



UNIVERSITÉ LIBRE DE BRUXELLES

ULB

Faculté de Lettres, Traduction et Communication

Analyse spatiale de l'adéquation entre les implantations des établissements scolaires et la répartition de la population

Modélisation et enjeux de qualités des données

GATHOYE Matias

Mémoire présenté sous la direction de DE
VALERIOLA Sébastien en vue de l'obtention du
diplôme de Master en sciences et technologies de
l'information et de la communication.

Année académique 2022-2023

Table des matières

1	Introduction générale	3
1.1	Motivations	3
1.2	Objectifs du mémoire	4
1.3	Enjeux d'une bonne répartition	5
1.4	Analyse spatiale	5
1.5	Provenance et qualité des données	8
1.6	Plan de l'exposé	10
2	Table des matières commentées	12
3	Commentaire sur la bibliographie	17
4	Plan de travail	21

Chapitre 1

Introduction générale

1.1 Motivations

L'enseignement est pour moi une activité essentielle. Celle-ci, malheureusement, tend à être de moins en moins bien gérée par les autorités. Il n'y a qu'à observer toutes les protestations des étudiants du supérieur ou des différentes réformes scolaires qui font polémiques pour s'en rendre compte. Nous avons donc décidé de nous pencher sur le cas des implantations des écoles en Belgique. En effet, bien que l'enseignement soit pour moi important, notre domaine de recherche n'est pas en sciences de l'éducation, mais bien en sciences et technologies de l'information et de la communication. C'est pourquoi nous avons décidé d'effectuer une analyse spatiale afin d'aborder un nouvel angle sur l'accès à l'éducation.

Une autre des raisons qui nous ont poussés à nous intéresser à ce sujet fut également le fait de traiter d'analyse spatiale. En effet, ce type d'analyse peut s'avérer précieux dans le cadre de la compréhension de jeux de données, voire même indispensable pour traiter d'un problème. Nous nous souvenons encore des cartes lors de la pandémie du Covid-19 pour annoncer à la population quelles zones, quelles régions ou quels pays étaient en code rouge, orange ou jaune. Nous pourrions assimiler cette approche aux sciences géographiques, plus précisément à la géomatique, et nous pensons que ce ne serait pas totalement faux. Cette étude s'inspire et se sert des méthodes quantitatives notamment utilisées de la cadre de cette dernière discipline. Ainsi, ce mémoire est, comme le master en sciences et technologies de l'information et de la communication, interdisciplinaire. Il regroupe les domaines

et disciplines qui sont la qualité des données, bien que brièvement, la géomatique, les statistiques ainsi que la visualisation et l'analyse des données dans un sens plus large.

1.2 Objectifs du mémoire

Ce mémoire remplit ou a rempli plusieurs objectifs, qu'ils soient durant le processus de réalisation ou en termes de résultat. En effet, pour ce qui a été durant la réalisation de ce mémoire, ce travail reflète mon endurance à traiter d'un sujet spécifique et me permet de comprendre ce que c'est que de travailler pendant 2 ans sur un même thème. De plus, n'ayant pas suivi de cours à part entière sur les analyses spatiales, cette recherche m'a permis de traiter d'éléments nouveaux comme l'utilisation de R pour l'analyse spatiale, la compréhension de nouveau concept lié à la géographie et ses applications. J'ai également pu approfondir la méthode de Voronoï dans un contexte bien précis et des données actuelles.

D'un point de vue des résultats, je ne m'attends pas à ce que mon mémoire impacte fortement la méthode de construction d'école ainsi que leur choix d'implantation. Néanmoins, si par un hasard une personne venait un jour à lire mon mémoire, celui-ci pourrait y trouver un point de vue intéressant quant à la disposition des écoles en Belgique et de l'influence qu'elles jouent entre elles. Les résultats apporteraient également des idées quant à ce que permettent d'apporter comme analyse les diagrammes de Voronoï.

Ainsi, ce qui importe probablement le plus ici sont les méthodes en elles-mêmes ainsi que leurs mises en place. En effet, nous effectuons une étude sur un cas spécifique et il est intéressant de partager les méthodes utilisées pour résoudre d'autres problématiques. Ici, la difficulté du mémoire, et des méthodes de manière générale, est de travailler avec des polygones et d'effectuer des analyses dessus dans R. Ces polygones seront tantôt les contours des communes, tantôt les zones d'influence des écoles. Ces zones d'influence varieront par ailleurs en fonction de la méthode utilisée, ce qui fait que les opérations suivantes évoluent également.

1.3 Enjeux d'une bonne répartition

Je n'ai pas trouvé d'article traitant directement sur les enjeux de garder une bonne répartition de l'offre éducative. Néanmoins, nous pouvons utiliser notre bon sens et notre pragmatisme afin de comprendre les enjeux. En effet, nous pourrions partir du constat que plus une école est loin d'un individu, plus celle-ci offre un enseignement de moindre qualité qu'une école plus proche. Par exemple, entre une école A qui se trouve à 5 minutes de trajet et une école B qui se trouve à 1 heure, peu importe le moyen de déplacement, l'école A aura toujours une plus value en terme de proximité. Les écoles qui sont plus proche permettront à l'élève ou l'étudiant d'avoir un bien-être général plus élevé que ceux qui vont loin. En effet, l'individu qui va à l'école A aurait par exemple 55 minutes de sommeil en plus que s'il allait à l'école B, et ce tous les jours d'école. C'est également souhaitable pour les parents qui travaillent de n'avoir à faire qu'un petit détour de 5 minutes avant d'aller au travail. Nous pourrions même pousser l'idée encore plus loin en disant qu'une bonne répartition permettrait de diminuer le taux de décrochage scolaire. Bien que ce ne soit pas le seul facteur, avoir une école qui nous permettrait d'étudier un sujet qui nous passionne mais qui se trouve à 1 heure de route peut être très décourageant.

1.4 Analyse spatiale

Dans le cadre de ce master en sciences et technologies de l'information et de la communication (STIC), nous pourrions nous demander si réaliser un mémoire principalement sur des analyses spatiales ne serait pas hors de notre domaine. Néanmoins, les humanités numériques, qui font partie des STIC, ont des intérêts prononcés quant aux études menées avec des analyses spatiales. En effet, les humanités numériques spatialisées regroupent trois grands thèmes principaux dans lesquels l'analyse spatiale tient une place majeur¹.

Le premier cas est lors d'un traitement de l'information spatiale se trouvant dans des corpus de texte. Ceux-ci sont généralement fournis par des travaux en sciences humaines et sociales, principalement à la suite d'études littéraires. Dans les sciences humaines et sociales (SHS), les textes ont une place majeure dans bon nombre de ses

1. [6]

disciplines. Les corpus de textes sont actuellement travaillés de plus en plus grâce à des outils d'analyse qui permettent d'annoter les corpus. Il s'agit d'ajouter des couches d'informations structurées sur le texte étudié afin de pouvoir augmenter les possibilités d'analyse du dit corpus. Dans ces annotations, les éléments spatiaux et les informations géographiques sont des éléments importants pour ce processus d'analyse. Un des enjeux de la spatialisation de corpus textuels, afin de réaliser une analyse spatiale, est de comprendre comment les lieux, ainsi que les autres références spatiales sont désignés dans les domaines textuels. Dans ce cadre d'étude, il faut donc définir ce qu'est un lieu tout en prenant en compte les différents contextes d'utilisation de ces notions dans les textes. Un autre intérêt pour ce genre d'étude est qu'il permettrait de transformer les informations textuelles de localisation dans un format de coordonnées. Les objectifs des études dans le traitement de l'information spatiale dans des textes se portent notamment sur l'annotation automatique, l'évaluation des outils utilisés ainsi que les consignes d'annotation des données textuelles.

Le second thème est plutôt lié à la recherche en histoire. Il s'agit en effet de se plonger dans des problématiques d'acquisition, de spatialisation et de diffusion de données géographiques du passé et du patrimoine. Nous pouvons donc, à partir de plans anciens, obtenir de nouvelles données géographiques ou alors, à partir de sources historiques diverses, nous pouvons réaliser une analyse géohistorique du passé des villes par exemple. Ce thème est, comme beaucoup d'activité en humanité numérique, pluridisciplinaire, car des collaborations entre des historiens, des géographes et des archéologues peuvent être observées en fonction des recherches entreprises. C'est également grâce à des recherches de ce style que la représentation spatiale numérique et la gestion de base de données historiques ont vu intégrer les outils et les méthodes de la géomatique dans la recherche historique. Les systèmes d'information géographique purement historiques sont nombreux de nos jours, bien qu'il y en ait bien moins pour l'Antiquité et le Moyen Âge que pour les autres périodes. On notera notamment le programme de recherche ALPAGE du côté des Français. Ce projet, mené entre 2006 et 2010, est un des précurseurs d'une analyse géomatique et diachronique d'un espace géographique. Celui-ci avait entre autres pour but de mettre en place des outils permettant de développer les recherches sur l'espace urbain parisien grâce notamment à un système d'information géographique

comprenant des données cadastrales, mais aussi historiques. On remarque donc bien l'intérêt que les sciences historiques peuvent porter aux données géographiques et que des travaux géohistoriques seraient envisageables dans beaucoup de domaines. Nous aurions pu avoir dans notre travail cette approche géohistorique si des données géographiques sur la localisation avaient été prises durant les années précédentes afin de réaliser une étude géotemporelle et comparé ainsi comparé la situation actuelle avec la situation antérieure. Malheureusement, ces données n'existent pas ou du moins n'ont pas été dénichées, donc nous nous contenterons de traiter de la situation actuelle.

Enfin, le troisième thème est abordé par l'angle de l'archéologie, par le traitement et l'usage de l'information spatiale. L'analyse spatiale des données numériques est aujourd'hui incontournable en archéologie. Celle-ci est un outil précieux dans la production et dans l'analyse de connaissances de cette discipline. L'archéologie a en effet toujours fait un usage extensif des apports de la géographie et de la cartographie. La discipline, qui entretient des rapports étroits avec la notion d'espace dans ses études, fut donc une des premières sciences historiques à avoir entrevu les intérêts que les méthodes et les outils de la géomatique pouvaient apporter aux interprétations spatiales des bases de données. À l'origine, l'utilisation de ces outils se faisait afin de placer des artefacts ou des fouilles sur des cartes, mais l'utilisation se développa et les outils sont maintenant utilisés pour de la prospection et des analyses permettant in fine de poser des questions beaucoup plus complexes aux bases de données. Aujourd'hui, il n'est plus question d'éviter d'utiliser les techniques de géomatique en archéologie préventive.

Le lecteur attentif remarquera que notre sujet ne rentre pas vraiment dans l'un de ces trois grands thèmes. En effet, je ne compte pas analyser un quelconque corpus de texte littéraire. Il n'est pas non plus question de données historiques, ne les ayant pas trouvées et nous sommes également bien loin de traiter d'archéologie dans notre étude. Nous désirons en effet nous concentrer sur l'étape qui vient après la création des données géographiques. En effet, les trois grands thèmes abordent surtout la question de la création et de la mise en relations des données géographiques avec les différents domaines. Néanmoins, après avoir récolté ces données, il faut les analyser. Nous comptons donc nous pencher plus particulièrement sur des méthodes

de géomatique qui pourrait, à l’avenir, être transférée dans l’un des différents domaines. Nous travaillons notamment avec des polygones et des découpages de zones qui portent certaines valeurs et qu’il faudra répartir en fonction des intersections, des superpositions et de beaucoup d’autres problématiques encore. Nous utiliserons notamment un diagramme de Voronoï pour débiter notre analyse. Celui-ci a l’avantage de découper des zones à partir de points sans que celles-ci ne se superposent, ce qui nous a permis de partir d’une première approche plus simple tout en acceptant les limites de cette méthode. L’utilisation d’autres méthodes est également à venir, mais nous vous laissons les découvrir dans les chapitres suivants.

1.5 Provenance et qualité des données

Les données principales ont été trouvées sur le site data.gov.be. Ce site regroupe un ensemble de données ouvertes. Ces données ouvertes sont des données sans informations sensibles ou personnelles. Elles sont généralement dans un format facile à traiter, comme un tableau Excel, un fichier CSV ou du XML. De plus, la totalité des données est en accès gratuit pour tout le monde. ”Mais qui publie les données ?” peut-on se demander. Il s’agit en fait de données de services publics comme les villes, les régions ou les autorités fédérales, mais également de certaines entreprises comme par exemple la TEC pour la localisation de leurs arrêts de bus.

Ces données sont disponibles pour plusieurs raisons, mais il y en a 2 principales. La première étant la plus importante : c’est la loi. En effet, il existe plusieurs lois et directives quant à la réutilisation des informations du secteur public qui oblige les services publics à rendre réutilisables certaines de leurs données. La seconde raison est que les autorités elles-mêmes recherchent souvent des données. Il est donc plus facile pour eux de les partager directement en ligne.

D’un point de vue de leur utilisation, celle-ci reste très libre. Il est possible de combiner des jeux de données entre eux, de développer ses propres sites web et applications sur base de ces données et même de les utiliser pour ce que les autorités n’avaient pas prévu. Dans le cadre de notre recherche, nous combinerons plusieurs jeux de données de différentes régions, mais aussi de différents thèmes.

Les deux jeux de données sur lesquels nous nous basons dans cette étude sont

”Fichier signalétique des établissements d’enseignement de la Fédération Wallonie-Bruxelles (FASE)” et ”Onderwijsaanbod in Vlaanderen en Brussel via POI-service”. Le premier, comme son nom l’indique, est un ensemble de données signalétique des implantations scolaires en Fédération Wallonie-Bruxelles. Le second regroupe le même style de données, mais pour la Flandre et la région de Bruxelles. Vous remarquerez que la communauté germanophone n’est pas reprise dans les données. Il ne s’agit pas d’un oubli, mais bien d’un manque de jeu de données pour cette communauté et nous avons pris la décision de ne pas rechercher les données dans le cadre du présent travail. Si ultérieurement, un jeu de données venait à naître, nous serions ravis de réappliquer les méthodes utilisées plus loin en ajoutant les nouvelles informations.

Les points importants de ces jeux de données que nous retiendrons sont la géolocalisation des bâtiments et qu’il peut y avoir un même établissement scolaire regroupant plusieurs implantations. Ces implantations peuvent être de différentes localisations, mais de différents niveaux d’enseignement également. Concernant l’interopérabilité de ces deux jeux de données, nous pouvons observer que la façon de référencer les écoles est différente. Pour la fédération Wallonie-Bruxelles, les établissements et implantations scolaires sont référencés par 2 numéros appelés ”FASE”. Un pour définir l’établissement et l’autre pour l’implantation. Cependant, il n’y a pas de différence quant au type d’enseignement, nous avons donc des cas où les deux numéros FASE sont les mêmes pour deux types d’enseignement différent. Par exemple, un établissement sur une même implantation aura les mêmes numéros FASE pour la partie qui donne un enseignement maternel et que pour la partie qui donne un enseignement primaire. La combinaison des numéros FASE ne forme donc pas d’identifiants uniques. Il faudra résoudre ce problème pour obtenir des analyses de qualités. De plus, la Flandre et les écoles flamandes à Bruxelles ne se réfèrent pas de la même manière. Ceux-ci utilisent un identifiant ”POI” qui est réellement unique, même au niveau des types d’enseignement. Nous devons donc potentiellement ajuster les données afin qu’elles se coordonnent entre elles. Par ailleurs, un des jeux de données étant totalement en français et l’autre en néerlandais, nous aurons éventuellement des problèmes pour identifier les établissements du même type. Je prends l’exemple des établissements de type ”Enseignement primaire or-

dinaire” qui se traduise par ”Gewoon lager onderwijs” en néerlandais. Il pourrait également exister des types d’enseignement qui existe dans une langue et qui ne trouve pas d’équivalent dans l’autre. C’est une problématique qu’il faudra traiter d’une manière ou d’une autre en faisant des choix. Nous tenterons de tenir compte des limites des modifications et surtout de les réduire au maximum.

Autre problématique : la géolocalisation. En effet, les jeux de données utilisent différents systèmes de coordonnées. Le flamand utilise le système Lambert, nous ne savons pas exactement lequel, et le WGS84 alors que le jeu de données wallon utilise le système anglo-saxon. Nous verrons donc s’il est nécessaire de choisir l’un d’entre eux et donc au besoin de rechercher comment passer de l’un à l’autre.

1.6 Plan de l’exposé

Nous avons eu dans cette introduction un avant-goût du sujet de ce travail. La suite de ce mémoire se prolonge sur trois chapitres principaux ainsi qu’une conclusion générale.

Dans le premier chapitre, intitulé ”Les données”, nous abordons les jeux de données utilisés plus en profondeur que dans notre introduction. Nous expliquons par ailleurs quelles ont été les éventuelles modifications apportées afin de réaliser la fusion des jeux de données. Le chapitre donne également un premier aperçu statistique des données avec divers tableaux reprenant par exemple le nombre d’écoles en moyenne par province, la commune qui comprend le plus d’écoles et bien d’autres pour se faire une première idée du contenu des données. Ensuite, nous présentons d’autres jeux de données qui auraient été utilisés pour les analyses. Nous présentons notamment les données sur la répartition de la population.

Le chapitre suivant ”Méthodes” explique les méthodes employées pour effectuer les diverses analyses. La première section portera sur la façon dont la population est répartie au sein d’une commune. Est-elle répartie uniformément ou sous forme de ”foyer” pour représenter les villes et villages ? Nous expliquons donc le choix fait. Ensuite, trois sections suivront pour présenter les trois méthodes utilisées pour analyser la répartition des écoles. Chacune d’entre elles porte sur sa mise en application dans R avec le jeu de donnée fusionné sur la répartition des établissements ainsi

que sur la façon dont nous avons traité la superposition avec la répartition de la population. Cette dernière peut varier en fonction de la méthode initiale employée.

Nous présentons un dernier chapitre contenant les résultats de cette étude. Nous commençons par un constat général sur l'état de la répartition en Belgique des établissements scolaires par rapport à la population. Par la suite, nous envisageons une section spéciale pour étudier le cas de Bruxelles. La capitale a en effet l'avantage d'avoir plusieurs communes en son sein et une concentration d'école élevée. Ceci permet donc d'effectuer des analyses plus approfondies sur cette zone. Ensuite, nous comparons les différentes méthodes afin de déterminer les différents avantages que les unes emportent sur les autres. Les limites de chacune des méthodes sont également abordées dans cette section.

Enfin, après la conclusion, vous trouverez le code utilisé lors de nos recherches ainsi que des figures intéressantes, mais qui auraient pollué le texte en lui-même de par leurs nombres conséquents. Nous vous souhaitons dès lors une bonne lecture en espérant vous faire découvrir de nouveaux aspects de l'analyse spatiale.

Chapitre 2

Table des matières commentées

J’ai réalisé une première ébauche de table des matières pour structurer mon travail. Celle-ci se trouve aux deux pages suivantes. Pour ce travail, j’ai décidé d’essayer d’appliquer la structure proposée IMRaD, c’est-à-dire Introduction, Méthodes, Résultats et Discussion/Conclusion. J’ai cependant pris la liberté de rajouter un chapitre intitulé ”Données”. En effet, les données jouant un rôle important dans ce mémoire, il me semble approprié de s’arrêter durant un chapitre afin d’examiner celles-ci avant de les utiliser avec les méthodes.

Je ne vais pas trop m’attarder sur les détails du chapitre introductif, celui-ci ayant déjà été rédigé. Je signalerais cependant que celui-ci évoluera probablement. En effet, pour la section ”Enjeux d’une bonne répartition”, j’aimerais étoffer un peu plus la section, mais je n’ai pas encore trouvé d’article ou d’études reflétant mon propos. En ce qui concerne la section ”Analyse spatiale”, je suis très satisfait d’avoir pu donner des cas d’applications dans le cadre des humanités numériques, mais j’aimerais également une explication plus générale. Encore une fois, faute de référence satisfaisante, je n’ai pas réellement pu la mettre en place. Enfin, pour ”Provenance et qualité des données”, j’hésite encore à le déplacer au chapitre suivant. Néanmoins, je trouvais que réaliser une première ébauche de comment j’aborderais la problématique des données serait intéressant à mettre dans l’introduction.

Concernant la structure du chapitre ”Les données”, j’avais envisagé trois sections. La première est pour expliquer comment j’ai fusionné les jeux de données sur les établissements. En effet, comme vu dans le chapitre sur la provenance et la qualité des données, nous avons des problématiques d’interopérabilité. Je consa-

crerai donc une section afin d'expliquer les choix qui ont été faits et le résultat obtenu. Ensuite, nous donnerons un premier aperçu statistique, comme le nombre d'écoles secondaire, quelle commune/province comporte le plus d'écoles, etc. Enfin, nous listerons les autres données utilisées comme la population des communes, nous pourrions imaginer également le classement des écoles favorisée, défavorisée dans le cadre d'une analyse supplémentaire.

Le gros morceau du travail se trouvera dans les méthodes. Nous commencerons par présenter comment nous avons réparti la population au sein des communes. Est-ce que nous la répartissons de manière uniforme ou sous forme de foyer qui représente les diverses villes et villages ? Ensuite, nous expliquerons les trois méthodes utilisées. Il n'est pas certain que nous arrivions à toutes les mettre en place, mais je voulais quand même les présenter pour montrer le potentiel du mémoire. Je ne suis cependant pas satisfait des sous-sections "applications" et "Superposition des deux répartitions". En effet, je trouve intéressant de présenter la mise en applications de chacune des méthodes et de la différencier de la façon dont nous effectuons la superposition, mais je trouve la structure redondante. Il faudra que je remodèle cela légèrement, mais ce n'est rien de bien grave non plus. D'autres méthodes/analyses pourraient également être réalisées comme par exemple une analyse selon le classement des écoles favorisées et défavorisées comme mentionné plus haut.

Le dernier gros chapitre présentera les résultats. Ceux-ci seront déjà légèrement présentés à la fin de chaque méthode, mais je trouve intéressant de réaliser un constat plus général sur la situation aperçue au regard des différentes méthodes. J'envisage également un Zoom sur le cas de Bruxelles. En effet, la ville étant découpée en commune, mais aussi très dense en école, il s'agirait d'un cas d'étude à part entière intéressant à prendre en compte pour y voir plus clair. Ce chapitre se terminerait ensuite par une comparaison des méthodes mises en place ainsi que des explications concernant les limites de chacune.

Le mémoire se terminerait ensuite par une conclusion générale. J'ajouterai ensuite une annexe reprenant le code utilisé lors des diverses méthodes ou transformations et j'envisage également une annexe contenant des figures. En effet, ces dernières seront peut-être nombreuses et je veillerai à ne pas polluer le texte général d'un surplus d'images. Évidemment, bien que celle-ci ne soit pas reprise dans la

table des matières, nous retrouverons à la fin du travail la bibliographie reprenant les ouvrages, les articles et les sites consultés.

Table des matières

1	Introduction générale	4
1.1	Motivations	4
1.2	Objectifs du mémoire	5
1.3	Enjeux d'une bonne répartition	5
1.4	Analyse spatiale	6
1.5	Provenance et qualité des données	9
1.6	Plan de l'exposé	11
2	Les données	13
2.1	Fusion des jeux de données	13
2.2	Premier aperçu statistique de la fusion	13
2.3	Autres données utilisées	13
3	Méthodes	14
3.1	répartition de la population sur une commune	14
3.2	Le diagramme de Voronoï	14
3.2.1	Application	14
3.2.2	Superposition des deux répartitions	14
3.3	Distance à vol d'oiseau	15
3.3.1	Application	15
3.3.2	Superposition des deux répartitions	15
3.4	Distance en temps de trajet en voiture	15
3.4.1	Application	15
3.4.2	Superposition des deux répartitions	15

4 Résultats	16
4.1 Constat général	16
4.2 Zoom sur le cas de Bruxelles	16
4.3 Comparaison et limites des méthodes	16
5 Conclusion	17
Annexes	18
A Code	19
B Figures	20

Chapitre 3

Commentaire sur la bibliographie

Vous trouverez ci-après ma première bibliographie. Veuillez déjà excuser la mise en page de celle-ci. En effet, la note ne se met pas à la fin dans certains cas et la fin de certaines bibliographies se retrouve après le commentaire. Je n'ai malheureusement pas réussi à faire mieux avec LaTeX, mais je préférais ce format plutôt que de renvoyer chaque fois à la page suivante et donc de vous faire des aller-retour incessants.

Ma bibliographie est très orientée vers les méthodes que je vais mettre en place. On y retrouve des manuels, des articles ainsi que des explications générales des concepts. Je trouve cependant qu'il me manque des références en termes de qualité des données. En effet, ne sachant pas trop encore comment s'aborde la qualité des données, je ne savais pas vraiment vers quoi me diriger et je ne voulais pas mettre des articles juste pour remplir la bibliographie. Je pense que l'interopérabilité jouera ici un rôle important, mais je devrai m'entretenir avec madame Boydens à ce sujet pour réellement comprendre la direction que je prends. Mis à part cela, j'ai l'impression d'avoir une base bibliographique qui me permettra d'aborder les thèmes et concepts importants de ce mémoire. Je vous laisse la découvrir ci-après.

Bibliographie

- [1] Vandy BERTEN. « Géocodage : contourner les lacunes d’OpenStreetMap (partie 1) | Smals Research ». In : *Smals Research* (2020).

Ce document me permettra de me rendre compte des éventuelles contraintes que je pourrais rencontrer dans mon travail en utilisant OpenStreetMap. Je pense que je serai obligé d’utiliser OpenStreetMap malgré les limitations car il s’agit quand même d’un outil en données libres, ce qui est bien plus pratique de google maps.

- [2] Vandy BERTEN. « Géocodage : contourner les lacunes d’OpenStreetMap (partie 2) | Smals Research ». In : *Smals Research* (2020).

La suite prolonge l’explication des contraintes qui pourraient être rencontrées mais propose également des solutions efficace de géocodage.

- [3] Vandy BERTEN. « La jointure spatiale, la clé de l’analytique géographique | Smals Research ». In : *Smals Research* (2021).

Cet article explique un concept fondamental des analyses spatiales, à savoir la jointure spatiale. En effet, il est possible que certaines zones ne soient pas attribuées de façon exactes à certaines communes, comme par exemple les zones boisées ou inhabitées. L’auteur propose donc des approches afin de résoudre ces problématiques lors de divers analyses spatiales, ce qui sera très certainement pertinent dans notre cas.

- [4] Vandy BERTEN. « Peut-on toujours atteindre une maternité en 30 minutes? (partie 2) | Smals Research ». In : *Smals Research* (2021).

Pour une raison qui m’échappe, LaTeX décide d’afficher la partie 2 avant la partie 1, soit. Les deux articles ”peut-on atteindre une maternité en 30 minutes?” apporte une première approche quand à la façon dont nous pourrions calculer les distances atteignables par un véhicule partant d’un point dans un

temps donné. Dans notre étude, cela nous permettra d’obtenir une première idée sur la façon de réaliser la section ”Distance en temps de trajet en voiture”. Bien que la recherche de cet article aie été réalisé en python, il me semble tout à fait plausible de réaliser une étude similaire dans R.

- [5] Vandy BERTEN. « Peut-on toujours atteindre une maternité en 30 minutes ? | Smals Research ». In : *Smals Research* (2021).

La première partie de l’article. Les raisons de l’utilité de cet article ayant été abordée plus haut, je ne répéterai pas ce qui a déjà été dit.

- [6] Carmen BRANDO et al. « Introduction. Humanités numériques et analyses spatiales : enjeux et perspectives ». fre. In : *Humanités numériques (Lille)* 3 (2021).

Cet article fut très intéressant pour comprendre l’intérêt des analyses spatiales dans le cadre de ce master. Il m’a notamment été d’une grande utilité pour la rédaction de la section ”analyse spatiale” car il offrait vraiment un panorama des applications en humanités numériques de ce style d’analyse.

- [7] Michael DORMAN. *Introduction to Spatial Data Programming with R*.

Ce livre en ligne, qui m’a été transmit par monsieur De Valeriola, me sera très utile dans la mise en place du code du travail. Il me fournira également des explications sur les concepts et probablement les limites des modèles. Il s’agit donc d’un livre qui sera très orienté vers la partie technique de notre étude, ce qui s’avère nécessaire dans un travail d’une telle envergure.

2023. URL : <https://geobgu.xyz/r/>.

- [8] Théophile EMMANOUILIDIS. « Modélisation de répartition d’élèves et délimitation de secteurs scolaires en milieu urbain ». In : *BELGEO (Leuven)* (2017).

Ce document a pour but d’apporter une approche sur la façon dont notre étude pourrait se développer et des points importants à analyser. En effet, le titre et le contenu de cet article est étrangement ressemblant au sujet abordé dans mon mémoire. Néanmoins, cette étude a été réalisée en France, sur la ville de Lausanne plus particulièrement, nous pouvons dès lors étudier la Belgique sans souci vu que l’article ne traite que d’une ville de France alors que nous nous attaquons à notre pays tout entier (excepté la communauté germanophone).

- [9] Thierry FEUILLET, Etienne Edouard COSSART et Hadrien COMMENGES. *Manuel de géographie quantitative : concepts, outils, méthodes*. fre. Cursus.

Ce manuel a pour but de m'apporter une compréhension sur la façon dont la géographie quantitative fonctionne. Quels sont les principaux concepts à connaître ainsi que les méthodes qui peuvent être intéressante à mettre en place.

Armand Colin, 2019, p. 237-.

- [10] Marina L GAVRILOVA. *Generalized Voronoi diagram : a geometry-based approach to computational intelligence*. T. 158. Studies in computational intelligence.

Ce livre est beaucoup plus complet et généraliste quant aux diagrammes de Voronoï. Ce n'est pas sûr que je le lirai en entier mais certains chapitres pourraient être utiles quant à la compréhension de certaines formules et applications. L'introduction me sera par exemple probablement fort utile afin d'expliquer précisément ce qu'est un diagramme de Voronoï mais également d'énoncer dans quels cas de figures nous nous trouvons dans les diagrammes de Voronoï. Springer, 2008.

- [11] Thomas Fisher ROBERT GARRETT. *Voronoi Diagrams with ggvoronoi*.

Ce petit site permet d'expliquer le package ggvoronoi. Il s'agit encore de littérature technique et, bien qu'il n'est pas encore sûr que ce sera la méthode choisie, il peut être intéressant de considérer l'option avec ce petit article qui a l'air d'expliquer clairement la mise en place d'un diagram de Voronoï.

2022. URL : <https://cran.r-project.org/web/packages/ggvoronoi/vignettes/ggvoronoi.html> (visité le 12/05/2023).

Chapitre 4

Plan de travail

Durant août et septembre, je compte réaliser la partie sur les données. J'identifierai exactement les problématiques qui se présentent dans chacun des jeux de données afin de réaliser une fusion la plus parfaite possible. Je poserai donc principalement durant cette période mes éventuelles questions à madame Boydens pour identifier correctement les enjeux.

À partir de septembre/octobre, il est assez compliqué de planifier réellement un planning, car, dans la mesure du possible, je serai en train de réaliser le stage de master 2. Celui-ci, je l'imagine, occupera une bonne partie de mon temps, mais je compte bien essayer d'avancer du mieux possible dans la façon de répartir la population dans les communes. Après le stage, vers décembre, nous commencerons à mettre en place le diagramme de Voronoï. Je ne sais pas exactement évaluer le temps que cela prendra, mais je pense que le réaliser pour mi-février me semble largement plausible.

Ensuite, il me semble raisonnable de dire qu'il me faudra moins de temps pour réaliser le premier modèle vu que celui-ci se basera sur légèrement sur la structure du premier. Je pense donc que la période mi-février à mi-mars suffira au chapitre distance à vol d'oiseau. De mi-mars à mi-avril, nous réaliserons la partie distance en temps de trajet en voiture.

Enfin, je prévois deux semaines pour la rédaction des résultats et de la conclusion, peut-être trois semaines si j'ai du retard. Enfin, entre fin-avril et mi-mai, mi-mai étant la date de rendu, je compte relire le mémoire, ajuster les derniers petits détails et corriger l'orthographe.

Il est assez difficile, je trouve, de prévoir exactement des deadlines, mais de manière générale, je compte me tenir à cette base de programme qui structure dans les grandes lignes la rédaction de mon travail. Je tiens également à spécifier que je rédigerai le mémoire au fur et à mesure que j'avancerai dans mes recherches.