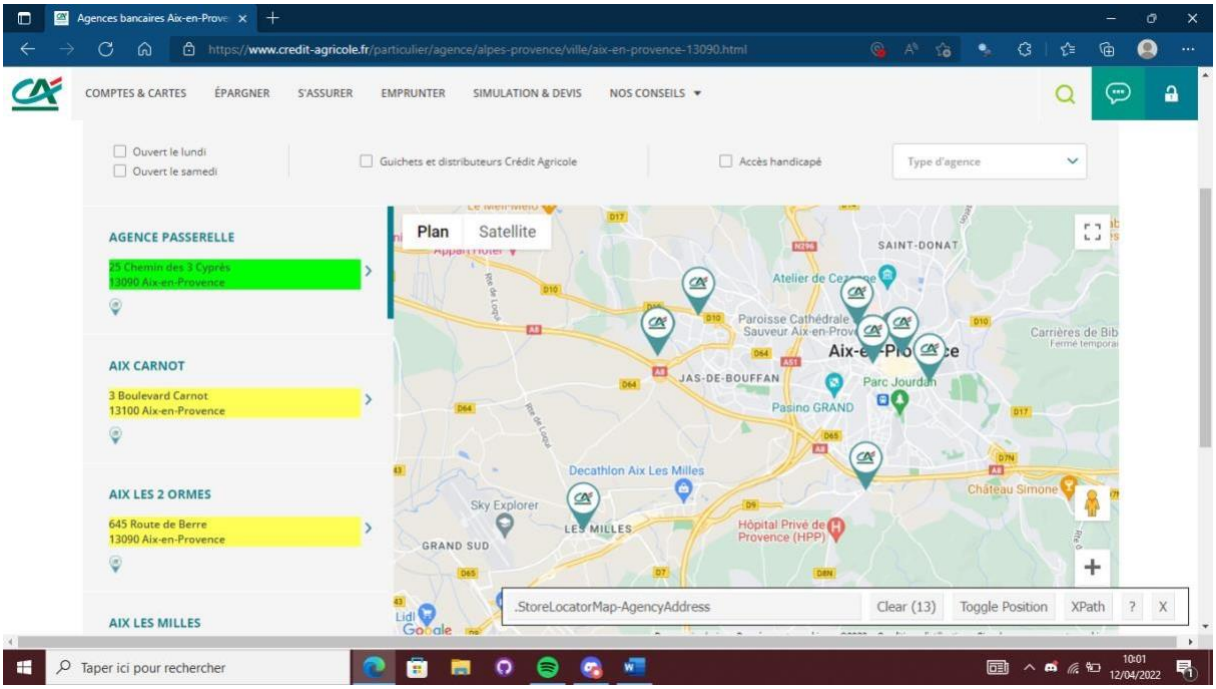

Figure X : Exemple de récupération d'adresse pour le Crédit Agricole à l'aide de l'outil selectorGagnet

Sur l'exemple ci-dessus, vous pouvez voir que la première adresse (en vert) a été sectionnée, les adresses qui suivent (en jaune) sont mises en évidence car l'extension reconnaît que ce sélecteur est du même type que celui sélectionné en vert. Dans la boîte en bas à droite, nous retrouvons le nom de l'adresse sélectionnée : « .StoreLocatorMap-AgencyAddress » et le nombre de sélecteur similaire disponible sur cette page (ici 13). Ainsi en utilisant ce nom nous allons pouvoir récupérer toutes les informations sélectionnées en vert et jaune, c'est-à-dire les treize adresses.

Afin d'automatiser la récupération de données, nous avions besoin de comprendre la logique qui reliait tous les liens. En effet, les pages ayant toutes à peu près la même structure, une fois que l'on possède une boucle générant chaque lien, il est assez aisé de récupérer les données en même temps. Seule subtilité, pour chaque banque nous avons dû adopter une technique différente.
Ce qui suit vous explique donc banque par banque la récupération de lien.

Le site du Crédit Agricole présente les régions puis dans chaque région les villes par ordre alphabétique où contiennent au moins une agence. Lorsque les villes contiennent des arrondissements, ils sont détaillées (Paris 01 par exemple). Ainsi pour mieux comprendre, voici l'url qui est sous la forme suivante :

https://www.credit-agricole.fr/particulier/agence/region**/ville/**ville-code postal**.html**

Dans ce cas, nous avons dû utiliser un fichier annexe contenant les codes postaux associés aux villes de France. Sur le site, certaines villes ont plusieurs codes postaux ce qui entrainent des doublons d'adresses. Ses doublons seront bien sur effacés à la fin de la récupération.

Les sites de la Banque Populaire et de la Société Générale sont construits de la même manière. Les urls sont de la forme :

https://agences.nom de la banque.fr**/banque-assurance/agences-**departement-numero du département

Nous avons fait un web scraping pour récupérer la liste de chaque département ainsi que leur numéro. A partir de ce lien, nous pouvons récupérer toutes les adresses d'un département et ainsi récupérer toutes les adresses des agences présents pour ses banques.

Le site du Crédit Mutuel est composé d'une liste de département puis des villes contenant une agence dans un département sélectionné. Dans une troisième page imbriquée nous avons une liste d'adresses d'agences de la ville choisie ainsi que d'autres aux alentours. Ça ne pose pas de problèmes en revanche cela entraine une quantité très importante de doublons d'adresses.

Le Crédit Mutuel de Bretagne est composé de plusieurs pages également imbriquée, les noms des départements puis le nom des villes avec leur codes postaux et enfin sur une troisième page, l'adresse exacte de l'agence. Nous avons dû formater l'url pour accéder à la page des villes et à celle de l'adresse.

Pour la BNP Paribas, le site utilisé divise également les agences en département. Si le nombre d'agences est important, la page est divisée en plusieurs pages. L'url est sous la forme suivante :

https://www.moneyvox.fr/pratique/agences/bnp-paribas/numero du département/ s'il y a plusieurs pages il faut rajouter un « numéro de la page/ »

Après avoir réalisé ce web scraping nous avons fait un premier tri des adresses car nous avions un nombre conséquent de doublons. Cela a été causé par le format du site ou encore du fait que l'agence était partagée en différente parties (par exemple : l'agence à Montélimar 23 rue Raymond Daujat qui a une partie « classique » et une partie réservée aux entreprises et apparait donc deux fois dans les adresses récupérées).

Tous les liens générés et toutes les adresses récupérées par banques, nous nous retrouvions avec une base de données comportant le nom de la banque, son type ainsi que son adresse. Or pour ce projet nous voulions et nous avions besoin des points exactes des banques, c'est-à-dire nous avions besoin de la longitude et la latitude de chaque adresse. Une dernière étape était donc primordiale.

Nous avons utilisé le package ***BanR*** qui, grâce la fonction geocode, nous a permis de convertir une adresse française en longitude et latitude. Cependant, il se pouvait que cette fonction ne nous renvoyait pas de résultats ou nous renvoyait une fausse conversion, ce qui était dû au

fait qu'elle ne reconnaissait pas l'adresse entrée en paramètre. Ainsi pour chaque adresse de banque erronée, il a fallu les modifier manuellement.

Une fois tout cela réalisé, nous avons dû faire un deuxième tri des doublons d'adresses. En effet, certaines adresses de la même banque étaient situées au même endroit du fait que certaines adresses changeaient seulement de codes postaux mais désignaient la même agence. Ainsi pour pallier cette erreur, nous avons utilisé le couple longitude-latitude pour que ce couple soit unique. Nous obtenons après cela, toutes banques confondues, 14704 agences différentes en France métropolitaine.

Socio-économique

Une autre tâche à réaliser était la récolte de toutes nos variables socio-économiques. Nous disposions déjà de données concernant certaines variables. Cependant nous avons dû les remettre au goût du jour pour que cela puisse correspondre le plus possible aux données actuelles.

Nous avons récolté les différentes variables socio-économiques sur le même principe que la récupération des coordonnées des banques. Cette récupération s'est divisée en quatre parties.

Premièrement, 23 variables ont été récoltées sur le site de l'Insee. Ces dernières caractérisant les revenus et pauvreté des ménages en 2019.
Afin de récupérer ces variables pour chaque zone d'emploi, nous avons automatisé le processus en utilisant la même technique que pour la récupération des adresses des banques. Nous avons utilisé l'url. Ce dernier est de la forme suivante :

https://www.insee.fr/fr/statistiques/6037462?geo=ZE2020-zoned'emploi

Comme vous pouvez le constater, ce lien ne présente pas de difficultés particulières. Nous devions simplement disposer de toutes les zones d'emplois et réaliser une boucle récupérant toutes les données. Ci- dessous un exemple réalisé pour la première zone d'emploi.

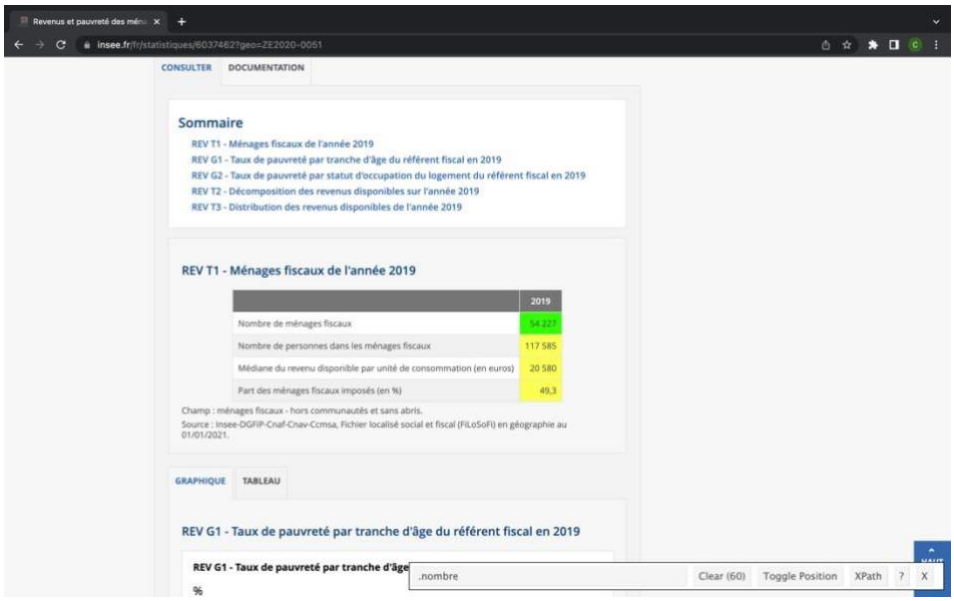

Figure X : Exemple de la récupération des revenus et pauvreté des ménages en 2019 pour la zone d'emploi 0051 à l'aide de l'outil selectorGagnet

Une fois les données récupérées nous avons dû les modifier pour obtenir la forme et le type que nous désirions. Tout cela a été possible à l'aide du package ***stringr.***

Deuxièmement, nous avons dû récupérer 7 variables caractérisant les emplois datant de 2018. Ces dernières étaient présentes dans un fichier Excel présent sur le site l'Insee. Vous pouvez retrouver ce fichier au lien en *Référence X*. Nous ne voulions que certaines informations présentes dans ce fichier c'est pourquoi nous avons réalisé un petit code qui nous a permis de récupérer uniquement les variables qui nous paraissaient les plus pertinentes.

Troisièmement, la récolte des données concernant le taux de chômage en 2020 a été réalisé de la même façon que les variables caractérisant les emplois. Un fichier Excel, présent sur l'Insee, répertoriant tous les taux de chômage depuis 2003 jusqu'à 2020. Vous pouvez retrouver ce fichier au lien en *Référence X.*

Enfin, concernant la récupération des informations sur la population active en 2017, nous l'avons réalisé exactement de la même manière que pour les variables sur le revenus et pauvreté des ménages. Voici l'url :

https://www.insee.fr/fr/statistiques/4515512?sommaire=4515574&geo=ZE2020-zoned'emploi

Une fois avoir réalisé notre web scraping et nettoyage des données pour chaque type de variables, nous avons réalisé une concaténation de toutes ces données et nous obtenons le jeu de données ci-dessous (présenté dans *Présentation des données*) :

Zone d'emploi 2020	Libellé	Nombre de ménages fiscaux	Nombre de personnes dans les ménages fiscaux	Médiane du revenu disponible par unité de consommation (en euros)	Part des ménages fiscaux imposés (en %)	Taux de pauvreté (en %) - Ensemble
0051	Alençon	54227	117585	20580	49,3	14,7
0052	Arles	62318	142025	20000	51	20,3
0053	Avignon	124955	283612	20380	53,2	19,7
0054	Beauvais	115005	280642	21750	59,2	12,9
0055	Bollène-Pie	33292	77648	20860	53,7	16,9
0056	Cosne-Cour	32046	64053	20590	50,9	14,8
0057	Dreux	57941	144689	21760	58,8	13,5
0058	La Vallée de	35093	76761	20050	49,7	14,5
0059	Mâcon	71748	159987	21840	57,5	11,7
0060	Nevers	74340	151259	20640	50,9	15
0061	Nogent-le-	29036	62245	20800	51,9	13,3
0062	Redon	37166	83420	20480	48,2	12,4
0063	Ussel	36621	73575	20070	46,1	15,3
0064	Valréas	28744	60135	19980	48,8	19,2
1101	Cergy-Vexir	240753	621050	24050	70,1	12,6
1102	Coulommi	29966	74765	22960	63,5	10,2
1103	Etampes	43559	108263	24320	69,6	9,4
1104	Evry	231140	604133	22510	66,7	15,6
1105	Fontainebl	82031	194807	23500	65,3	12,3
1106	Marne-la-V	182135	470897	24540	71,8	11,1
1107	Meaux	79055	202680	23520	67,3	11,3
1108	Melun	77562	192996	22800	65,1	13,5
1109	Paris	2869474	6490476	24280	69,3	17,2
1110	Provins	23517	57412	21830	59,8	12,1
1111	Rambouille	32400	79335	28400	78	5,1
1112	Roissy	336302	950838	19900	60	22
1113	Saclay	214741	537494	25700	73,6	11
1114	Seine-Yveli	317815	811660	25560	71,7	12
1115	Versailles-S	241381	600648	28270	77	8,3
2401	Blois	76154	170788	21990	58,2	12,6
2402	Bourges	92496	196199	21460	55,1	13,4
2403	Chartres	89358	208566	23160	64	10
2404	Châteaudu	25380	55446	20850	52,3	13
2405	Châteauro	93502	192187	20430	49,1	14,7
2406	Chinon	19994	43400	21340	52,3	12,5
2407	Gien	36219	79568	21170	54,9	13,6
2408	Loches	25780	53914	20520	46,6	13,3
2409	Montargis	56611	126422	20700	54	15,7

Figure X : Data frame caractérisant les données socio-économiques

Toutes les données maintenant récoltées et archivées dans les bases de données, nous avons pu commencer à jouer avec.

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)*.

```
Simple feature collection with 6 features and 5 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: 2.184161 ymin: 44.4716 xmax: 5.047536 ymax: 49.77987
Geodetic CRS:  WGS 84
   code             libelle ze2020                lb_clst part_rg                       geometry
1 07176          Planzolles   8402                Aubenas    <NA> MULTIPOLYGON (((4.1667 44.4...
2 07192          Rochepaule   8420 Les Sources de la Loire    <NA> MULTIPOLYGON (((4.495502 45...
3 08101         La Chapelle   4421                  Sedan    <NA> MULTIPOLYGON (((5.017189 49...
4 18284          Villegenon   2407                   Gien    <NA> MULTIPOLYGON (((2.597565 47...
5 18285 Villeneuve-sur-Cher   2402                Bourges    <NA> MULTIPOLYGON (((2.210953 46...
6 18286         Villequiers   2402                Bourges    <NA> MULTIPOLYGON (((2.784898 47...
```

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.

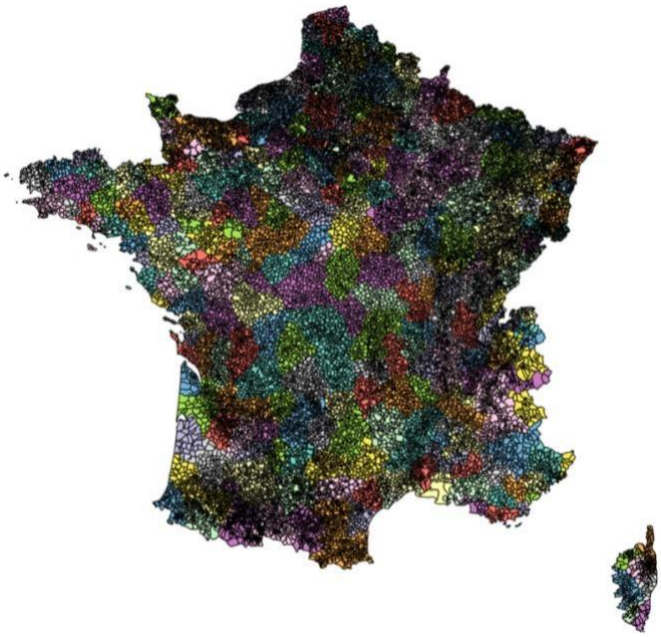

Figure X1 : Tracé des communes colorées par zone d'emplois

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***, qui 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'emploi. 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 obtenu le tracé présent dans la Figure X2.

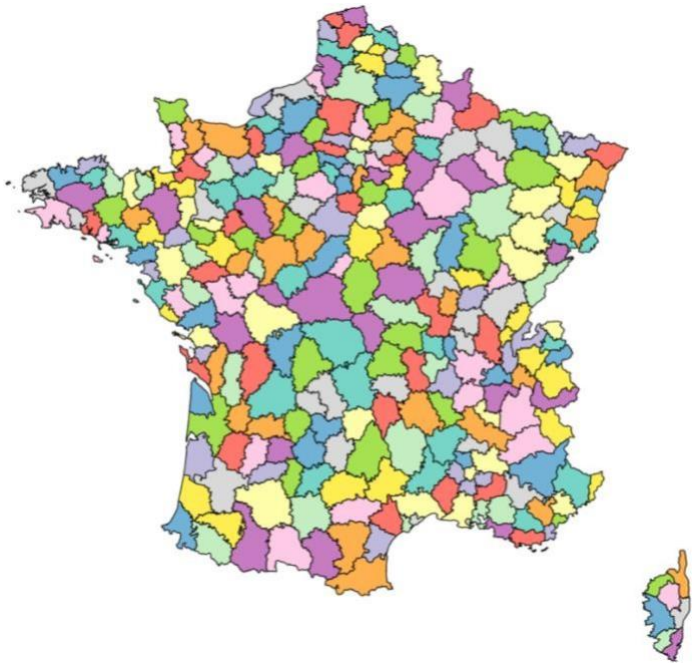

Figure X2 : Tracé des zones d'emploi

Une fois les coordonnées des zones d'emplois bien définies, nous avons commencé à regarder comment nous pourrions ajouter les variables socio-économiques ainsi que la position géographique des banques. Ces données furent récupérées par une partie du groupe pendant que l'autre travaillait sur la carte.
Les premières données que nous avons recueilli étaient celles des banques. Nous avons donc commencé par tracer les positions géographiques des banques. Vous pouvez en voir un exemple sur la Figure X3.

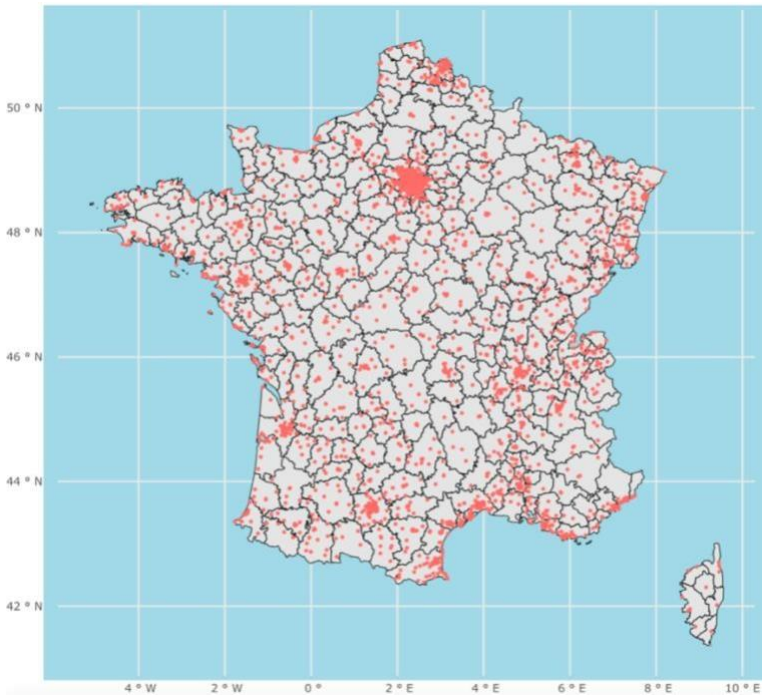

Figure X3 : Position géographique des banques Banque Populaire

Après avoir compris comment afficher la position des banques, nous avons voulu colorer les zones d'emploi en sélectionnant un critère socio-économique.
Pour tracer ces cartes, nous avons dans un premier temps utilisé le package ***ggplot2*** qui permet une certaine aisance pour superposer des objets provenant de classe différente et de jeux de données différents. Cela fût très utile dans notre cas car nous avions d'un côté un jeu de données de classe ***sf*** (*cf. shapefile zones d'emploi*) et de l'autre une simple data frame contenant les données socio-économiques par zone d'emploi.
Ainsi, nous avions enfin une carte permettant de visualiser la position géographique des banques en fonction de critères socio-économiques par zone d'emploi. Vous pouvez en voir un exemple dans la Figure X4.

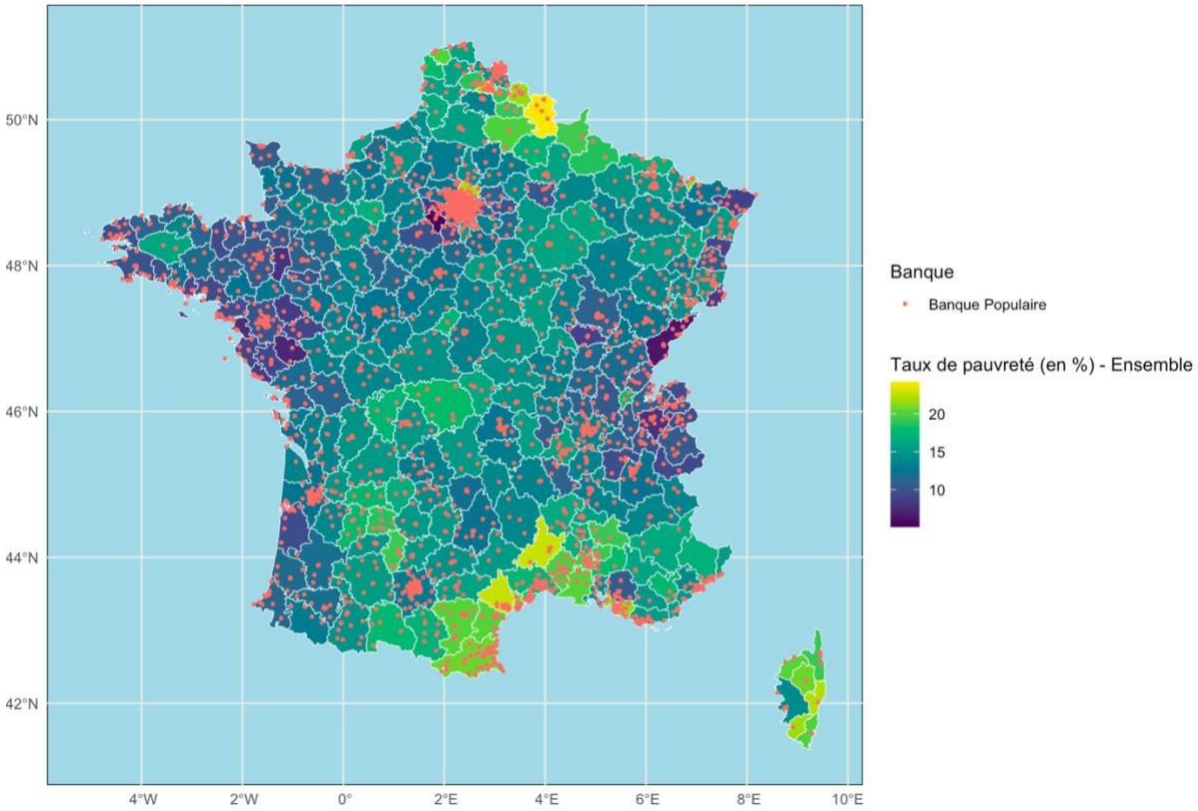

Figure X4 : Position géographique des banques Banque Populaire en fonction du taux de pauvreté par zone d'emploi

Le but étant de fournir une application de visualisation de données, nous avons voulu rendre ces cartes interactives : qu'on puisse zoomer, qu'on puisse sélectionner seulement une partie de la carte, qu'on puisse bouger sur la carte, etc… Et pour cela, il faut passer par le package ***plotly***. Grâce à la fonction *ggplotly*(), nous pouvons transformer un objet *ggplot* en objet *plotly*, ce qui nous a permis d'obtenir des cartes interactives et dynamiques. Toutefois, ce type de représentation ne nous fournit pas toutes les informations. En effet, nous avons bien une représentation globale de la France avec la position des banques associées aux critères socio-économiques, mais nous ne savons plus quelle zone d'emploi est laquelle. C'est pour cela que par la suite, nous avons fait le choix d'ajouter sur la carte, le code associé à chaque zone d'emploi. Par ailleurs, ce type de carte ne nous fournit pas le nom des villes.

Afin d'aller plus loin dans la représentation, nous avons voulu essayer de produire des cartes directement via ***plotly***. Pour cela nous sommes passés par l'API *Mapbox*.

Mapbox est une entreprise américaine spécialisée dans la cartographie en ligne. Elle a développé pour de nombreux langage une API permettant d'accéder aux services de *Mapbox* dont nombreux d'entre eux utlisent des données d'*OpenStreetMap*.

Pour utiliser l'API sous ***R***, il faut créer un compte *Mapbox* afin d'avoir accès à un *token*. C'est ce dernier qui nous permettra d'avoir accès à tous les services. Nous avons donc regardé comment afficher la position des banques en fonction de critères socio-économiques par zone d'emploi. Afin de comprendre comment fonctionne l'API avec ***plotly***, nous avons divisé le travail en trois. Nous avons tout d'abord regardé comment afficher les banques sur la carte, puis regardé comment colorer les zones d'emploi en fonction de critères socio-économiques, et enfin essayer de combiner les deux.
Vous pouvez voir le résultat sur la Figure X5.

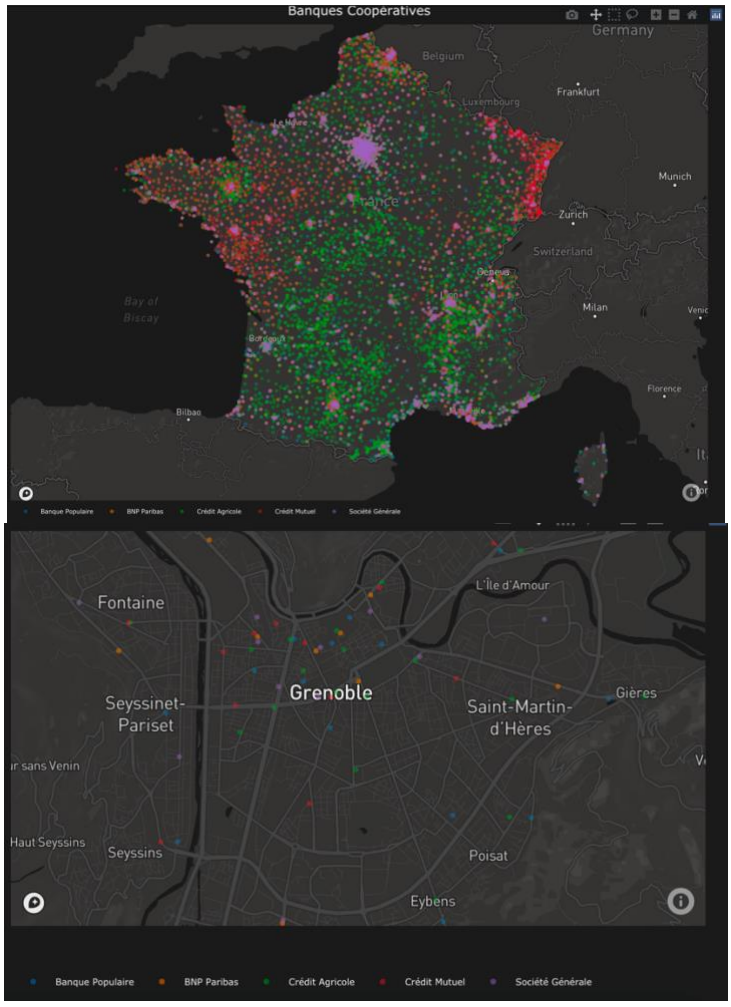

Figure X5 : Position géographiques de toutes les banques

Nous voyons qu'avec *Mapbox*, les cartes sont plus précisent. En effet, nous avons le nom des villes, les noms des rues, les reliefs et si nous zoomons beaucoup, nous pouvons même voir la forme des bâtiments. De plus, ce type de carte est plus ergonomique et l'utilisation est plus aisée et plus ludique que la version précédente.

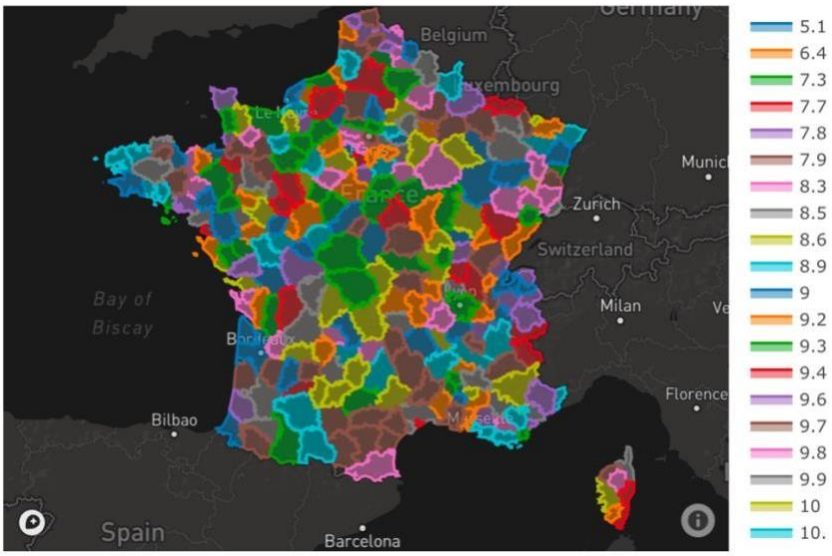

Figure X6 : Représentation du taux de pauvreté par zone d'emploi

Nous avons par la suite tracé les zones d'emploi en fonction de critères socio-économiques. Mais comme vous pouvez le voir sur la Figure X6, la carte n'est pas lisible. Le code couleur n'a aucun sens et la légende n'est pas lisible. ***Plotly*** lit les pourcentages comme des valeurs discrètes et non continues. C'est pour cela que les valeurs sont ainsi affichées dans la légende. Il est sans doute possible de pouvoir jouer sur les paramètres afin d'ajouter une palette de couleurs et de rendre les valeurs continues, mais nous voyons bien qui si nous combinons la carte dans la Figure X5 et celle de la Figure X6, le résultat ne serait pas très lisible. Nous n'avons d'ailleurs pas réussi à combiner les deux cartes.

La principale différence est qu'avec ***ggplot***, nous traçons la carte sur un fond vierge alors qu'avec ***plotly*** et *Mapbox*, nous affichons les points et traçons les points sur une carte de base. Comme nous avons pu le voir, *Mapbox* offre de nombreuses fonctionnalités qui permettent une lecture simple est aisée des cartes. Toutefois, pour le type de représentation que nous voulons, c'est-à-dire afficher la position géographique des banques ainsi que les variables socio-économiques, la version ***ggplot*** semble plus adaptée. Toutefois, si nous voulons avoir plus d'informations sur la position des banques, alors la version avec *Mapbox* serait mieux.

Pour aller plus loin, nous réaliserons également des cartes seulement à l'échelle des zones d'emploi. Ces dernières seront réalisées à l'aide de ***ggplot*** et détaillés dans la partie *Analyse*.

Interface shiny

Création package

Analyse de la gestion

Difficultés rencontrées

Pendant la réalisation de notre projet, nous nous sommes confrontés à de nombreuses difficultés.

En premier lieu, le web scraping d'une des banques a posé problème : celui du Crédit Agricole.

En effet, tout d'abord la récupération des adresses a été compliquée par le fait que le code postal des villes avait été intégré à l'URL menant aux informations d'une agence. Pour pallier ce problème, nous avons dû faire recourt à une base de données permettant d'associer à un code postal un nom de ville. Ainsi, il suffisait de tester les différentes combinaisons entre le nom d'une ville et les codes postaux correspondants, ce qui a considérablement rallongé le temps d'exécution du web scraping.

Enfin, une fois les adresses récupérées, il s'agissait de déterminer la latitude et la longitude de chacune d'elles, en utilisant le package ***BanR***. Cependant, nombreuses étaient les adresses pour lesquelles ce package ne trouvait pas de coordonnées géographiques, notamment les centres commerciaux. De ce fait, nous avons dû corriger toutes ces adresses une par une en utilisant Google Maps (soit en les rectifiant, soit en choisissant une adresse à côté), afin de trouver leur localisation.

En second lieu, la deuxième difficulté à laquelle nous avons eu à faire était la réalisation de la carte.

Tout d'abord, afin de faciliter la compréhension et la manipulation de shapefiles, nous avions dû installer le package ***sf***. L'installation de ce package fût compliquée. En effet, cette librairie fonctionne avec des dépendances qui doivent directement être installées sur l'ordinateur. Parmi celles-ci, nous y retrouvons GDAL, RGEOS et PROJ. Il a donc d'abord été nécessaire d'installer ces dépendances avant de pouvoir installer ***sf***.

Une fois installé, nous avons mis du temps à comprendre comment fonctionnaient les objets ***sf*** et comment les manipuler.
Lors de la fusion des polygones, la colonne *geometry* n'était plus composée d'objets de classe *sfc MULTIPOLYGON* mais *POLYGON*. Or pour tracer les cartes, il faut que la colonne *geometry* soit de classe *sfc MULTIPOLYGON*. Il nous a donc fallu comprendre qu'il fallait caster la colonne pour pouvoir obtenir des cartes.

Enfin, l'une des principales difficultés fût la compréhension et la manipulation du package ***plotly*** combiné à l'API *Mapbox*. Tout d'abord, il est très compliqué de combiner plusieurs variables sur une même carte. Comme expliqué dans la partie *Réalisation*, obtenir les zones d'emploi colorées en fonction de critères socio-économiques avec position des banques par-dessus donnerait quelque chose illisible. Toutefois, nous avons trouvé la possibilité de tracer les frontières des zones d'emploi et d'y ajouter les positions géographiques des banques à l'intérieur.

Pistes d'améliorations

Bien que notre projet ait été mené à bien, quelques améliorations auraient pu être faites si nous avions eu plus de temps.

Premièrement, nous aurions pu récupérer les positions des agences d'autres banques, notamment celles de la Banque Postale ou encore du Crédit Lyonnais.

En second lieu, une carte dynamique, réalisée grâce à la librairie ***plotly,*** aurait pu être conçue. En effet, nous sommes parvenus à faire deux cartes : une représentant les positions des banques et une autre permettant de visualiser les critères socio-économiques par zone d'emploi. Cependant, ce que nous n'avons pas réussi à faire est de superposer ces deux cartes.

Ensuite, nous aurions pu rajouter les données d'outre-mer. En effet, l'ensemble des banques pour lesquelles nous avons étudié les positions est généralement présent en Martinique, Guadeloupe, Guyane, ainsi qu'à la Réunion.

En plus de ça, si cela avait été possible, il aurait été intéressant de récupérer les données socio-économiques et les répartitions des agences pour différentes années, afin de comparer ces données en 1970 et en 2022 par exemple, ou bien d'apprécier l'évolution de ces dernières. En effet, il est impossible d'obtenir ces données pour des années passées, donc une étude rétrospective est impossible aujourd'hui. Cependant, si le web scraping de ces données est réalisé, par exemple, chaque année et qu'elles sont sauvegardées et stockées, il sera désormais possible d'effectuer des comparaisons temporelles. En revanche, il faut garder à l'esprit que les zones d'emploi changent de temps en temps.

Pour terminer, nous pourrions nous intéresser à d'autres pays et réaliser la même étude sur le Québec par exemple, en considérant les banques présentes sur le territoire, telles que la Banque Royale du Canada ou encore la Banque Scotia.

Durabilité du projet

Pour aller plus loin, nous allons proposer une solution pour récupérer les données à jour pour faire perdurer notre projet dans le temps.

Concernant le web scraping des banques, le fichier *web_scraping_banques_V2.R* fonctionnera pour récupérer les adresses si la structure des sites internet ne change pas. Pour récupérer les couples longitude-latitude, notre code risque de ne plus marcher s'il y a de nouvelles agences. Comme énoncé dans la partie précédente, certaines adresses ont été changées manuellement et il faudra certainement refaire cette manipulation pour des nouvelles adresses.
Pour la récupération des zones d'emplois, il suffit de récupérer le dossier compressé sur l'INSEE qui contient le fichier shapefile qui lui-même contient les polygones des zones d'emploi actuelles.

Enfin, pour mettre à jour les données socio-économiques, il faut rechercher les pages internet qui contiennent les informations voulues et télécharger les fichiers .xlsx à jour. Il faut ensuite modifier le fichier *webscraping_socio_V1.R* en actualisant les urls et les lectures de fichier .xlsx. Il est possible que ça ne soit pas suffisant si la structure des pages pour les mêmes données ou celle des fichiers change et alors il faudra plus modifier le code. Si lors d'une mise à jour il y a plus ou moins de critères socio-économiques par rapport à avant ou que l'ordre des critères change, il faudra aussi changer le nom des colonnes du fichier des données socio-économiques pour que ça soit cohérent.

Conclusion

Annexe

Analyse

Utilisation de notre git

Références