

Rapport Méthodologique

Comment vote-t-on en fonction de notre accès à internet ?

I. La démarche

1. Choix du sujet

Nous avons choisi d'associer à la fracture numérique l'étude du vote dans la région Auvergne Rhones Alpes.

Tout d'abord le choix de la région Rhones Alpes semble propice à révéler des fractures numériques étant donné un relief important, les Alpes à l'est et le Massif Central à l'ouest, car ils présagent des disparités naturelles en termes d'infrastructures numériques. De plus, la présence d'une frontière avec la Suisse apporte une diversité que nous ne trouvons pas dans certaines régions françaises. Ces paramètres ont donc un potentiel à révéler des différences dans les critères que nous avons choisi d'étudier.

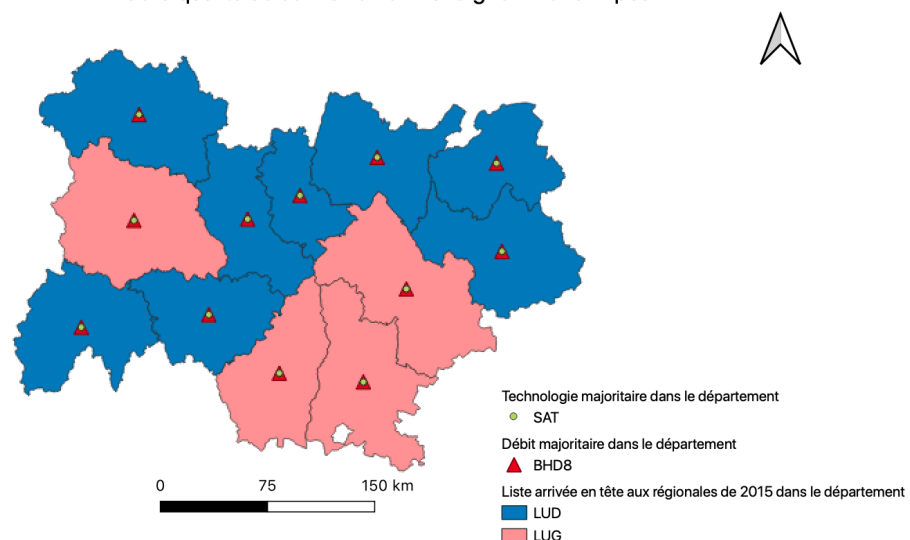
En particulier, nous avons choisi de centrer nos recherches sur les élections régionales pour deux raisons. D'une part, les élections municipales comportent des candidats trop divers et parfois difficiles à caractériser politiquement. D'autre part, les européennes n'ont pas encore assez d'électeurs pour être la représentation la plus fidèle des tendances politiques. Étant donné que l'étude se centre sur la région Auvergne-Rhône-Alpes, le choix des élections régionales semble le plus adapté.

Finalement, l'association du vote et de la fracture numérique peuvent faire écho à un phénomène bien connu et étudié de façon empirique, le vote rural. On peut en effet spéculer que les zones définies comme rurales, c'est-à-dire éloignées des centres urbains et des axes de communication avec une faible densité de population, seraient aussi des zones avec de mauvaises connexions internet ou du moins un choix de technologies plus faible.

2. Une observation limitée de la polarisation classique du vote Droite/Gauche et de la fracture numérique

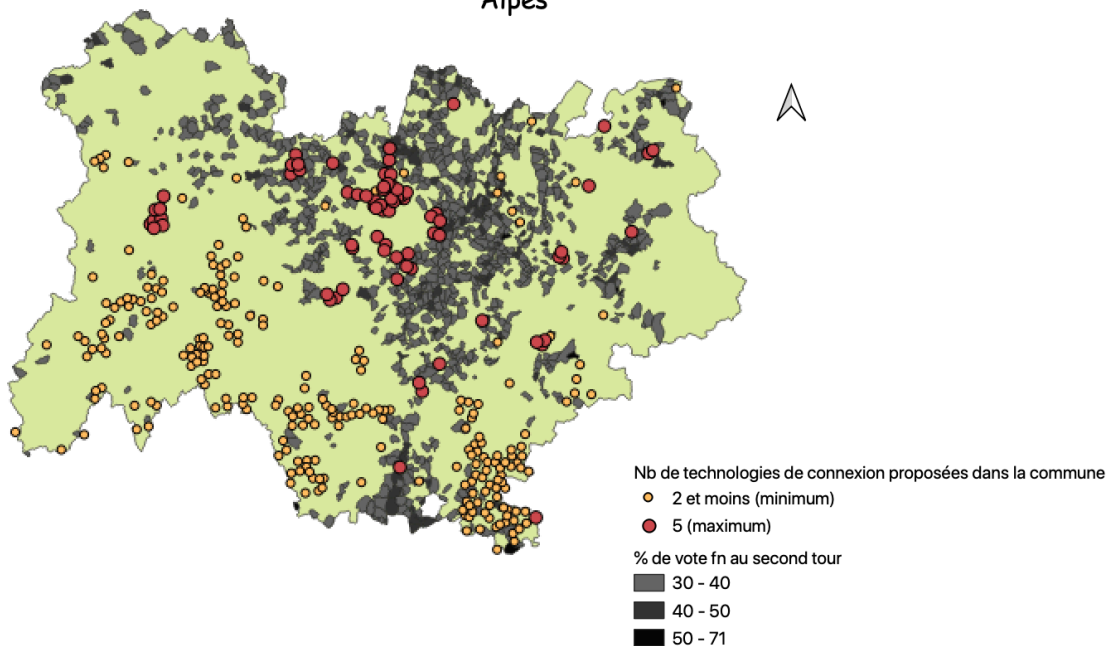
Aussi bien au niveau départemental, communal et infra-communal, nous n'observons pas de corrélation réelle entre l'orientation du vote sur un axe gauche/droite et la qualité de la connexion à internet. La carte suivante illustre ce propos :

Carte des résultats aux deuxième tour des élections Régionales de 2015 et de la qualité de connexion en Auvergne-Rhône-Alpes



Comme nous pouvons le voir au niveau départemental, une observation des pratiques de vote en région Rhones-Alpes sous le prisme de la scission droite/gauche ne donne pas de résultat satisfaisant. Également, se porter sur une vision macroscopique s'avère également limitée (cf II – Etapes). Nous avons donc décidé de modifier notre angle d'observation et nous porter sur les pratiques de vote contestataires : le vote Front National, l'abstention et le vote blanc. Ce dernier élément, présent en nombre trop limité pour une étude (environ 3% en moyenne), a été abandonné par la suite. Une approche plus microscopique et basée sur les indicateurs de vote nouvellement définis nous donne une vision plus claire et pertinente de nos indicateurs :

Répartition de l' offre de connexion par ville et du vote FN de plus de 30% au second tour des élections régionales de 2015 dans la région Auvergne-Rhone-Alpes



II. Etapes et difficultés

1. Visualisation du jeu de données des connexions

Nous avons commencé par projeter les données brutes. Une réflexion a ensuite été entreprise sur la manière de les transformer afin d'avoir une représentation fidèle de notre sujet, notamment définir à partir des différentes variables ce qu'a priori est une bonne ou une mauvaise connexion.

2. Récupération des données de vote

Les données de vote furent récupérées sur le site data.gouv.fr sur le profil du Ministère de l'Intérieur (adresse non disponible en raison de l'incendie des serveurs à Strasbourg).

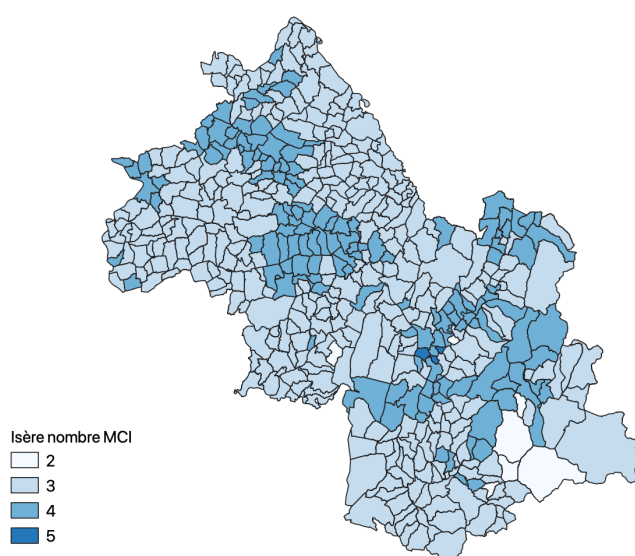
3. Nettoyage, formatage et transformation

A. Données de connexion

L'enjeu avec les données de connexion était de définir la qualité d'une connexion internet. A partir des deux jeux de données disponibles, plusieurs hypothèses pouvaient être faites.

Le nombre de modes de connexions disponibles fut établi comme le premier critère pouvant départager des communes plus ou moins connectées. Avec Python, une colonne `nbr_mci` a calculé le nombre de colonnes `log_*` étant supérieure à 0.

Nombre de modes de connexion proposés dans les communes de
Isère



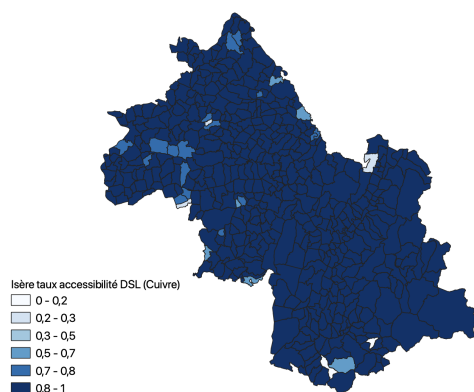
La visualisation ci-dessus permet une distinction de certains territoires mais l'analyse des données quantitatives contient quelques nuances sur l'accessibilité aux technologies proposées. En effet tous les logements et locaux professionnels dans une commune ne sont pas systématiquement équipés, par exemple de la fibre, alors que leur commune la propose.

Nous avons ainsi défini un nouvel indicateur, le taux d'accessibilité par mode de connexion ainsi qu'un taux d'accessibilité global qui calcule la moyenne de tous les taux précédents. Toujours sous Python, le nouvel indicateur calcule la proportion de `log_*` par rapport au nombre d'entités.

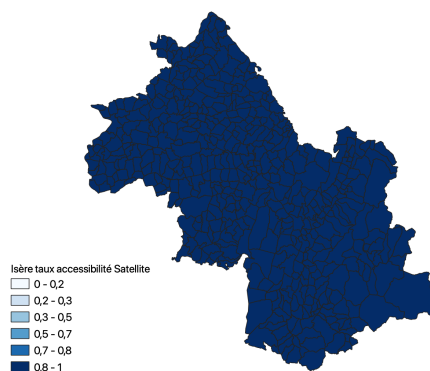
Sur la page suivante, les cartes des taux d'accessibilité nous permettent d'observer une totale couverture de la DSL ou du satellite, et l'absence de couverture pour les réseaux radios et la quasi-absence du câble. La fibre et la 4G présentent en revanche une répartition plus hétérogène, ce qui nous permettra de les choisir comme paramètres plus tard.

Le taux global est finalement assez peu pertinent étant donné l'homogénéité de plusieurs technologies qui pèsent dans la moyenne.

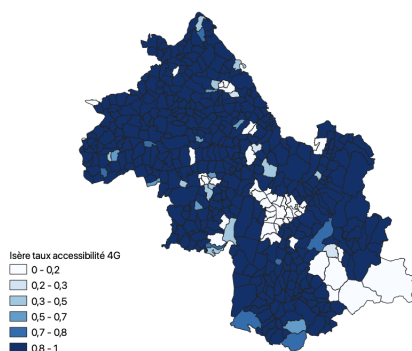
Taux d'accessibilité à la DSL des logements et locaux professionnels dans les communes de l'Isère



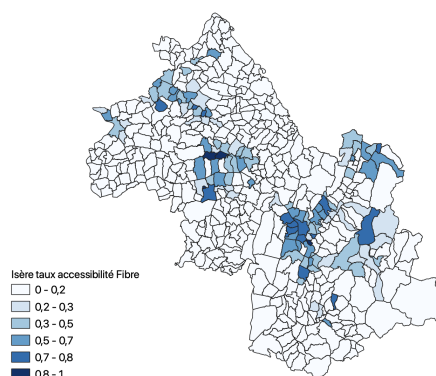
Taux d'accessibilité au Satellite des logements et locaux professionnels dans les communes de l'Isère



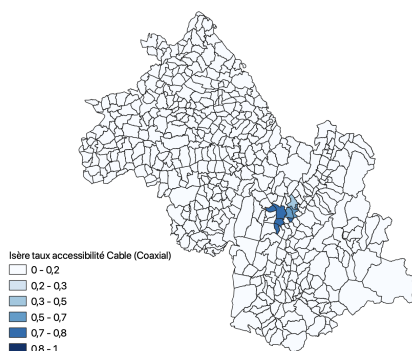
Taux d'accessibilité à la 4G des logements et locaux professionnels dans les communes de l'Isère



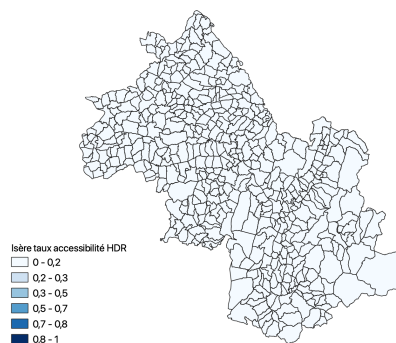
Taux d'accessibilité à la Fibre des logements et locaux professionnels dans les communes de l'Isère



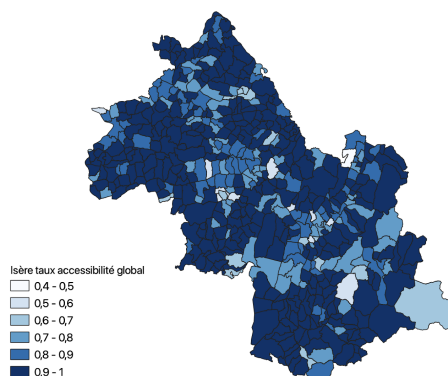
Taux d'accessibilité au Câble des logements et locaux professionnels dans les communes de l'Isère



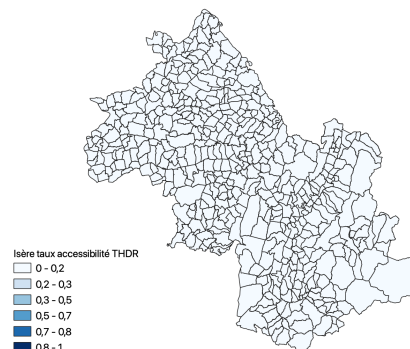
Taux d'accessibilité au HDR des logements et locaux professionnels dans les communes de l'Isère



Taux d'accessibilité global aux modes de connexion proposés dans les communes de l'Isère



Taux d'accessibilité au THDR des logements et locaux professionnels dans les communes de l'Isère



Les données par immeubles ont été quant à elles nettoyées suite à l'aggrégation en supprimant tout d'abord les champs limitation et saturation, puis en ne sélectionnant que la classe de débit descendant la plus élevée pour chaque immeuble. L'indicateur que nous allons représenter est donc cette classe de débit maximum convertie en Mbit/s.

B. Données de vote

Les données de vote, bien que complètes, présentèrent un réel challenge à leur formatage dans un format de tableau adéquat à une projection spatiale et avec les données voulues. Comme le pourcentage des résultats chaque liste, le parti arrivé en tête, l'abstention ainsi que les votes blancs nous intéressaient, une manipulation de tableau assez importante a été faite sous excel afin d'avoir des données pleinement utilisables. Aucun nettoyage n'a été à prévoir puisque ces données se doivent d'être parfaitement remplies car validées par le ministère de l'Intérieur sous la réserve du Conseil Constitutionnel.

Pour ce qui est de la projection, là aussi un travail important a été nécessaire. Au niveau départemental et communal, le principal souci a été les jointures dans Qgis avec d'autres couches comme celle des connexions internet. Seul le nom des communes étant renseigné dans les données de vote, toute jointure ne pouvait que se faire par le nom de celles-ci. Ces noms ayant été écrits par des services différents, l'orthographe des communes présentait une logique différente et par conséquent une disphase pour toutes les communes avec accents, tirets et majuscules. « Rechercher et Remplacer » d'Excel fut un outil précieux dans cette mise en conformité des noms, et les communes résistantes aux réformes furent traitées à la main.

De même, le problème s'est posé pour les bureaux de votes de Grenoble, mais avec une couche de difficulté supplémentaire. Les bureaux de vote du fonds de carte trouvé sur le site de l'agglomération de Grenoble n'étaient pas identiques aux bureaux de vote renseignés dans le jeu de données des résultats des élections. Il a donc fallu, par le biais d'indices géographiques et de noms de rues, trouver les bons bureaux de vote par déduction¹

III. Outils utilisés

Nous avons décidé d'utiliser R, Python et Qgis. Le choix de cet éparpillement provient du fait que nous sommes à l'aise en Python et également en Qgis, mais que ce projet est une bonne occasion de pratiquer R et notamment l'exploration de données tabulaires. Par manque de temps, nous avons préféré rester sur Python pour les traitements avancés et Qgis pour la visualisation.

¹ L'option « appel à un ami » ne marchait pas car ces derniers n'habitent pas dans les quartiers concernés

IV. Limites

Notre analyse aurait pu être plus poussée sur plusieurs aspects.

Tout d'abord nos jeux de données sur le vote ne contiennent que l'année 2015 et uniquement les élections régionales. Nous aurions pu faire une étude sur une période plus longue et voir l'évolution avec internet. Idéalement, nous voudrions faire une expérience naturelle, en visualisant par exemple une carte du vote FN avant et après l'arrivée d'internet. De même nous pourrions voir des variations en fonction du type d'élections.

Par manque de temps, nous n'avons 'zoomé' que sur le département de l'Isère et de Grenoble, mais il aurait pu être intéressant de représenter nos cartes en tenant compte des extrêmes. Par exemple, sur les données de connexion, en prenant la ville la plus et la moins connectée. De même sur le vote FN.

De plus, nous nous sommes servi de Python principalement, bien qu'ayant fait une tentative avec R. L'utilisation de Shiny aurait été adéquat pour explorer fracture numérique et vote en changeant des variables tel que l'année, le type d'élections, le nombre de modes de connexion disponibles, l'accessibilité aux technologies, etc.

Finalement, des données d'ordre socio-démographiques seraient sûrement très utiles pour comprendre et interpréter nos résultats, ainsi qu'une expertise en sciences humaines et sociales.