

Photo de couverture :

Panneau trilingue de certaines localités du Tyrol italien
avec dans l'ordre le ladin, l'allemand et l'italien
Photo : K. Bouchir

Remerciements

Je souhaiterais remercier mon Directeur de mémoire, le Professeur François Bavaud, pour son enthousiasme, sa bonne humeur, sa disponibilité et pour m'avoir accompagné durant toutes les étapes du mémoire, ainsi que le Maître d'enseignement et de recherche Isaac Pante qui a accepté de devenir expert pour ce travail. Je tiens aussi à remercier les 20 participants aux tests, qui m'ont été d'une aide certaine sans oublier ma relectrice Nadia, ainsi que Céline, Fatima et Daniela et bien d'autres qui m'ont soutenu et avec qui les séances de travail pour le mémoire et durant la maîtrise ont permis de concilier études, joie, bonne humeur et entraide. Je tiens pour finir à remercier les membres de ma famille ainsi qu'Aziza et Christelle de m'avoir soutenu tout au long de mes études.

Table des matières

1	Introduction	4
2	Méthodologie statistique	6
2.1	Les données	6
2.2	La loi de Zipf	8
2.3	Les indices liés à la courbe de Lorenz	8
2.3.1	L'indice d'Herfindahl	8
2.3.2	Le coefficient de Gini	9
2.4	Les indices de ségrégation	10
2.4.1	L'indice de ségrégation de Duncan	10
2.4.2	L'indice de dissimilarité de Duncan	10
2.4.3	L'indice de Morill	11
2.4.4	L'indice de Wong	11
2.4.5	L'indice de Gorard	12
2.5	Les autocorrélations spatiales	13
2.6	Un nouvel indice	15
3	Méthodologie visuelle	18
3.1	Visualisation de statistiques sophistiquées dans leur calcul	18
3.2	Procédure d'expérimentation	18
3.3	Choix et justification de la procédure	21
3.4	Cadre de l'expérimentation	22
3.5	Traitement des données	26
4	Résultats	27
4.1	La loi de Zipf	27
4.2	L'indice d'Herfindahl	29
4.3	Gini et Duncan : les étrangers mieux répartis que les nationaux ?	31
4.4	Duncan ou Gorard : quelles différences ?	32
4.5	Les premiers résultats spatialisés : l'indice de Morill	34
4.6	Un indice qui se veut complet (Wong)	35
4.7	La prise en compte des altitudes	36
4.8	Les indices de Moran	37
4.9	Les autocorrélations spatiales avec les cadrans de Moran	41
4.10	La visualisation comme outil éducatif	44
5	Conclusion	50
6	Annexes	51
6.1	Lexique et données	51
6.2	Liste des codes	59
7	Bibliographie	74

1 Introduction

Longtemps présente dans les diverses sociétés, la ségrégation spatiale est aujourd'hui plus que jamais au cœur des débats d'intégration. Cependant, si elle n'est visible qu'à travers la présence de communautés homogènes sur un territoire donné, elle n'indique statistiquement pas forcément une ségrégation sociale liée à l'ethnie, à la religion ou à un autre critère social, mais généralement à cause de revenus différents. Plusieurs travaux ont déjà été établis sur le sujet de la préservation des minorités linguistiques par exemple sur l'ensemble des Alpes dans leur généralité. C'est pourquoi le Tyrol italien apparaît comme un lieu idéal pour entreprendre des recherches sur la ségrégation spatiale linguistique. En effet, cette région d'Italie du Nord, composée du Trentin et du Tyrol du Sud, possède trois communautés linguistiques reconnues à savoir les germanophones, les italophones et les ladinophones.

Toutefois, bien que la démographie du Tyrol italien soit plutôt isolée, il semblerait que les italophones aient tendance à voir leur population décroître alors que dans les régions avoisinantes, ce phénomène concerne plutôt les minorités. Ernst Steinicke, professeur de géographie à l'université d'Innsbruck parle même d'assimilation des minorités dans les zones périphériques du Tyrol italien. Pour le cas du Tyrol italien, le plus surprenant est que ce soit après la Deuxième Guerre que les effectifs des communautés changent. Avant cela, leur nombre était plus ou moins stable (Steinicke et Ladins, 2001). Selon Hobsbawm, la « frontière » linguistique entre les groupes linguistiques n'a même plus varié depuis le Haut-Moyen Age (Hobsbawm, 1993). Ce qui est observé aujourd'hui, c'est qu'aucune ethnie n'augmente sa présence dans le Tyrol italien.

Cependant, les germanophones et ladinophones ont tendance à conserver leur proportion de locuteurs dans le Tyrol du Sud. En effet, si cette région est la seule qui permet la préservation de ses minorités, c'est parce qu'elle est une des seules d'Italie à bénéficier d'un statut d'autonomie en termes économiques et culturels. Celui-ci a permis de renforcer l'identité ethnique des germanophones et ladinophones. Il a aussi permis d'apaiser les tensions qu'il y avait entre l'Italie et l'Autriche quant au sort de ces minorités, l'Autriche voulant la sauvegarde des germanophones. Cependant, il demeure que la population italophone tend à décroître dans cette région. Ce phénomène pose toutefois un problème, car il est difficile à interpréter. En effet, la baisse de population peut être due soit au fait que le taux de natalité de cette ethnie est faible, soit à cause d'une émigration ou de l'usage d'une méthode statistique inappropriée pour représenter ce phénomène. Néanmoins, il semblerait qu'il y ait une forme de ségrégation spatiale au travers de laquelle les communautés linguistiques tendent à se concentrer en un lieu indépendamment du caractère rural ou urbain de l'espace. (Steinicke et Ladins, 2001).

Ainsi, l'utilisation de méthodes plus avancées que les simples statistiques descriptives, peut apporter une réponse du moins partielle aux interrogations exprimées précédemment. Comme il s'agit donc d'analyser des voisinages à l'échelle communale, les autocorrélations spatiales paraissent être l'outil le plus approprié pour analyser des territoires statistiquement divers en tenant compte de la spatialité. Ceci ne peut se faire avec de simples indices dits de « ségrégation spatiale », car la spatialité n'apparaît pas dans ces calculs. Ils permettent juste d'apprécier la ségrégation de manière purement statistique pour des territoires pouvant être positionnés aléatoirement dans l'espace.

La ségrégation spatiale pose un deuxième problème général. En partant d'une méthode basée sur

l'autocorrélation spatiale, il faut également une façon de générer une représentation permettant la compréhension directe des résultats. Beaucoup d'essais se sont faits à travers des cartes choroplèthes, ponctuelles ou encore par courbes de niveaux (Chessel et Thioulouse, 2003). Bien que la carte choroplèthe soit assez souvent utilisée comme par exemple dans le travail d'Oliveau sur l'Inde (Oliveau, 2010), il est nécessaire d'utiliser une méthode plus adaptée à la compréhension de toute personne, y compris non-académique. Ainsi, le but de cette partie serait de pouvoir générer une application de géovisualisation à des fins éducatives permettant d'instruire des personnes lambda sur des statistiques sophistiquées comme l'autocorrélation spatiale.

2 Méthodologie statistique

2.1 Les données

Etant donné le contexte politique du Tyrol italien (situation géolinguistique notamment), chacune des deux régions (Trentin et Tyrol du Sud) possède son propre organisme de collecte des données. Pour une analyse optimale, il faudrait avoir accès au niveau le plus fin du maillage de ces régions. Heureusement, la plupart des données se font à l'échelle des districts pour le Trentin et des communes pour le Haut-Adige. Comme les communes correspondent au maillage le plus fin, nous avons opté pour la recherche de ces données prioritairement. Ainsi, le *Landesinstitut für Statistik*, qui est l'office sub-régional des statistiques de Tyrol du Sud, publie des données qui vont jusqu'à l'échelle communale. Parmi elles, celles concernant la proportion de locuteurs par commune paraît la plus intéressante pour ce mémoire. Ce jeu de données est composé d'une matrice dans laquelle apparaissent les unités spatiales pour les lignes et les variables pour les colonnes. Les unités spatiales sont en fait les communes au nombre $n = 116$ et les variables correspondent à la proportion des différents groupes ethnolinguistiques au nombre de $p = 3$ pour l'allemand, l'italien et le ladin en pourcentage de présence dans la commune (Autonome Provinz Bozen SÜDTirol/Provincia autonoma di Bolzano Alto Adige-Landesinstitut für Statistik ASTAT, 2015). Seuls les citoyens italiens, c'est-à-dire les personnes possédant la nationalité italienne sont recensées dans ces données. Ceci permet de conserver le nombre de variable linguistique à 3 groupes (Autonome Provinz Bozen SÜDTirol/Provincia autonoma di Bolzano Alto Adige-Landesinstitut für Statistik ASTAT, 2015, Bevölkerung 3/popolazione 3).

Quant au nombre d'habitants total pour l'année 2011, il s'élève officiellement à $N = 511'688$. Cet indicateur ne sera pas de très grande utilité sachant que les données que nous possédons concernent les résidents possédant la nationalité italienne. Cependant, il existe des données qui prennent en compte uniquement les « Inländer » des communes, c'est-à-dire les citoyens. En règle générale, ce mot peut signifier les résidents permanents (y compris certains étrangers), mais dans ce cas-ci, seuls les nationaux sont pris en compte. Ces derniers sont au nombre de $N = 467'333$. De plus, ces données sont collectées chaque dix ans depuis l'année 1971 et montrent ainsi l'évolution linguistique des trois groupes ethniques. La base de données montre également le nombre d'étrangers par commune (Autonome Provinz Bozen SÜDTirol/Provincia autonoma di Bolzano Alto Adige-Landesinstitut für Statistik ASTAT, 2015). Dans l'ensemble, nous pouvons tenir compte de $a = 4$ communautés à savoir les germanophones, les italoophones, les ladinophones et les étrangers. Cependant, en incluant les étrangers, il devient nécessaire d'adapter les données en pourcent que nous possédons et rétablir leur proportion non plus en fonction du nombre de citoyens, mais en fonction du nombre d'habitants total. Le nombre d'étrangers, nous l'avons déjà. Quant aux autres communautés, il suffit de multiplier le taux de chacune des communautés d'une commune par le nombre de citoyen de la commune et d'arrondir les résultats. Par vérification, certaines incertitudes se sont introduites dans les résultats, mais l'erreur étant assez faible, elle peut être négligée. En effet, en vérifiant à chaque fois que la somme des personnes de chaque communauté linguistique soit égale au nombre de citoyens, nous pouvons restituer le nombre de chacun des représentants des communautés linguistiques avec une incertitude maximale de plus ou moins une personne pour les communes concernées (une quinzaine). Ceci permet de considérer les nouvelles données comme étant fiables pour les futurs usages.

Toutefois, il n'est pas possible de procéder ainsi pour les données du Trentin. En effet, pour cette

sous-région, la définition des groupes linguistiques est différente sans compter la méthode utilisée pour déclarer la présence ou non d'un groupe linguistique dans les communes, au nombre de $n = 217$. Les données, qu'offrent le *Servizio statistica della Provincia Autonoma di Trento*, ne déclarent que la présence de groupes ethniques autres que les italophones. A la différence du Tyrol du Sud qui oblige la population à déclarer la langue d'usage de chaque personne, dans le Trentin il n'y a aucune obligation de le faire. A propos de groupes ethniques, la reconnaissance d'une communauté linguistique est différente dans le Trentin. En effet, les groupes sont les suivants : Italiens, Ladins, Mochènes et Cimbres. Comme pour le Sud-Tyrol, c'est la langue qui définit l'ethnie. Toutefois, les statistiques ne parlent pas d'Allemands, mais de dialectes germanophones à savoir le mochène et le cimbre. Pour des besoins d'uniformisation des données, nous choisissons d'agrèger les données de ces deux dialectes pour reconstituer une zone germanophone. Ainsi, si dans une commune, il y a la présence de ces deux dialectes, les pourcentages seront additionnés. Par ailleurs, comme pour les données du Sud-Tyrol, seuls les résidents des communes sont pris en compte dans les statistiques linguistiques. Cependant, contrairement au Tyrol du Sud, où les étrangers sont clairement séparés des statistiques linguistiques, dans le Trentin ils sont pris en compte et ceci complique l'uniformisation des données avec le Tyrol du Sud.

En outre, les statistiques partent du principe que le Trentin est majoritairement italophone, ce qui explique l'absence de statistique sur cette population. En effet, ne sont répertoriées que les minorités régionales susmentionnées (*Servizio statistica della Provincia Autonoma di Trento*, 2014). Par conséquent, il a fallu faire quelques concessions dans l'approche de ces données. Nous sommes partis du principe que dans les données concernant les minorités linguistiques (germanophones et ladinophones), il n'y avait pas d'étrangers, car il s'agit de minorités linguistiques historiques. C'est pourquoi par relation mathématique, les étrangers se trouvent dans la population qui n'apparaît pas dans les statistiques, soit les italophones. Ainsi, en soustrayant au nombre d'habitants total les minorités linguistiques susmentionnées ainsi que le nombre d'étrangers (*Servizio statistica della Provincia Autonoma di Trento*, 2014 ; *Servizio statistica della Provincia Autonoma di Trento*, 2013), nous obtenons les italophones ayant la nationalité italienne. On pourrait penser faire l'inverse en prenant pour base les données du Trentin et modifier celles du Sud-Tyrol. Or, il demeurerait un problème quant aux étrangers, car aucune indication ne mentionne leur langue dans les données du Tyrol du Sud. C'est pourquoi nous préférons la première façon d'uniformiser. Ainsi, nous avons la variable $p = 4$ modalités pour la proportion de chaque communauté par commune (*Servizio Statistica della Provincia Autonoma di Trento*, 2014).

Le calcul de certains indices nécessitera également l'usage d'altitudes communales A_i . Pour se faire, il est nécessaire de connaître quel est le lieu défini pour représenter l'altitude communale. Heureusement, le Trentin et le Haut-Adige possèdent la même définition à savoir le siège de la commune, soit la mairie. Ces données étant issues de trois rapports comparés, elles ne nécessitent aucun réajustement (*Instituto nazionale di statistica ISTAT*, 2017 ; *Gobbi et Thurner*, 2013 ; *Servizio Statistica della Provincia Autonoma di Trento*, 2012)

Afin de pouvoir représenter les résultats, il est nécessaire d'avoir des géodonnées. Par chance, le site internet *Global administrative aeras* (*Hijmans*, 2009), publie gratuitement des fichiers *shape* de tous les pays avec diverses échelles de maillages. Ainsi, les données géographiques des communes tyroliennes sont accessibles sans avoir à les créer nous-mêmes. Il demeure cependant juste quelques modifications à faire. En effet, comme le site date de 2009, certaines communes ont fusionné. C'est pourquoi deux

fusions de communes n'ont pas été prises en compte dans le maillage à savoir celle de Ledro (composée de Bezzacca, Pieve di Ledro, Concei, Molina di Ledro, Tiarno di Sopra et Tiarno di Sotto) et celle de Comano Terme (composée de Bleggio Inferiore et de Lomaso). Cette fusion ne s'applique donc que pour les données statistiques de 2011. Toutefois, nous pouvons également nous servir des nouvelles géodonnées mises en place par l'*Instituto nazionale di statistica* (ISTAT) (2017, 11 janvier). Ces données mises en ligne au mois d'avril 2017 possèdent l'avantage de s'adapter à l'année désirée et intègrent déjà les valeurs des périmètres et des surfaces. Grâce à ArcGIS et l'option «intersecter» que propose le programme, il est possible de sortir un fichier ne contenant que les limites des communes. Ces dernières peuvent ensuite être introduites dans le logiciel QGIS qui propose de calculer la longueur de ces frontières grâce à un des algorithmes du logiciel. Ainsi, les longueurs interfrontalières, le périmètre et la surface de chaque commune serviront de données pour les calculs.

2.2 La loi de Zipf

La loi de Zipf également connue sous le nom de «loi rang-taille» servait initialement à la classification de mots par fréquence. Il s'agissait d'un travail de linguistique. Cependant, en reprenant cette loi, il serait possible de classer les communes tyroliennes par leur population. La loi consiste à observer une tendance statistique au sein d'une distribution. Ainsi, en classant les communes dont une communauté linguistique est la plus présente vers celles où elle est la moins présente, nous obtenons un graphique, dont la pente permet d'apprécier la rigueur de la distribution. La formule se présente de la manière suivante (Kirby, 1985) :

$$F_{ia} = \frac{N_{ia}}{r_{ia}^k} \quad (1)$$

où

N_{ia} = population de la communauté a dans la commune i

r_{ia} = rang de la commune déterminé par l'ordre décroissant des communautés a des communes i

k = constante de la fonction

F_{ia} = fréquence de la commune i pour la communauté a

Afin de pouvoir déterminer ce paramètre k , il est nécessaire de prendre le logarithme. C'est pourquoi le graphique final se présentera avec une échelle logarithmique allant de 0 à 1 pour la population, soit pour l'axe des ordonnées. La formule prendra alors la forme suivante :

$$\log_{10}(F_{ia}) = k \cdot \log_{10}\left(\frac{N_{ia}}{r_{ia}}\right) \quad \text{ou} \quad k = \frac{\log_{10}(F_{ia})}{\log_{10}\left(\frac{N_{ia}}{r_{ia}}\right)} = \frac{\log_{10}(F_{ia})}{\log_{10}(N_{ia}) - \log_{10}(r_{ia})} \quad (2)$$

A partir de là, nous avons une idée générale de la distribution des communautés dans le Tyrol italien. Les disparités exposées à la suite des opérations mathématiques aideront à l'interprétation des résultats des indices qui suivront.

2.3 Les indices liés à la courbe de Lorenz

2.3.1 L'indice d'Herfindahl

Dans la suite des analyses, il est nécessaire d'avoir plus d'indices pour caractériser le phénomène de ségrégation spatiale. Ainsi, l'indice d'Herfindahl peut être utilisé à cette fin pour autant que ses paramètres

soient redéfinis. En effet, à la base, l'indice d'Herfindahl sert à représenter la concentration des activités économiques d'un secteur sur un territoire donné. En modifiant et en adaptant la formule, nous obtenons les équations suivantes (Houdebine, 1999) :

$$\text{Pour chaque commune : } H_i^* = \frac{H_i - 1/m}{1 - 1/m} \quad \text{avec} \quad H_i = \sum_{a=1}^m p_{ia}^2 \quad \text{où} \quad p_{ia} = \frac{N_{ia}}{N_{i\bullet}} \quad (3)$$

$$\text{Pour le Tyrol italien : } H^* = \frac{H - 1/m}{1 - 1/m} \quad \text{avec} \quad H = \sum_{a=1}^m p_{\bullet a}^2 \quad \text{où} \quad p_{\bullet a} = \frac{N_{\bullet a}}{N_{\bullet\bullet}} \quad (4)$$

où

m = nombre de communautés linguistiques total (ici 4)

$N_{\bullet\bullet}$ = population totale dans le Tyrol italien

$N_{\bullet a}$ = population totale de la communauté a au sein du Tyrol italien.

$N_{i\bullet}$ = population totale dans la commune i

p_{ia} = proportion de locuteurs de la langue a dans la commune i

$p_{\bullet a}$ = proportion de locuteurs de la langue a dans le Tyrol italien

H = indice de Herfindahl normal pour le Tyrol italien

H_i = indice de Herfindahl normal par commune

H^* = indice de Herfindahl normalisé pour le Tyrol italien

H_i^* = indice de Herfindahl normalisé par commune

Cette formule peut ainsi s'appliquer à la fois au Tyrol italien et à toutes les communes. Comme elle est normalisée, le résultat de l'indice reste compris dans l'intervalle $[0; 1]$ avec une faible concentration lorsque la valeur est proche de 0 et une forte lorsqu'elle est proche de 1 (Houdebine, 1999). Sans normalisation, le résultat serait compris entre $1/m$ et 1, c'est-à-dire $H_i \in [1/m; 1]$, puisque $\sum_a p_{ia} = 1$. Par conséquent, le résultat le plus faible que l'on puisse obtenir correspond à la situation dans laquelle chaque communauté a aurait la même proportion de représentants. Ceci correspond à l'indice de Herfindahl normal divisé par le nombre de communautés que comporte l'unité spatiale, soit H/m . A l'inverse, plus p_{ia} est proche de 1, plus H sera proche de 1.

2.3.2 Le coefficient de Gini

Par la suite, d'autres coefficients peuvent permettre de mettre en lumière la ségrégation spatiale. Le coefficient de Gini, très utilisé pour montrer l'inégalité de revenus entre les sexes, peut également s'appliquer à l'analyse de la ségrégation spatiale liée à la langue parlée. L'indice de Gini est compris entre 0 et 1. Plus le coefficient est proche de 0, plus la commune est égalitaire. A l'inverse, plus il est proche de 1 et plus elle est inégalitaire. Dans le cas du Tyrol italien, il s'agira d'apprécier la concentration de chaque communauté linguistique en sachant que plus l'indice est proche de 0 et plus la communauté linguistique est dispersée (Houdebine, 1999). Pour calculer ce nouvel indice, il est nécessaire d'avoir deux variables. La première est le nombre d'habitants par commune N_i . La seconde correspond au nombre N_{ia} d'une communauté a dans la commune i . Le coefficient de Gini se traduit de la manière suivante (Apparicio, 2000) :

$$G_a = \sum_{i=1}^n \sum_{j=1}^n \frac{N_i N_j |p_{ia} - p_{ja}|}{2N^2 \bar{p}_a (1 - \bar{p}_a)} \quad (5)$$

où

N_i = population totale de la commune i

N_j = population totale de la commune j

p_{ia} = proportion de la communauté a dans la commune i

p_{ja} = proportion de la communauté a dans la commune j

N = population totale dans le Tyrol italien

\bar{p}_a = proportion de la communauté a dans le Tyrol italien

G_a = coefficient de Gini pour la communauté a

2.4 Les indices de ségrégation

2.4.1 L'indice de ségrégation de Duncan

Maintenant que les indices de Herfindahl et de Gini sont explicités, les différents indices de ségrégation spatiale à proprement dits peuvent être introduits. D'abord, nous commencerons par évoquer les indices de Duncan. En effet, ce chercheur américain a essayé de caractériser la ségrégation spatiale entre les populations blanche et non-blanche dans les quartiers de villes américaines. Son indice de ségrégation possède l'avantage de pouvoir s'adapter à n'importe quelle échelle spatiale, car il s'agit d'un indice global. L'idée serait de pouvoir l'appliquer pour chacune des $p = 4$ communautés linguistiques à l'échelle communale et comparer vis-à-vis de la région entière du Tyrol italien. En outre, le résultat de l'indice est toujours compris entre 0 (dispersion spatiale maximale) et 1 (concentration spatiale maximale) La formule de l'indice de ségrégation spatiale s'exprime de la façon suivante (Duncan et Duncan, 1955 ; Office fédéral de la statistique, 2016) :

$$IS_a = \sum_{i=1}^n \frac{N_i |p_{ia} - \bar{p}_a|}{2N\bar{p}_a(1 - \bar{p}_a)} \quad \text{avec} \quad N = \sum_{i=1}^n N_i \quad (6)$$

où

N_i = population totale dans la commune i

N = population totale dans le Tyrol italien (= $N_{\bullet\bullet}$)

p_{ia} = proportion de la communauté linguistique a dans la commune i (= $\frac{N_{ia}}{N_{i\bullet}}$)

\bar{p}_a = proportion de la communauté linguistique a au niveau du Tyrol italien (= $\frac{N_{a\bullet}}{N_{\bullet\bullet}}$)

n = nombre de communes soit du Tyrol italien

IS_a = indice de ségrégation de Duncan

2.4.2 L'indice de dissimilarité de Duncan

Cet indice de ségrégation spatiale a l'avantage de pouvoir comparer la «marginalisation» générale en fonction de l'échelle choisie. Cependant, il n'y a aucune comparaison dans la formule entre les diverses communautés linguistiques. C'est pourquoi nous introduisons également l'indice de dissimilarité de Duncan. Il doit permettre de donner une idée de la présence de «marginalisation» entre deux communautés. Comme pour l'indice de ségrégation, son résultat est compris entre 0 et 1 avec la même signification. L'indice de dissimilarité s'articule de la manière suivante (Duncan et Duncan, 1955 ; Apparicio, 2000) :

$$ID_{ab} = \frac{1}{2} \sum_{i=1}^n \left| \frac{N_{ia}}{N_a} - \frac{N_{ib}}{N_b} \right| \quad (7)$$

où

N_{ia} = population de la communauté a de la commune i

N_a = population de la communauté a totale

N_{ib} = population de la communauté b de la commune i

N_b = population de la communauté b totale

ID_{ab} = indice de dissimilarité de Duncan

2.4.3 L'indice de Morill

L'indice de Duncan offre la possibilité d'apprécier la ségrégation spatiale et la dissimilarité spatiale, mais sans toutefois avoir de dimension spatiale. Ainsi, peu importe la place géographique qu'occupe un territoire (ici une commune), l'indice ne changera pas de valeur si ce territoire devait «changer de lieu». C'est pourquoi cette thématique a occupé beaucoup de géographes pour finalement donner lieu à un indice de Duncan spatialisé. Il s'agit de l'indice de Morill. La spatialisation de l'indice de ségrégation de Duncan réside dans le fait que les communes adjacentes sont prises en compte avec une valeur de $a_{ij} = 1$ s'il y a contiguïté spatiale et $a_{ij} = 0$ si elle est inexistante. Cet indice, comme celui de Duncan, est compris entre 0 et 1. La formule s'écrit de la manière suivante (Morill, 1991 ; Apparicio, 2000) :

$$ISM_a = IS_a - \frac{\sum_{i=1}^n \sum_{j=1}^n |a_{ij}(p_{ia} - p_{ja})|}{\sum_{i=1}^n \sum_{j=1}^n a_{ij}} \quad (8)$$

où

IS_a = indice de ségrégation de Duncan

a_{ij} = matrice d'adjacence

p_{ia} = proportion d'une communauté a dans la commune i

p_{ja} = proportion d'une communauté a dans une autre commune j

ISM_a = indice de ségrégation de Morill pour la communauté a

2.4.4 L'indice de Wong

Avec l'indice de Morill, le paramètre de spatialité a_{ij} a pu être introduit. Cependant, il ne tient compte que de l'existence d'une frontière commune entre deux unités spatiales. Ainsi, le résultat ne peut être que binaire : soit il y a contiguïté spatiale et par conséquent une valeur d'indice pour Morill, soit la partie de l'équation qui suit la soustraction est nulle et ainsi l'indice de Morill égale à l'indice de Duncan. Dans la lignée des recherches spatiales, certains géographes n'étaient pas satisfaits de l'indice de Morill, qu'ils jugeaient trop simple. Parmi eux, David Wong était d'avis qu'il était nécessaire de préciser les paramètres frontaliers des unités spatiales, car pour lui ce n'était pas seulement la présence de contiguïté spatiale qui importait, mais bien la taille de celle-ci. En effet, si toutes les frontières ont la même pondération, les interactions spatiales entre les communes sont considérées comme similaires. Or, ces interactions spatiales, qui sont au centre de la recherche de Wong, dépendent de la taille de la frontière entre les communes, car plus une frontière est grande et plus les interactions spatiales sont importantes. C'est pourquoi il propose, en plus de tenir compte des apports de Duncan et de Morill, d'ajouter trois paramètres à savoir la longueur des frontières intercommunales l_{ij} , le périmètre des communes concernées P_i/P_j et la surface de ces communes S_i/S_j . Par ailleurs, pour Wong, il était nécessaire de prendre également en considération la ségrégation spatiale non pas entre deux groupes, mais entre les différentes ethnies. C'est pourquoi il est possible de considérer son indice, bien qu'issu de la lignée des précédents, comme

complet. L'indice se présente de la façon suivante (Wong, 1993 ; Wong et Chong, 1998) :

$$ISW_a = IS_a - \frac{1}{2} \sum_i \sum_j L_{ij} |p_{ia} - p_{ja}| \cdot \frac{\frac{P_i}{S_i} + \frac{P_j}{S_j}}{2 \cdot \max(\frac{P}{S})} \quad \text{avec} \quad L_{ij} = \frac{l_{ij}}{P_j} \quad (9)$$

où

IS_a = indice de ségrégation de Duncan

l_{ij} = longueur de la frontière entre deux communes

L_{ij} = longueur de la frontière entre deux communes l_{ij} divisé par le périmètre de la commune P_j

p_{ia} = proportion d'une communauté a dans la commune i

p_{ja} = proportion d'une communauté a dans la commune j

P_i = périmètre de la commune i

P_j = périmètre de la commune j

S_i = surface de la commune i

S_j = surface de la commune j

$\max(\frac{P}{S})$ = ratio périmètre/surface maximal dans le Tyrol italien

ISW_a = indice de ségrégation de Wong pour la communauté a

2.4.5 L'indice de Gorard

Toujours dans la lignée de Duncan, un autre indice a trouvé son importance, cette fois-ci au début du XXI^e siècle. Il s'agit de l'indice de Gorard. Ce chercheur s'est basé sur l'indice de dissimilarité de Duncan ID_a et a tenté d'en exprimer les avantages et les inconvénients. L'indice de Duncan offrait en effet la possibilité de n'être lié à aucune forme de configuration de la population et cela permet de comparer entre elles différentes études faites par ce moyen. Cependant, trois critiques obligent à repenser la manière d'organiser l'indice. Premièrement, Duncan ne prend pas en compte toutes les interactions entre les unités spatiales (critique faite également par Morill et Wong). Deuxièmement, la taille des minorités ethniques doit avoir une certaine importance vis-à-vis du nombre d'unités spatiales. Troisièmement, les études comparatives sur d'autres territoires avec d'autres échelles demeurent impossibles aussi longtemps que la composition ethnique change dans le temps et l'espace. Gorard propose donc un nouvel indice qui serait la simple différence entre la proportion d'une communauté a d'une commune i par rapport à cette communauté a au niveau du Tyrol italien et la proportion de la population totale d'une commune i par rapport à la population totale du Tyrol italien. La formule s'écrit de la façon suivante (Gorard et Taylor, 2002) :

$$ISG_a = \frac{1}{2} \sum_{i=1}^n \left| \frac{N_{ia}}{N_a} - \frac{N_i}{N} \right| \quad (10)$$

où

N_{ia} = population de la communauté a dans la commune i

N_a = population de la communauté a dans le Tyrol italien

N_i = population totale de la commune i

N = population totale du Tyrol italien

ISG_a = indice de ségrégation de Gorard pour la communauté a

2.5 Les autocorrélations spatiales

Avec les indices de ségrégation spatiale, nous pouvons déjà avoir une représentation générale du phénomène social entre les communautés linguistiques. A présent, il faut observer la similarité du phénomène entre les communes, c'est-à-dire l'homogénéité linguistique des communes. C'est pourquoi, afin de pouvoir réaliser cette étude comparative, les mesures d'autocorrélation spatiale vont être utilisées. Cette dernière représente justement la corrélation d'un phénomène des régions (ici communes). Plus la corrélation est élevée et plus les communes sont voisines. Une absence d'indépendance entre les unités spatiales caractérise l'autocorrélation spatiale (Le Gallo, 2002). Afin de quantifier celle-ci, nous allons utiliser l'indice de Moran ou I de Moran. L'indice de Geary ou la semi-variance peuvent également être utilisés à cette fin. Toutefois, l'avantage du I de Moran est qu'il est plus répandu dans la littérature et certains logiciels, notamment en statistique, géostatistique et système d'information géographique, proposent même de le calculer. Il est aussi considéré comme plus fiable de par la robustesse de sa formule qui implique la moyenne. En effet, cette dernière peut éliminer les valeurs aberrantes lorsqu'elle est calculée. La formule du I de Moran se présente de la manière suivante :

$$I = \frac{\sum_i \sum_j e_{ij} (p_{ia} - \bar{p}_a)(p_{ja} - \bar{p}_a)}{\sum_i f_i (p_{ia} - \bar{p}_a)^2} \quad (11)$$

où

p_{ia} = proportion de locuteurs de la langue a dans la commune i

i = commune de référence

j = communes voisines du point i , définies par la matrice e_{ij}

e_{ij} = matrice d'échange, soit la matrice d'échange pondérée avec $e_{i\bullet} = f_i$ (prenant en compte la pondération f soit la population, et la matrice d'adjacence a_{ij} entre les communes)

\bar{p}_a proportion de la communauté linguistique a au niveau du Tyrol italien

I = indice de Moran global

Comme dans toute statistique qui se respecte, un test de significativité est nécessaire pour vérifier que les résultats soient attribuables à une simple fluctuation ou non. La fonction `delta_z` écrite dans le code R (placé dans les annexes) permet de calculer l'indice d'autocorrélation spatiale δ et la variable centrée réduite correspondante. Cette dernière fera office de test de significativité. Pour une marge d'erreur de $\alpha = 5\%$, la valeur du z-score z doit être supérieure à $1,96 = u_{0,975}$ pour être significative. En partant du principe que l'hypothèse H_0 correspond à une absence d'autocorrélation spatiale, si $|z| \geq u_{1-\alpha/2}$, alors H_0 est à rejeter (voir la définition de la variable centrée réduite sur (Bavaud, 2014)).

A partir de là, nous avons une mesure globale. En effet, le I de Moran ne permet pas de mesurer ce qui se passe localement pour chaque unité spatiale. C'est pourquoi Luc Anselin (1995), professeur à l'Université d'Etat de l'Ohio, a proposé des indicateurs locaux nommés LISA pour « indicateurs locaux d'association spatiale ». Par construction, le I de Moran local I_i prend ainsi la forme suivante (Oliveau, 2005) :

$$I_i = \delta_i = \frac{1}{f_i} \frac{\sum_j e_{ij} (p_{ia} - \bar{p}_a)(p_{ja} - \bar{p}_a)}{\sum_j f_j (p_{ja} - \bar{p}_a)^2} \quad (12)$$

où

p_{ia} = proportion de locuteurs de la langue a dans la commune i

i = commune de référence

j = communes voisines du point i , définies par la matrice e_{ij}

e_{ij} = matrice d'échange pondérée avec $e_{i\bullet} = f_i$ (prenant en compte la pondération f_i soit la population, et la matrice d'adjacence a_{ij} entre les communes)

f_i = pondération par la population de la commune i

f_j = pondération par la population de la commune j

\bar{p}_a = proportion de la communauté linguistique a au niveau du Tyrol italien

$I_i = \delta_i$ = Indice de Moran local

Cette mesure va permettre de connaître l'autocorrélation spatiale au niveau communale. Son résultat est normalement compris entre -1 et 1 :

1. Si l'intervalle correspond à $[-1; 0[$, l'autocorrélation spatiale est négative. Cela signifie que pour chaque variable géographique ayant une localisation voisine d'une unité géographique, la variable aléatoire affiche une valeur moins proche de celle entre deux champs de deux régions.
2. Si l'auto-corrélation vaut 0, elle est inexistante. Cela signifie que la répartition géographique d'une variable est entièrement due au hasard.
3. Lorsque l'intervalle est compris entre $]0; 1]$, l'autocorrélation spatiale est positive. Ceci se remarque par le fait que les valeurs des champs de deux régions sont plus proches que la valeur de deux régions choisies aléatoirement (Le Gallo, 2002).

Toutefois, il se peut, suivant les définitions adoptées, que l'autocorrélation spatiale dépasse l'intervalle $[-1; 1]$, ce qui fait que cette étendue n'est pas exhaustive. Par ailleurs, le I de Moran montre une absence d'autocorrélation spatiale lorsqu'il prend la valeur $\frac{-1}{(n-1)}$, parce que ce résultat correspond à la valeur attendue sous l'hypothèse d'absence d'autocorrélation (selon les définitions) (Oliveau, 2005). Par conséquent, le mieux serait d'adapter un peu la définition. Gardons le fait que $I_{\bullet} = -(n-1)^{-1}$ correspond à une absence de autocorrélation spatiale s'il n'y a pas de composante diagonale (c'est-à-dire $e_{ii} = 0$) et donc à une répartition aléatoire. A partir de là, les valeurs supérieures à cette formule montrent une corrélation spatiale positive (concentration des valeurs) et celles inférieures à une corrélation spatiale négative (dispersion des valeurs) (Aubry et Piégay, 2001).

Si l'autocorrélation spatiale est quantifiable, interpréter son existence est une chose différente. En effet, il existe deux raisons principales qui expliquent la présence d'autocorrélation spatiale :

1. Il existe des interactions entre deux espaces. Elles peuvent être de différentes natures : économiques, financières, liées au transport de biens et de personnes. Ces interactions spatiales ont en général une incidence très forte sur le résultat statistique des autocorrélations spatiales.
2. Elles peuvent aussi provenir d'erreurs soit dans les bases de données, soit de variables jugées trop corrélées omises. Dans ce cas, la méthode des autocorrélations spatiales peut permettre de déceler des défauts dans un modèle choisi. (Le Gallo, 2002)

En outre, il faut tout de même noter la différence entre le I de Moran local δ_i et l'indice de ségrégation de Duncan IS_a . En effet, le premier est une mesure spatiale à proprement dite dans le sens où il est possible d'obtenir des scores qui pourront être utilisés sous forme de coordonnées pour faire un diagramme de Moran par exemple. Il s'agit donc bien d'une mesure d'autocorrélation spatiale. L'indice de ségrégation de

Duncan IS_a n'a pas cette caractéristique spatiale, il permet seulement d'apprécier la ségrégation dans une commune par rapport à l'ensemble de la région. Il montrera donc la variabilité de locuteurs par commune. Il est également possible de déterminer comment pourrait évoluer la situation linguistique des communes du Tyrol du Sud à travers les chaînes de Markov. Cette méthode peut donner un résultat sur la proportion de locuteurs par commune dans l'avenir sur la base de la moyenne des tendances des cinq années de mesures sur un intervalle de cinquante ans (de 1971 à 2011). Nous aurions également pu utiliser les séries temporelles pour représenter le même phénomène, mais il aurait fallu avoir des observations sur cinquante voire cent ans en ayant des données pour toutes les années. Par conséquent et au vu du faible nombre d'années à disposition, ni les séries temporelles, ni les chaînes de Markov (temporel) ne seront prises en compte dans ce travail.

Par ailleurs, tous les indices précédents les autocorrélations spatiales (Herfindahl, Duncan, Morill et Wong) ainsi que tous ceux qui suivront ne nécessiteront pas de tests statistiques. En effet, ces derniers comparent des éléments statistiques (moyenne, pourcentages, etc.) sur un échantillon donné. Or, au vu des données, un test n'est pas nécessaire, car nous travaillons avec des populations entières et non pas avec un échantillon représentatif. Il n'y a par conséquent aucune fluctuation aléatoire qui pourrait être due à l'adéquation entre un échantillon représentatif et une population totale (Jougla, 1997). Les autocorrélations spatiales, elles, sont déterminées par le I de Moran qui «compare l'écart de la valeur d'un individu à la moyenne à la valeur des voisins de ces individus à la moyenne» (Oliveau, 2014). De plus, comme les autocorrélations spatiales sont dépendantes de l'âge du réseau t (Bavaud, 2014), qui correspond au degré de voisinage pris en compte (à $t = 1$, ce sont les communes limitrophes de la commune étudiée) et peut posséder plusieurs valeurs, un facteur aléatoire est ajouté et nécessite l'utilisation d'un test sous la forme de variables centrées réduites pour vérifier la significativité du degré de voisinage choisi.

2.6 Un nouvel indice

Il est intéressant de voir comment chacun des indices spatiaux caractérise les résultats. Cependant, bien qu'ils donnent beaucoup d'informations pour caractériser l'espace linguistique du Tyrol italien, il n'en demeure pas moins qu'une dimension n'a pas été prise en compte à savoir l'altitude. Pour ce faire, nous décidons de partir des progrès liés aux indices de Wong et de Gorard. En effet, la série de critiques que chaque chercheur a faite à son prédécesseur était justifiée. Cependant, chacun d'eux partait de l'indice de Duncan qui a donc été une base essentielle de développement de formule. Dans l'idée de caractériser le territoire de manière encore plus précise, il est nécessaire de prendre en compte, à la fois un maximum de paramètres de géographie physique et les associer aux paramètres de géographie humaine. C'est pourquoi il est nécessaire de prendre en compte le périmètre des communes, leur surface, mais également leur altitude, car elle peut limiter voire empêcher la communication de deux communes adjacentes. Évidemment, elle ne pourra sans doute pas prendre en compte toute les spécificités communales comme par exemple certaines routes principales essentielles qui permettent de faire communiquer des communes inaccessibles sans la main de l'homme à travers des tunnels ou des cols. Cependant, cette variable simple et très représentative d'un territoire a souvent été délaissée, pour ne pas dire oubliée, dans les divers travaux de géographie en général. Par ailleurs, nous pensons que l'indice de Gorard devrait remplacer celui de Duncan dans la formule de Wong, car ce n'est pas une différence entre deux groupes qui importe, mais bien d'observer le comportement des quatre communautés. Pour écrire la formule, le plus simple serait d'intégrer l'altitude dans le rapport entre le périmètre et la surface. En effet, du point de vue des unités mathématiques, cela permettrait d'absorber la présence d'une unité en l'occurrence les mètres par mètres

carrés.

$$ISKB_a = ISG_a - \frac{1}{2} \sum_i \sum_j L_{ij} |p_{ia} - p_{ja}| \cdot \frac{\frac{P_i \cdot A_i}{S_i} + \frac{P_j \cdot A_j}{S_j}}{2 \cdot \max(\frac{P \cdot A}{S})} \quad \text{avec} \quad L_{ij} = \frac{l_{ij}}{P_j} \quad (13)$$

où

ISG_a = indice de ségrégation de Gorard

l_{ij} = longueur de la frontière entre deux communes

L_{ij} = longueur de la frontière entre deux communes l_{ij} divisé par le périmètre de la commune P_j

p_{ia} = proportion d'une communauté a dans la commune i

p_{ja} = proportion d'une communauté a dans la commune j

P_i = périmètre de la commune i

P_j = périmètre de la commune j

S_i = surface de la commune i

S_j = surface de la commune j

A_i = altitude de la commune i

A_j = altitude de la commune j

$\max(\frac{P}{S})$ = ratio périmètre/surface maximal dans le Tyrol italien

$\max(\frac{P \cdot A}{S})$ = ratio périmètre*altitude/surface maximal dans le Tyrol italien

$ISKB_a$ = indice de ségrégation de Wong tenant compte des altitudes pour la communauté a

Cependant, l'ajout de l'altitude peut également se faire en reprenant la même forme que pour l'indice de compacité ajouté par Wong (1993) dans sa formule. Il s'agit d'un rapport dans lequel le numérateur contient la somme des altitudes des communes adjacentes et un dénominateur qui ne tient compte que de l'altitude maximale dans la région du Tyrol. Cependant, étant donné que cette fraction s'ajoute aux membres précédents, il demeure un risque de faible immuabilité vis-à-vis de l'indice de Gorard. Ce risque peut également apparaître pour les formules qui suivent.

$$ISKB_a = ISG_a - \frac{1}{2} \sum_i \sum_j L_{ij} |p_{ia} - p_{ja}| \cdot \frac{\frac{P_i}{S_i} + \frac{P_j}{S_j}}{2 \cdot \max(\frac{P}{S})} \cdot \frac{A_i + A_j}{2 \cdot \max(A)} \quad \text{avec} \quad L_{ij} = \frac{l_{ij}}{P_j} \quad (14)$$

Toutefois, une autre façon de construire l'indice peut être encore plus pertinente. Basé sur la soustraction à 1 des altitudes de deux communes $A_i - A_j$ par rapport à leur addition $A_i + A_j$, c'est-à-dire $1 - \frac{A_i - A_j}{A_i + A_j}$, il permettrait également de bien caractériser le phénomène et posséderait l'avantage par simplification d'avoir un numérateur ne dépendant que de l'altitude adjacente. Cependant, cet élément de l'indice et plus précisément son dénominateur pose problème si les deux altitudes des communes valent chacune zéro, même si dans la pratique, les chances pour que deux communes valent zéro pile soient très peu probables dans le monde et le sont encore moins au Tyrol italien.

$$ISKB_a = ISG_a - \frac{1}{2} \sum_i \sum_j L_{ij} |p_{ia} - p_{ja}| \cdot \frac{\frac{P_i}{S_i} + \frac{P_j}{S_j}}{2 \cdot \max(\frac{P}{S})} \cdot \frac{2 \cdot A_j}{A_i + A_j} \quad \text{avec} \quad L_{ij} = \frac{l_{ij}}{P_j} \quad (15)$$

C'est justement pour pallier à ce problème qu'une ultime mise en forme de l'indice sera mentionnée. L'idée d'un numérateur constitué de la soustraction des altitudes des deux communes adjacentes se maintient. Cependant, le dénominateur deviendrait la valeur maximale de la soustraction (en valeur absolue)

de deux communes adjacentes et serait invariant par rapport à la définition de l'altitude 0.

$$ISKB_a = ISG_a - \frac{1}{2} \sum_i \sum_j L_{ij} |p_{ia} - p_{ja}| \cdot \frac{\frac{P_i}{S_i} + \frac{P_j}{S_j}}{2 \cdot \max(\frac{P}{S})} \cdot \frac{|A_i - A_j|}{\max_{ij} |A_i - A_j|} \quad \text{avec} \quad L_{ij} = \frac{l_{ij}}{P_j} \quad (16)$$

3 Méthodologie visuelle

3.1 Visualisation de statistiques sophistiquées dans leur calcul

De nombreux travaux en statistiques spatiales ont utilisé les autocorrélations spatiales afin d'enrichir l'analyse des données. En général, deux formes de visualisation ressortent quasi systématiquement dans la littérature à savoir une cartographie choroplèthe des I de Moran locaux ainsi que celle des diagrammes de Moran, plus répandues. Or, étant donné qu'il s'agit d'indicateurs sophistiqués dans leur calcul, il n'est pas toujours évident d'en saisir toute la subtilité, du moins il n'est pas donné à tout le monde de pouvoir en comprendre l'essentiel. C'est pourquoi la question de fond de la visualisation se base bien plus sur la compréhension directe des autocorrélations spatiales que sur une éventuelle amélioration esthétique des représentations déjà existantes. Pour arriver à un tel résultat, nous prétendons dès le départ que la visualisation des statistiques sophistiquées est accessible à tout public, et en particulier aux personnes sans connaissance des autocorrélations spatiales. Pour ce faire, deux visualisations seront générées et accessibles sur le compte «Niwetsch», répertoire nous appartenant et dont les productions sont nôtres, du site Github. La première (Niwetsch, 2018) contient une carte représentant des autocorrélations spatiales locales (pour les germanophones, italophones, ladinophones et étrangers) dans le Tyrol italien ainsi que la proportion d'étrangers par commune. De plus, chaque valeur des communes est reliée à un diagramme ternaire montrant la proportion de chaque communauté linguistique. Pour la deuxième visualisation (Niwetsch, 2018), que l'on peut appeler «contre-visualisation», il s'agit de montrer des statistiques jugées plus faciles à saisir, en l'occurrence celles de l'indice de Herfindahl par commune et complétées par des cartes de la proportion de chaque communauté (germanophones, italophones, ladinophones, étrangers) par commune. Les deux visualisations seront connectées au même diagramme ternaire et un bouton d'information viendra expliquer brièvement ce que sont les autocorrélations spatiales et l'indice de Herfindahl.

3.2 Procédure d'expérimentation

Lorsque l'on procède à une expérimentation de cette envergure, plusieurs questions liées aux procédures doivent être posées. Nous possédons en effet deux visualisations, mais faut-il les présenter l'une après l'autre ? Si oui, dans quel ordre ? Sinon faut-il montrer une visualisation à une partie des évaluateurs et l'autre, aux autres ? Il existerait ainsi une variable indépendante qui correspondrait au type de visualisation dans notre cas et une variable dépendante offrant la possibilité de mesurer l'expérimentation (le niveau de compréhension des statistiques par exemple) (Martin, 2008). Au niveau de l'expérimentation, il existe deux méthodes (Figure 1) : la première, nommée «plan à groupes indépendants» ou «plan inter-sujets» (*Between-Subjects Experiments*), consiste à n'exposer qu'un seul «niveau» de la variable indépendante (qu'une seule visualisation) à chaque expérimentateur. La seconde, intitulée «plan à mesures répétées» ou «plans à groupes appariés» ou «plan intra-sujets» (*Within-Subject Experiments*), exige que tous les «niveaux» de la variable indépendante leur soient montrés. Ces deux méthodes seront expliquées, mais pas comparées entre elles. Ainsi, pour trancher entre ces deux plans, il est nécessaire de faire le point sur le type d'analyse que nous comptons faire ainsi que de déterminer les avantages et les inconvénients de chacune. En commençant par le plan à groupes indépendants, le principal avantage est que l'expérimentateur n'ayant à examiner qu'une seule variable indépendante (visualisation), il ne reçoit aucune influence qui serait en mesure d'altérer son jugement. Cet avantage permet également de collecter plus de données que s'il fallait juger deux expériences. Il offre en outre la possibilité de réduire le temps d'expérimentation pour chaque participant. En effet, la motivation des expérimentateurs après chaque essai

tend à s'éteindre très rapidement et se confirme par la baisse du nombre de participants (Martin, 2008 ; Légal, s.d.).

Cependant, le plan à groupes indépendants possède également quelques désavantages. A commencer par le plus important à savoir la différence de caractéristiques entre les groupes en fonction du niveau de la variable indépendante. Ainsi, si les groupes n'ont pas de similarité sur certaines caractéristiques, il demeure un risque de falsification des résultats. Pour diminuer la probabilité d'arriver dans un cas de contraste intergroupe, l'usage de la randomisation (attribution aléatoire de la place de chaque individu dans un groupe) offre une solution simple et pratique. Ce choix au hasard peut se faire de différentes façons comme jeter une pièce ou le tirage de billet dans un chapeau. Une autre possibilité qui est combinable avec la randomisation serait d'augmenter l'échantillon choisi pour chaque groupe. Ainsi, plus le nombre de personnes augmente, moins l'assignation des individus dans un groupe ne générera des résultats biaisés, car les différences intergroupes seront minimisées.

A côté du plan à groupes indépendants, il existe le plan à mesures répétées. Il n'est pas forcément considéré par tous les scientifiques comme une méthode supérieure, mais il offre tout de même plusieurs possibilités. Un des premiers avantages par opposition à l'expérience inter-groupe est qu'elle nécessite peu d'expérimentateurs. Les deux méthodes possèdent ainsi une différence de participants issue d'un quotient, dont la valeur correspond au nombre de niveaux de l'expérience inter-groupe. Ainsi, pour un plan à mesures répétées, il faut M participants ; pour une expérience inter-groupe comprenant deux niveaux, il faudra $M * 2$ participants. Ainsi, comme le plan à mesures répétées nécessite moins de participants, la durée totale de l'expérimentation sera plus courte en particulier lorsque l'expérience nécessite un pré-entraînement comme c'est le cas dans les essais liés à la mémoire ou à la cognition. Le plan à mesures répétées offre également des avantages statistiques. En effet, avec cette méthode, il est plus aisé de pouvoir repérer une «différence réelle» dans le cas où «la différence entre niveaux est grande ou si la variabilité estimée au sein des niveaux est petite» (Martin, 2008).

Cependant, le plan à mesures répétées n'offre pas que des points positifs, si bien que certains psychologues mettent en doute son utilité jusqu'à prétendre qu'il ne peut s'utiliser que dans des domaines bien restreints (Poulton, 1973). Outre le problème de l'irréversibilité induite par l'expérience acquise lors du premier essai, nommée *carry-over effect* (effet contagieux), l'ordre dans lequel apparaissent ces expérimentations influence également les participants. L'effet d'ordre (*order effect*) contribue à modifier leur comportement et une compensation méthodologique (les expérimentations doivent être pratiquées dans un certain ordre prédéfini) devient nécessaire. Ainsi, le plan à mesures répétées n'est approprié que pour les essais liés aux sensations ou à la perception (Martin, 2008).

Bien qu'il existe une série de procédures pour mener à bien le plan à mesures répétées, il faut tout de même également tenir compte d'un dernier aspect pour le plan à groupes indépendants. Etant donné qu'il peut exister des différences individuelles au sein des groupes, l'idée serait de pouvoir diminuer leurs effets à l'aide d'un plan à groupes correspondants (*matched-groups design*). Il s'agit en fait d'affecter les participants aux caractéristiques similaires au sein d'un même groupe. Toute forme de dissimilarités est éliminée si deux conditions sont respectées à savoir une randomisation ainsi qu'un nombre de participants élevés. Cependant, il demeure un risque d'ignorer des caractéristiques importantes qui deviendraient des variables indépendantes et nuiraient à la véracité de l'expérience. Ainsi, la similarité entre les individus de

Procédures d'expérimentation

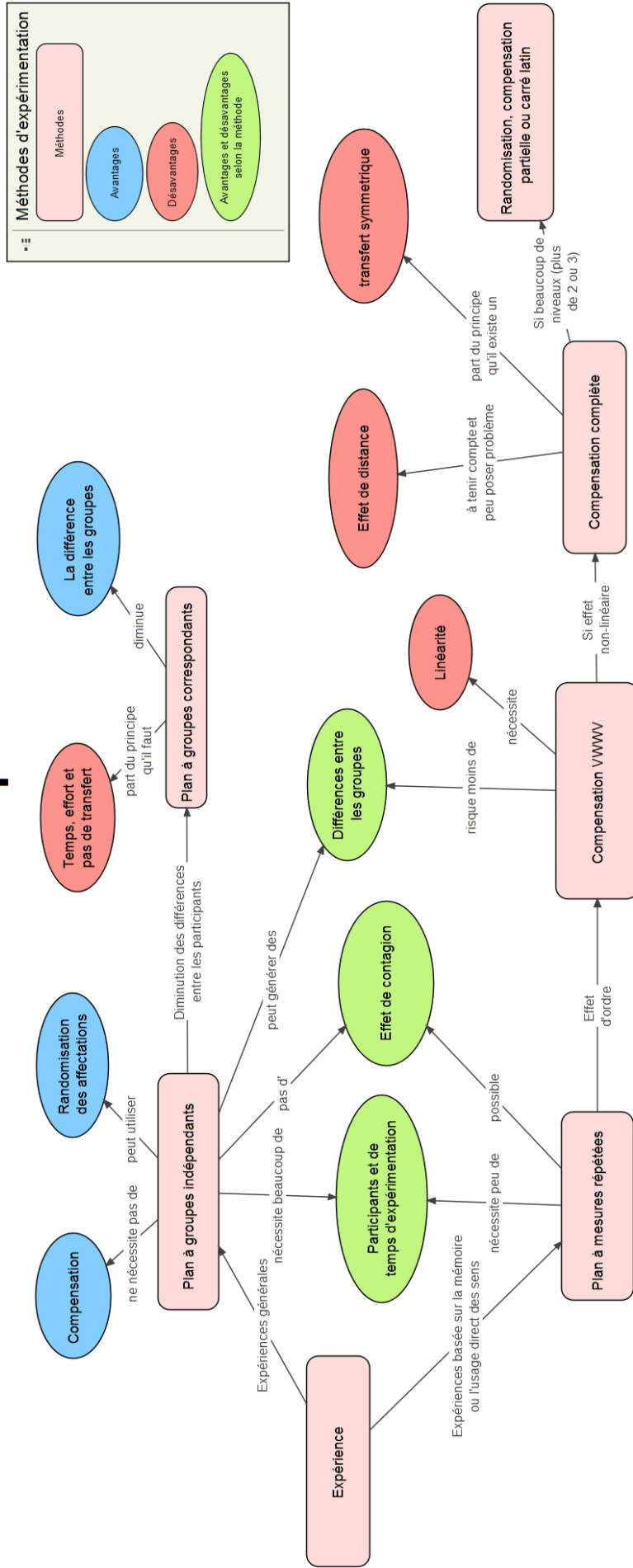


FIGURE 1 – Schéma d'expérimentation résumant les principaux plans de l'ouvrage de David Martin. Seule la partie supérieure a été retenue pour la procédure d'expérimentation

chaque groupe doit être basée sur une variable dépendante fortement corrélée. Il est également possible d'utiliser des tests statistiques pour savoir si une différence de valeur entre les groupes est due à la variable indépendante ou bien au hasard, la méthode étant plus sensible entre ces deux résultats.

Le désavantage du plan à groupes correspondants est qu'il prend du temps à être exécuté. Il faut à vrai dire une première phase de préparation et une phase d'examen (l'expérience en elle-même). Cependant, le plan à groupes correspondants n'est pas une nécessité et comme il peut s'utiliser avec des répercussions positives ou négatives, il est du ressort de l'expérimentateur de décider de son application ou non.

Au final, le plan à groupes indépendants possède l'avantage de n'avoir qu'un seul niveau de variable indépendante à tenir compte si bien que le participant n'est pas affecté par une expérience précédente susceptible de brouiller les résultats. De plus, les temps pour chaque session sont en général plus courts. En outre, les compensations ne sont pas nécessaires, mais l'usage du plan à groupes correspondants et de l'affectation aléatoire permet de diminuer les différences entre individus et d'éviter les valeurs aberrantes. Cependant, des différences entre les groupes sont toujours possibles et il est nécessaire d'avoir plus de participants. Bien que les temps des sessions soient plus courts, celui de l'ensemble de la procédure est plus lent (préparation). Les correspondances sont chronophages et fatigantes et partent du principe qu'il n'y a pas d'influence issue de cette opération.

Le plan à mesures répétées possède l'avantage d'avoir besoin de moins de participants, de pouvoir réduire le temps de préparation et d'avoir des différences entre les groupes plus faibles. Cependant, l'effet d'ordre impose une compensation qui complexifie le processus d'expérimentation.

3.3 Choix et justification de la procédure

Après avoir fait un état des lieux des principaux modes d'expérimentation qui existent en psychologie, il faut à présent choisir lequel serait le plus adapté pour l'expérience de visualisation qui nous concerne. Le plan de la Figure 1 résume de manière générale à la fois les avantages ainsi que les inconvénients de chacune des méthodes. A première vue, le plan à groupes indépendants présente des caractéristiques beaucoup plus faciles à saisir et permet une organisation plus aisée que pour le plan à mesures répétées. Sans réénumérer les qualités et les défauts de chaque plan, la décision peut se faire en fonction de ce que nous pouvons nous permettre et de ce que l'expérience exige pour obtenir les résultats les plus généraux possibles. Ainsi, nos expérimentations ne requièrent pas forcément un nombre limité de participants (au contraire). De plus, le temps n'est pas un problème, car nous ne possédons que deux niveaux de variable indépendante. Cependant, au niveau de la variabilité entre les groupes, il faudra définir assez à l'avance dans quelle catégorie se situeront les participants. Si d'une manière générale, le plan à mesures répétées est plutôt avantageux au niveau des éléments de sa constitution (nombre de participants, temps d'expérimentation et variabilité entre groupes), il possède une procédure beaucoup plus complexe que le plan à groupes indépendants. Ainsi, la première question est de savoir s'il faut plutôt utiliser une méthode «simple» dont les éléments constituent les variables inconnues (plan à groupes indépendants) ou à l'inverse privilégier ces éléments au détriment d'une procédure plus complexe et dépendante de plusieurs facteurs comme le type de compensation. Le second critère concerne le type d'expérience que nous allons effectuer. Martin (2008) conseille d'utiliser le plan à mesures répétées essentiellement pour les expériences liées à la mémoire, à la cognition ou à l'apprentissage. Dans notre cas, aucune de ces

variables psychologiques ne nous intéresse. L'expérience de visualisation est beaucoup plus axée sur la réactivité (temps de réaction), la compréhension (représentations de données statistiques sophistiquées dans leur calcul), de la confiance par rapport à la source et de la satisfaction (vis-à-vis du «produit»). Dans cette optique, le plan à groupes indépendants reste le plus adapté à la dite expérience.

3.4 Cadre de l'expérimentation

Au niveau du cadre de l'expérimentation en lui-même, chaque participant (une vingtaine au total) sera amené à tester l'une ou l'autre des visualisations en fonction d'un tirage au sort, qui déterminera le groupe auquel il appartiendra. Les séances se dérouleront dans une bibliothèque, lieu dans lequel les dérangements sont atténués et où le calme régnant offre à chacun des testeurs la concentration la plus optimale pour cet exercice.

Au niveau du recrutement des testeurs, nous avons lancé un appel à participants sur divers réseaux sociaux et avons accepté directement les personnes qui avaient répondu présent pour l'expérience. Le message annonçait simplement que nous recherchions des candidats pour tester une application à but éducative pour un temps d'environ 20 minutes et que le questionnaire qu'ils devront répondre était totalement anonyme. Ces personnes font partie de notre cercle plus ou moins proche auquel se sont ajoutés d'autres candidats liés à ces individus. Au vue de la nécessité de répartir les testeurs de manière équitable en particulier pour les sexes, certains candidats se sont vus refuser leur participation lorsque le nombre nécessaire de représentants de chaque sexe était atteint, soit 10 par catégorie.

Par ailleurs, les expériences se sont déroulées en plusieurs temps. D'abord, il est à notre sens nécessaire que les testeurs prennent connaissance du fonctionnement des applications. C'est pourquoi une période d'environ 5 minutes est donnée dans un premier temps pour découvrir l'ensemble de l'interface. Dès le départ, il est demandé au testeur d'annoncer ouvertement qu'il a fini d'explorer l'application. Une fois cette tâche accomplie, chaque candidat est amené à comprendre les statistiques sophistiquées qu'il a en face de lui. Puis, il est nécessaire de vérifier et d'aider les testeurs à se situer notamment en fonction de la compréhension des statistiques sophistiquées et du diagramme ternaire. Durant toute l'expérimentation, le testeur est autorisé à parler librement ou à poser des questions, qui n'obtiennent réponse que si elles concernent le questionnaire. Si une question a un rapport avec la compréhension des statistiques sophistiquées, le candidat est amené à vérifier par lui-même s'il n'a pas oublié d'explorer des éléments de l'application nécessaires pour sa compréhension. Pour les autocorrélations spatiales, les questions sont les suivantes : Comment comprenez-vous la zone brune foncée homogène chez les ladinophones ? Comment comprenez-vous la couronne bleue qui entoure la zone brune foncée chez les ladinophones ? Pourquoi y a-t-il une limite de communes blanches entre le Trentin et le Tyrol du Sud pour les germanophones ? Pourquoi le Trentin apparaît-il en brun clair pour les germanophones ? Pourquoi un îlot de communes bleues se distingue pour les germanophones dans le Trentin ? Ces questions requièrent l'usage du bouton informatif et du diagramme ternaire, qui montre les proportions des trois communautés indigènes par commune. Les questions pour ce graphique sont les suivantes : Comment comprenez-vous les trois axes du diagramme ternaire ? Pourquoi la grande majorité des communes est située sur ces axes ? Pourquoi existe-t-il une grande concentration de points sur les sommets (bouts) des axes et peu au milieu ? Que signifie une commune qui se situe sur un sommet et une au milieu ? Où devrait se situer une commune qui aurait une proportion de chacune des trois communautés linguistiques égale ? Avec toutes ces constatations, existe-t-il d'une manière générale une ségrégation spatiale ? De plus, pour l'indice de

Questionnaire sur la visualisation des données

Répondez au questionnaire durant ou après l'expérimentation de la visualisation. Merci d'entourer la réponse pour les propositions Homme / Femme, Oui / Non et 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10, 1 étant la note la plus médiocre et 10 la meilleure. **Ce questionnaire est anonyme, n'écrivez donc pas votre nom !**

Questions de base

Sexe : Homme / Femme

Visualisation : Pile / Face

Âge : _____

1. Connaissez-vous le Tyrol italien ou en avez-vous déjà entendu parler ? Oui / Non
2. Saviez-vous où le Tyrol italien se situait avant d'utiliser la visualisation ? Oui / Non
3. Connaissez-vous les différences linguistiques dans la région avant? Oui / Non
4. Aviez-vous des connaissances préalables dans le domaine des autocorrélations spatiales ou de l'indice de Herfindahl? Oui / Non

Questions sur la visualisation

5. L'affichage des éléments vous paraît-il rapide ? 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10
6. L'interaction est-elle suffisamment rapide ? 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10
7. Les fonctionnalités vous semblent-t-elles pratiques? 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10
8. Comprenez-vous les thématiques statistiques ? 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10
9. Comprenez-vous le diagramme ternaire ? 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10
10. Les informations vous semblent-elles correctes? 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10
11. Les informations vous semblent-t-elles complètes? 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10
12. Utiliseriez-vous cette visualisation en tant que référence pour un travail, une publication ou tout simplement comme une source fiable ? 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10
13. Votre curiosité a-t-elle était stimulée ? 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10
14. Cette visualisation est-elle satisfaisante? 1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / 9 / 10
15. Quels sont les points positifs de cette visualisation?

16. Quels sont les points négatifs de cette visualisation?

17. Avez-vous des remarques ou des suggestions d'améliorations?

FIGURE 2 – Questionnaire sur la visualisation

Herfindahl, les questions prennent la forme suivante : Comment comprenez-vous la définition de l'infobulle du bouton ? Que signifie les communes qui s'affichent en vert ? Que signifie les communes qui s'affichent en beige ? Où se situent les communes où la présence des quatre communautés est la plus égalitaire ? Voyez-vous un lien entre l'égalité des communes et la répartition sur le diagramme ternaire ?

En outre, chaque participant a entre 20 et 30 minutes pour à la fois tester la visualisation ainsi que pour répondre au questionnaire (Figure 2) qui accompagne l'expérimentation. Le questionnaire contient diverses questions relatives à la fois au type de personne et à ses connaissances de base et des questions sur l'application en elle-même. Ainsi, seules deux questions sont considérées comme personnelles à savoir le sexe et l'âge du testeur. Ensuite, il s'agit de sélectionner l'application qu'il devait tester. Le choix se présente sous la forme de «pile» pour la visualisation de l'indice de Herfindahl et «face» pour celle des autocorrélations spatiales. Puis, 4 questions traitent des connaissances de base des testeurs. Elles permettent de vérifier si la personne a déjà entendu parler (question 1) ou connaît (question 2) le Tyrol italien, les trois communautés linguistiques historiques (question 3) et une des mesures statistiques (autocorrélation spatiale locale ou indice de Herfindahl selon la visualisation reçue) (question 4).

Par la suite, 10 (5 à 14) questions composent l'appréciation des testeurs quant aux caractéristiques spécifiques de l'application. La question 5 traite de la rapidité de l'affichage des éléments (carte, graphique, etc.). La question 6 est basée sur la rapidité d'exécution de l'interactivité entre les deux éléments (par exemple, en passant la souris sur la carte, le graphique ternaire affiche des informations et les points changent de couleur ou en double-cliquant sur la carte, un zoom sur la commune sélectionnée est activé). La question 7 traite de la praticité des fonctionnalités comme l'usage d'un bouton d'informations sur les statistiques qui affiche une infobulle ou l'affichage des noms de communes par effleurement des unités spatiales ou encore des informations complémentaires offertes par le diagramme ternaire sur la composition ethnique dans la commune sélectionnée. Il s'ensuit de la question 8, très essentielle pour ne pas dire la plus importante, celle de la compréhension des thématiques statistiques à savoir l'indice de Herfindahl ou l'autocorrélation spatiale locale. Tout ce que présentent les applications est sensé permettre aux testeurs de comprendre les informations que donnent les cartes. La question 9 évalue la compréhension du diagramme ternaire et doit permettre ainsi d'estimer s'il est nécessaire pour comprendre les statistiques sophistiquées dans leur calcul ou au contraire pose plus de problèmes quant à son objectif. La question 10 vérifie si la personne peut faire confiance aux informations qui lui sont présentées. Il ne s'agit pas de savoir si les informations sont réellement correctes ou non, mais si pour les personnes qui les utilisent, elles pourraient être correctes. La question 11 révèle si les testeurs ont estimé que les informations qui leur étaient présentées sont suffisamment complètes pour atteindre l'objectif de compréhension des statistiques sophistiquées ou manque de détails ou d'éléments pouvant faciliter la vocation des applications. La question 12 met en lumière le niveau de confiance qu'auraient les utilisateurs quant aux applications présentées. Elle permet de voir si les testeurs seraient prêts à utiliser ces visualisations pour des besoins de publications ou comme référence fiable dans le domaine de la ségrégation spatiale. La question 13 vérifie la curiosité des testeurs durant le processus d'expérimentation. Finalement, la question 14 met en évidence la satisfaction des utilisateurs quant à l'objectif de compréhension des statistiques sophistiquées.

Ces questions sont suivies de 3 autres sur les aspects positifs, négatifs et les suggestions d'amélioration des applications. Une énième question, en l'occurrence la 18^e, n'apparaissant pas dans le questionnaire, a été tout de même prise en compte. Elle s'énonce de la manière suivante *Avez-vous trouvé*

cette visualisation instructive ? A travers cette interrogation, le but n'est pas de savoir si les utilisateurs ont compris les informations, mais s'ils auraient potentiellement appris quelque chose de nouveau et ainsi si les applications pourraient avoir un but pleinement éducatif, ne mettant pas seulement en jeu la compréhension, mais l'instructivité. Concrètement, durant les expérimentations, il a été demandé à chaque testeur d'inscrire la question manuellement et d'y répondre de la même manière que les 10 autres. Ainsi, nous avons 11 questions pour lesquelles le testeur sera amené à donner une note comprise entre 1 (médiocre) et 10 (excellent), précisant ainsi son sens critique et permettant par la suite d'obtenir des valeurs numériques facilitant les calculs lors de la phase suivante : le traitement des données des questionnaires.

Parmi ces 11 questions à appréciation numérique, certaines pourraient être perçues comme suffisamment similaires pour passer pour inutiles. Il s'agit des questions 1 et 2 ; 5 et 6 ; 10 et 12. Or, même s'il la manière dont elles sont posées paraît proche, leur signification diverge légèrement. En effet, ces questions sont complémentaires et non pas redondantes. Ainsi, entre la première et la deuxième, la différence réside respectivement dans l'existence du Tyrol italien et dans sa localisation géographique pour chaque testeur. Cependant, ces deux questions s'impliquent : si l'utilisateur n'a jamais entendu parler du Tyrol italien, il y a normalement très peu de chance qu'il sache où il est localisé. Ceci serait en général vrai pour n'importe quelle région dans le monde, mais pour celle qui nous intéresse, la définition de sa localisation réside dans son nom. Ainsi, *Tyrol* fera penser à l'Autriche et *italien* à l'Italie dirigeant «naturellement» la personne vers une zone au située au Nord-Est de l'Italie. Pour ce qui est des questions 5 et 6, il est peut-être moins évident de sentir la différence, car elle est en effet très minime. La question 6 a pour but de préciser la question 5 sur la rapidité des interactions directes avec les actions qu'elle effectue, tandis que la question 5 montre l'affichage d'une manière générale. Cependant, la question 6 aurait pu suffire à elle seule. Finalement, les questions 10 et 12 présentent la fiabilité respectivement de manière prompt et générale, et comme base pour diffuser un contenu ou une information à d'autres personnes à travers une publication par exemple. Ces explications complémentaires seront transmises oralement durant les expérimentations en cas d'incompréhension.

En outre, les seules questions relatives à la personne elle-même concerne le sexe et l'âge, variables nécessaires afin de diminuer les «discriminations» potentielles quant à l'accessibilité éducative de l'application. Sur les deux applications, il y aura autant d'hommes que de femmes qui testeront sa validité. Ainsi, au total, 10 femmes et 10 hommes répartis par 5 représentants de chacun des sexes pour avoir une égalité parfaite, seront confrontés à une des visualisations. Au niveau de l'âge, il n'est pas forcément facile et nécessaire d'avoir une répartition équivalente entre «seniors» et «jeunes adultes», mais au moins avoir un représentant par sexe par visualisation. En effet, les seniors ont moins tendance (par peur ou méconnaissance) à accepter de participer à des expériences dans lesquelles l'informatique est centrale. Il faut ainsi compter avec un nombre de participants seniors plus faible que les jeunes adultes. Ainsi, au minimum, il faudra 4 personnes ayant plus de 50 ans (et par conséquent 16 qui ont plus de 20 et moins de 35). Elles seront réparties de manière à ce qu'un homme et une femme seniors participent à la première visualisation et l'autre homme et l'autre femme seniors à la seconde. Nous avons ainsi un représentant senior des deux sexes pour chaque visualisation. De plus, pour des raisons de praticité, nous mettons un âge minimum et maximum pour autoriser l'accès aux expérimentations. Ainsi, seule toute personne ayant entre 20 et 70 ans pourront participer aux examens. Cette tranche d'âge d'une cinquantaine d'années approche également respectivement l'âge de la majorité (souvent comprise entre 18 et 21 selon les pays) et l'âge de la retraite (entre 60 et 70 ans selon le pays). Ainsi, mineurs et retraités sont exclus.

3.5 Traitement des données

Une fois les expérimentations terminées, les vingt questionnaires seront rassemblés. A travers l'usage de méthodes quantitatives, diverses caractéristiques devront ressortir, le but étant de vérifier quel groupe de population a eu un accès facile aux ressources éducatives de l'application et lequel a eu plus de peine à acquérir le savoir véhiculé. Dans un premier temps, des statistiques descriptives (moyenne, médiane, écart-type, minimum et maximum) seront utilisées pour avoir une vision générale des aspects positifs et négatifs de l'application. 11 questions possèdent une valeur numérique comprise entre 1 et 10. Chaque question sera ainsi traitée indépendamment des autres en fonction de la valeur estimée par chacun des testeurs. Comme les deux applications contiennent des thématiques très différentes, les données seront traitées de manières séparées en fonction de la visualisation à laquelle les utilisateurs ont été confrontés. Ceci permettra une comparaison des résultats entre les deux applications et ainsi de voir si les autocorrélations spatiales, méthode statistique jugée difficile à comprendre, seraient autant comprises que l'indice de Herfindahl, méthode statistique jugée facile à comprendre. Afin de représenter les statistiques descriptives de manière plus évidente, des boîtes à moustache seront utilisées. Ces dernières présentent l'avantage de montrer plusieurs quantiles (le minimum, le premier décile, le premier quartile, le deuxième quartile (médiane), le troisième quartile, le neuvième décile et le maximum). Mis côte à côte, ils permettent d'observer les profils de chacune des questions pour une visualisation donnée. A cela s'ajoutera les valeurs des moyennes pour chaque question. De manière absolue, des valeurs supérieures à 8 pour la moyenne et la médiane permettront d'affirmer que les objectifs de l'application sont atteints. Entre 6 et 8, des améliorations seront à effectuer (le questionnaire laisse la possibilité d'énoncer les points positifs, négatifs et les améliorations à apporter aux applications). En dessous de 6, l'objectif annoncé par la question n'est pas atteint et il faudra repenser en entier la visualisation afin de l'intégrer.

Afin d'avoir la comparaison entre les visualisations, un test de Student (test t) sera utilisé pour valider les analyses entre les deux visualisations. Par la suite, tout le processus de traitement (boîtes à moustache et test de Student) devra être répété pour ne prendre en compte que les hommes, puis que les femmes, puis les personnes de plus de 50 ans et finalement celles de moins de 35 ans. Un test ANOVA (test f) sera également utilisé entre les deux visualisations et de prendre deux facteurs (le sexe et l'âge). Ceci permettrait d'observer les effets principaux et l'existence ou non d'une interaction entre ces deux facteurs.

4 Résultats

A présent que les aspects méthodologiques ont été exposés, il est possible d'apprécier les résultats issus des calculs et de représentations cartographiques attachés.

4.1 La loi de Zipf

Avant de se lancer dans tout calcul d'indice, il est nécessaire de bien évaluer les données que nous possédons. Etant donné que nous travaillons plutôt dans un domaine démographique, la loi de Zipf offre la possibilité d'avoir un bon aperçu de la distribution des communautés linguistiques du Tyrol italien. Il est possible de calculer cette observation à l'aide du logiciel R, dont le code provient du site internet de *Stack Exchange* (2017). Ce code (Mpiktas, 2017) situé dans les annexes de ce travail, calcule également la méthode des moindres carrés qui apparaît sous forme de droite rouge (Perron, 2003) et l'estimateur du maximum de vraisemblance sous forme de ligne verte (Chesneau, 2017). Ces droites s'ajoutent aux graphiques de la loi de Zipf. Nous commencerons par analyser les données de population en valeurs absolues, puis celle en valeur relatives.

Ainsi, nous débutons l'analyse avec les germanophones. Pour éviter des opérations mathématiques aberrantes, telles que $\log_{10}(0)$, sur R, nous avons décidé de retirer certaines valeurs. En effet, dans 150 des 333 communes (Table 1), la présence des germanophones est inexistante. Il y a donc en tout 150 rangs de Zipf, qui représentent un peu moins de la moitié des communes. Pour le nombre de germanophones (Figure 3a), nous constatons que la courbe décroît un peu au fur et à mesure que le rang augmente, puis chute de manière brutale à partir du rang 90. La méthode des moindres carrés montrent une valeur de 1,31, tandis que pour celle du maximum de vraisemblance, elle est de 0,78 (Table 1). Pour la loi de Zipf pour le taux de germanophones par communes (Figure 3b), la fonction est constante jusqu'au rang 100, à partir duquel une chute rapide de la fréquence est enregistrée. De plus, la valeur des moindres carrés vaut à présent 1,13 pour un maximum de vraisemblance à 0,33 (Table 1) du fait de la forme plus arrondie de la courbe de données. Par ailleurs, la droite des moindres carrés et celle du maximum de vraisemblance sont très éloignées l'une de l'autre autant pour les taux que pour les nombres de germanophones. Cette configuration s'explique par deux facteurs. Le premier vient du fait que la population germanophone est essentiellement concentrée dans le Tyrol du Sud et est «presque inexistante» dans le Trentin. La seconde est que là où les germanophones sont présents, ils sont en général très majoritaires, car ils vivent plutôt dans les campagnes. Même dans les villes du Tyrol du Sud, ils sont minoritaires.

Pour les italophones, il faut également tenir compte du fait qu'ils sont présents dans 332 des 333 communes que compte le Tyrol italien (Table 1). Au niveau du nombre d'italophones, la décroissance de la courbe est beaucoup plus régulière (Figure 3c). Ce n'est qu'à partir du rang 200 qu'elle chute de manière plus importante. Cette tendance régulière peut également être appréciée par les deux droites qui sont très proches l'une de l'autre tout le long de la fonction. La droite des moindres carrés vaut 1,26 et celle du maximum de vraisemblance 1,07 (Table 1). La configuration de cette courbe montre par conséquent que les italophones sont beaucoup mieux répartis dans les diverses communes du Tyrol italien que les germanophones. Cependant, il est tout de même possible d'apprécier grosso modo le fait que les 217 premiers rangs représentent essentiellement des communes du Trentin et les 115 derniers des communes du Sud-Tyrol (Table 1). En effet, le Tyrol du Sud possède des communes beaucoup plus homogènes de germanophones que le Trentin, dominé plutôt par les italophones et dont les communes sont plus

nombreuses (217 pour le Trentin et 116 pour le Sud-Tyrol). Toutefois, les pourcentages d'italophones par commune (Figure 3d) indiquent une configuration toute autre, comparable à celle des germanophones. En effet, jusqu'au rang 200, la fonction est très constante, puis chute brusquement après. Les moindres carrés montrent une valeur inférieure à 1 (0,96) et le maximum de vraisemblance est très bas (0,37) (Table 1), ce qui dénote bien du caractère très brusque de la chute de la courbe de données. Ainsi, bien que les italophones soient présents dans toutes les communes, il y en a 200 dans lesquels ils sont très présents et 133 où ils sont minoritaires voire très minoritaires. A nouveau il s'agit des variantes subrégionales entre le Trentin et le Tyrol du Sud. De plus, le maximum de vraisemblance suit plutôt la tendance des 200 premières communes tandis que les moindres carrés sont très influencés par les 132 suivantes, d'où la disparité entre les droites comme dans le cas des germanophones.

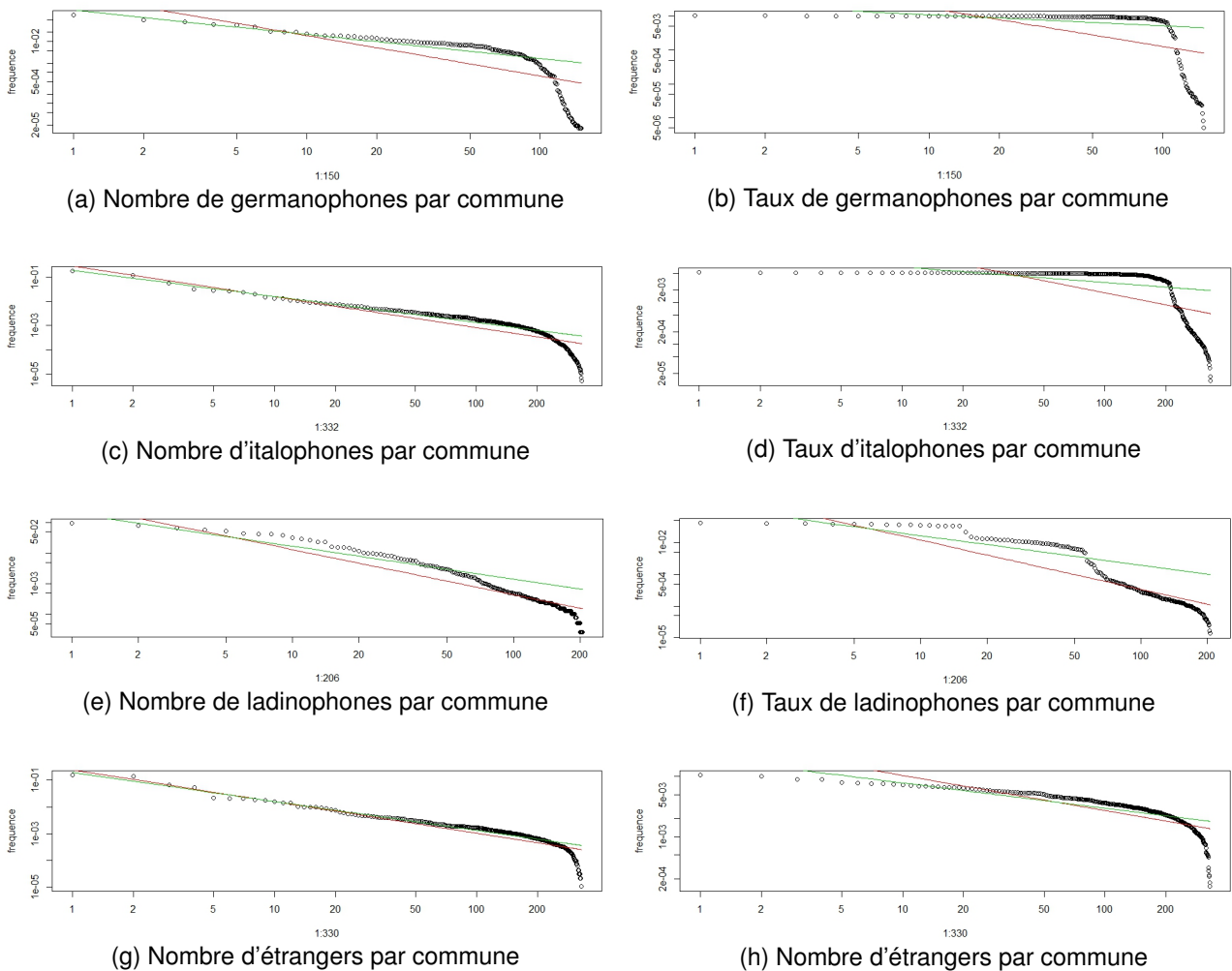


FIGURE 3 – Loi de Zipf pour chaque communauté (nombre et taux) en échelle logarithmique. La ligne rouge correspond à la méthode des moindres carrés et la verte à l'estimateur du maximum de vraisemblance

La courbe du nombre de ladinophones (Figure 3e) est certainement la plus rectiligne de toutes les communautés analysées. Ce côté très droit peut également être apprécié à travers une droite des moindres carrés à hauteur de 1,48 et une autre droite de maximum de vraisemblance à 1,08 (Table 1). En effet, les ladinophones sont d'une manière générale plus présents dans les communes du Tyrol italien que les germanophones (206 communes pour les ladinophones contre 150 pour les germanophones). Cependant, les ladinophones ont une présence très différente d'une commune à l'autre, mais leur nombre reste assez faible en comparaison des deux autres communautés précédentes. En effet, il ne dépasse pas 3720 dans

la commune qui en compte le plus en termes absolus (commune d'Ora/Auer). Concernant le taux de ladinophones (Figure 3f), la courbe se fractionne en trois parties. La première, jusqu'au rang 16, montre des communes dans lesquelles ils sont largement dominants (plus de 90% en général) et concerne essentiellement les régions à l'Est du Tyrol italien (en grande partie dans le Sud-Tyrol et au Nord-Est du Trentin). Puis, jusqu'au rang 57, il y a les communes dans lesquelles ils sont moyennement présents (supérieur à 10%). Comme il s'agit essentiellement de communes du Nord-Ouest du Trentin, ils auront plutôt tendance à partager leur territoire avec les italoophones. Pour finir, la dernière tranche de la courbe montre les communes dans lesquelles ils sont très peu présents (moins de 10%). Entre les deux courbes de régression, la distance reste assez courte du fait de la faible rondeur de la courbe de données. Cependant, les moindres carrés valent 1,55, ce qui est plus fort que pour le nombre de ladinophones. Il en est de même pour le maximum de vraisemblance avec 0,93 (Table 1).

	Germano.		Italo.		Ladino.		Etran.	
	Nbre	%	Nbre	%	Nbre	%	Nbre	%
Moindres carrés	1,31	1,13	1,26	0,96	1,48	1,55	1,18	0,58
Max. vraisemblance	0,78	0,33	1,07	0,37	1,08	0,93	1,08	0,42
Présence communes	150		332		206		330	

TABLE 1 – Synthèse de la méthode des moindres carrés et du maximum de vraisemblance par communauté (germanophones, italoophones, ladinophones, étrangers) en nombre et proportion, et du nombre de communes où une communauté est présente

Pour les étrangers, leur présence se fait dans 330 des 333 communes que compte le Tyrol italien (Table 1). Au niveau de leur nombre (Figure 3g), la configuration est encore différente. Les moindres carrés valent 1,18 et le maximum de vraisemblance 1,08. De plus, les deux droites sont presque confondues. Il est intéressant de retrouver le même phénomène que pour les taux de ladinophones à savoir des paliers plus ou moins marqués de la courbe de données. Les trois premiers paliers (jusqu'au rang 12) désignent les grandes villes du Tyrol italien (les deux premières étant respectivement Trento et Bolzano/Bozen). Aussi, plus le rang augmente et plus l'espace entre les communes se resserre. Ceci est dû au fait que la proportion d'étrangers s'adapte à la taille de la commune citadine, cependant elle reste «assez» indifférenciée dans les campagnes, dans lesquelles leur proportion dépendra plutôt du type de commune (résidentielle, touristique, d'activité, etc.). Avec les taux d'étrangers (Figure 3h), la courbe est assez rectiligne jusqu'à la 300^e commune après laquelle elle chute. Les villes sont les premières à être représentées puis les campagnes, pour lesquelles le rang de fréquence ne varie plus beaucoup en fonction de l'augmentation du rang. Les moindres carrés valent 0,58 ce qui est assez faible, bien que les deux droites de régression se rejoignent. Le maximum de vraisemblance est quant à lui de 0,42 (Table 1).

En définitive, la loi rang-taille proposée par Zipf montre assez bien la configuration spatiale et offre ainsi une meilleure appréhension des analyses liées aux divers indices de caractérisation du territoire.

4.2 L'indice d'Herfindahl

En regardant des cartes de répartition linguistique, nous pouvons apprécier la distinction assez claire entre un Trentin très italoophone, un Haut-Adige très germanophone et une partie orientale (chevauchant ces deux régions) très ladinophones. Toutefois, ce genre de visualisation ne montre que le groupe dominant dans chaque région. C'est pourquoi, il est impossible de cette façon d'apprécier l'ensemble des répartitions communales, car aucune information n'est donnée sur les groupes plus faiblement présents. L'exemple de Merano/Meran est certainement le plus typique. En effet, classée germanophone en ré-

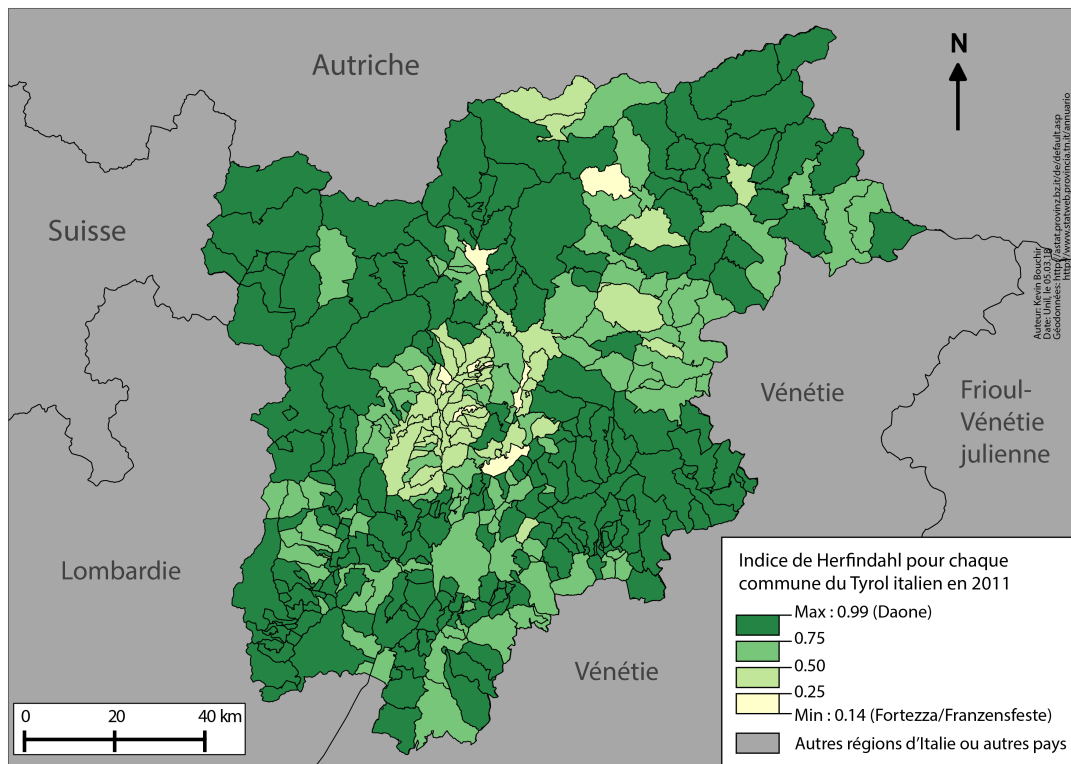


FIGURE 4 – Coefficient d'Herfindahl par commune. Plus l'unité spatiale est foncée plus une communauté est dominante

gion du Sud-Tyrol, elle possède pourtant presque autant de germanophones (42, 49%) que d'italophones (41, 30%). Pour pallier à ce problème, l'indice d'Herfindahl normalisé (3) peut offrir une première approche pour en apprécier les concentrations en fonction de chaque communauté (prise en compte en même temps dans la même unité spatiale). Pour rappel, il s'agit d'un indice qui varie de 0 à 1 pour montrer respectivement la concentration et la dispersion au sein de l'unité spatiale. Cependant, il faut comprendre par là que mathématiquement une unité spatiale est concentrée si les proportions de chacune des communautés sont très différentes avec notamment une d'elles qui est largement dominante. Le fait que l'indice d'Herfindahl mette les proportions de chaque locuteur au carré signifie que plus il sera petit et plus il aura tendance à être petit au carré et diminuera moins vite si le nombre est proche de 1 (le carré de 0,1 est 0,01 alors que celui de 0,99 est de 0,9801).

L'indice d'Herfindahl possède l'avantage de pouvoir s'appliquer à toutes les formes d'unité spatiale, car il ne fait pas de comparaison entre deux échelles contrairement aux indices qui suivront. C'est pourquoi deux échelles ont été choisies à savoir celle du Tyrol italien et celle de chaque commune afin de mieux apprécier la situation générale et locale de la ségrégation spatiale au Tyrol italien. Ainsi, l'indice général normalisé possède une valeur de 0,22 et montre une concentration peu importante ou plutôt une égalité assez forte dans la répartition des communautés (assez dispersées). Sa valeur de concentration plutôt faible laisse à penser qu'il y a beaucoup de zones dans lesquelles les communautés sont bien réparties et assez peu dans lesquelles une d'elles dominerait. Ceci est théorique et se baserait sur le fait que la plupart des territoires possèdent la même composition ethnique.

Cependant, il est aussi possible d'interpréter ce résultat comme le fruit de la «zonation» entre les différentes communautés. Chacune d'elles posséderait ainsi une zone (un district par exemple) dans laquelle elle est largement majoritaire et ainsi la ségrégation serait bien présente. Etant donné le contexte particu-

lier du Tyrol italien et le doute entre les interprétations bien que la seconde option ait plus de chance d'être réelle, il est nécessaire d'analyser la situation à une échelle plus fine, celle des communes (3) étant la plus privilégiée pour ce travail. Ainsi, ce niveau d'échelle laisse entrevoir une configuration assez proche de la deuxième hypothèse montrée à travers l'indice d'Herfindahl normalisé global (4) (Figure 4). L'interprétation principale qui peut se faire est au niveau du centre du Tyrol italien. En effet, il semblerait que ces communes possèdent une plus grande égalité entre les communautés que dans la majeure partie du Tyrol italien. De plus, quelques communes peuvent être considérées comme très égalitaires comme par exemple Fortezza/Franzensfeste, Merano/Meran ou Salorno/Salurn. En outre, entre Salorno/Salurn et Merano/Meran, toute une série continue de communes possède un indice d'Herfindahl assez faible. Parmi, elles se trouve Bolzano/Bozen, le chef-lieu du Sud-Tyrol. Bressanone/Brixen, autre commune de la région, possède également un indice assez bas. Pour certaines communes, le phénomène d'égalité est en général lié à une forte proportion d'étrangers dans les zones plutôt campagnardes ou à une forte présence germanophones pour les communes citadines du Tyrol du Sud. Dans une moindre mesure, la présence des ladinophones «implique» également d'avoir un indice plus bas, mais ils sont par contre égalitaires non pas avec une, mais deux autres communautés, en général les italophones et les étrangers. Ceci montre bien qu'entre les germanophones et les ladinophones, il y a assez peu de contact, car ce sont des communautés historiques longtemps présentes dans le Tyrol italien. Les italophones et les étrangers sont beaucoup plus en contact, car ils sont les nouveaux-venus issus d'après les deux guerres mondiales.

4.3 Gini et Duncan : les étrangers mieux répartis que les nationaux ?

L'indice de Gini G_a qui d'habitude sert pour analyser les disparités de revenus notamment entre les classes riches et pauvres ou entre hommes et femmes ou pour certaines études sur les populations, peut également servir pour des analyses géolinguistiques. Pour rappel, plus l'indice est proche de 0, plus la communauté a est dispersée et vice versa pour 1. En général, il est plutôt attendu que les indigènes soient plus dispersés que les étrangers. Cependant, le cas du Tyrol italien nous montre tout l'inverse et de manière assez forte. En effet, les germanophones (0,92) et les ladinophones (0,93) sont extrêmement proches de la concentration parfaite. Les italophones suivent de près avec (0,82), mais montrent tout de même une plus grande dispersion. Et par opposition aux indigènes, les étrangers (0,28) sont les plus dispersés et de manière plus radicale que les italophones. Ce phénomène pour le moins inattendu pourrait choquer à première vue. Cependant, en se penchant sur le cas particulier du Tyrol italien, une certaine logique ressort. En effet, la loi de Zipf avait montré une forme de concentration très forte pour les ladinophones et les germanophones. Ils sont normalement présents dans les communes dans lesquelles ils sont proportionnellement et numériquement majoritaires. Ces communes sont également en général campagnardes et la concentration de ces communautés se fait respectivement dans l'Est du Tyrol italien et dans le Sud-Tyrol. Les italophones, eux, possèdent un indice de Gini élevé, mais plus bas que les deux autres groupes indigènes, car ils sont plutôt concentrés dans le Trentin, mais possèdent d'importantes communautés dans le reste du Tyrol italien et notamment dans les communes citadines du Sud-Tyrol. Au final, ce sont les étrangers qui sont les mieux répartis (en nombre et en proportion), car ils sont présents dans quasi toutes les communes (330 sur 333) comme l'a montré la loi de Zipf, mais surtout dans les communes citadines.

	Gini	Duncan	différence
germanophones	0,92	0,80	0,12
italophones	0,82	0,69	0,13
ladinophones	0,93	0,85	0,08
étrangers	0,28	0,21	0,07

TABLE 2 – Coefficient de Gini et indice de ségrégation de Duncan pour chaque communauté

En comparaison avec les résultats du coefficient de Gini, l'indice de ségrégation de Duncan, IS_a , montre des résultats relativement proches, mais avec un coefficient plus faible de l'ordre de 0,07 à 0,13 points. Ainsi, pour les germanophones (0,80), les italophones (0,69), les ladinophones (0,85) et les étrangers (0,21), ces coefficients s'expliquent de la même manière que pour ceux de Gini. Il faut également mentionner que les deux indices possèdent une formule très similaire, quoique distincte. En effet, Duncan ne tient compte que de la somme des i communes, tandis que Gini prend en compte également la somme des j communes (non forcément adjacentes), car il fait la comparaison constante entre 2 communes du Tyrol italien. Duncan, lui, ne considère qu'une seule commune à la fois, mais la compare directement à l'échelle du Tyrol italien. Aussi, le nombre d'habitant de ce dernier se retrouve multiplié par lui-même chez Gini, alors qu'il est non transformé pour Duncan. D'une manière générale, l'indice de Gini contient un facteur $\sum_j^n \frac{N_j}{N}$ en plus de l'indice de Duncan. En soit, la valeur finale de ce facteur est 1, vu que $\sum_j^n N_j = N$. Or, combiné avec les autres éléments de l'indice de Gini, il montre un résultat assez différent de l'indice de Duncan. Ainsi, estimer la différence entre ces deux indices ne fait pas forcément sens du fait de leur construction respective, mais l'amplitude de leur différence pour chaque communauté montre que plus la ségrégation spatiale est faible, plus la valeur des deux indices sont proches d'une manière générale. Etant donné que l'indice de Gini compare chacune des communes entre elles par rapport à l'ensemble de la région, il paraît plus approprié pour apprécier les inégalité de répartition des communautés.

4.4 Duncan ou Gorard : quelles différences ?

Dans la phase de développement des indices de ségrégation spatiale, Morill (1991) et Wong (1993, 1998) s'étaient basés sur l'indice de dissimilarité de Duncan (1955). Cependant, Gorard (2002), une décennie après, a contesté la validité de cet indice. C'est sur cette base que nous allons vérifier la pertinence de ces publications. Etant donné que l'indice de dissimilarité de Duncan (7) compare entre elles deux communautés, il en ressort une matrice 4x4 :

Duncan gentilé/gentilé	germanophones	italophones	ladinophones	étrangers
germanophones	-	0,83	0,91	0,64
italophones	0,83	-	0,86	0,27
ladinophones	0,91	0,86	-	0,84
étrangers	0,64	0,27	0,84	-

TABLE 3 – Matrice représentant les indices de dissimilarité de Duncan entre chaque communauté

Selon les résultats de la Table 3, la dissimilarité entre les différentes communautés est d'une manière générale très forte. Les résultats les plus importants concernent essentiellement les citoyens italiens. En effet, l'indice est de 0,83 entre les germanophones et les italophones, de 0,86 entre italophones et les ladinophones et monte jusqu'à 0,91 entre les germanophones et les ladinophones. Le fait que ces deux dernières communautés aient une valeur aussi importante vient essentiellement du fait que dans la zone du Sud-Tyrol où ces communautés sont les plus importantes, elles atteignent un niveau d'homogénéité

tel qu'une forme de ségrégation pourrait être suspectée. Les italophones, bien que concentrés dans le Trentin, se «mélangent» dans passablement de communes. Quant aux étrangers, ils sont moins présents dans les zones ladinophones, d'où la valeur de 0,84 et sont plus présents dans les villes que dans les campagnes. L'indice descend à 0,64 dans les zones germanophones du fait de la présence de certaines grandes villes comme Bolzano/Bozen, Merano/Meran ou encore Bressanone/Brixen. C'est par contre avec les italophones que les étrangers sont le plus souvent présents. La valeur de 0,27 (la plus faible), montre que la dissimilarité entre ces deux communautés existe, mais reste très basse en comparaison aux autres. Le fait que, d'une manière général, il y ait plus de villes dans le Trentin et qu'elles soient plus grandes que dans le Tyrol du Sud, fait ressortir un tel résultat.

Ces valeurs issues de l'indice de dissimilarité de Duncan donnent déjà une petite idée de la ségrégation qui peut exister au sein de ce territoire. Cependant, comme l'indice de Gorard est plus récent et est basé sur celui de Duncan, «contesté», les résultats attendus devraient être plus «pertinents» que ceux issus de Duncan. Comme l'indice de Gorard (2002) (10) propose de comparer chacune des communautés à la population totale. Il en ressort donc quatre résultats :

	Indice de Duncan (moyenne)	Indice de Gorard
germanophones	0,79	0,55
italophones	0,65	0,30
ladinophones	0,87	0,82
étrangers	0,58	0,19

TABLE 4 – Valeurs de la moyenne des indices de dissimilarité de Duncan et indice de Gorard pour chacune des communautés

Les valeurs sont d'une manière générale plus basse pour l'indice de Gorard que pour celui de la moyenne des Duncan (Table 4). Ce dernier est calculé à partir des Duncan dans lesquels apparaissent la communauté en question (moyenne de trois indices de Duncan) issue de la Table 3. Entre les deux Duncan (Table 2 et 4), les résultats sont assez proches sauf pour les étrangers, qui ont une différence de 0,37. Concernant l'indice de Gorard, les germanophones possèdent une dissimilarité assez neutre (0,55) du fait qu'ils occupent la moitié du Tyrol italien, dans laquelle les communes sont très homogènes et sont très peu présents dans le Trentin. Les italophones, avec 0,30, possèdent une distribution plus pondérée et par conséquent leur répartition est plus hétérogène dans le Tyrol italien, mais s'ils sont plus présents au Trentin, ils sont en général majoritaires dans les villes du Sud-Tyrol. Quant aux ladinophones, la valeur de l'indice de Gorard, 0,82 est assez proche de celle de Duncan, 0,87. L'interprétation reste la même à savoir les ladinophones sont concentrés dans des communes où ils sont très majoritaires et sont très peu présents dans les autres. Ces deux extrêmes sont représentés par un indice de Gorard très élevé (le plus élevé de tous) et démontre par conséquent une ségrégation importante soit par contrainte (concurrence avec les autres communautés et replis dans leurs «bastions») soit par volonté (faible ouverture sur les autres communautés). Finalement, les étrangers sont ceux qui sont le mieux répartis dans le Tyrol italien. Ils sont présents essentiellement dans les villes (grandes villes) et très partiellement dans certaines communes touristiques. L'indice de Gorard à 0,19 (valeur la plus faible) témoigne de l'hétérogénéité dans leur répartition.

Au final, les italophones et les étrangers sont les communautés qui sont le mieux réparties dans le Tyrol italien et sont également celles qui vivent le plus dans les villes. A l'inverse, les germanophones

et les ladinophones sont beaucoup plus concentrés. Les premiers dans le Sud-Tyrol, les seconds à l'Est du Tyrol italien avec une présence plus forte dans le Sud-Tyrol. Ce sont également des communautés plutôt rurales et cela peut expliquer leur tendance à l'homogénéité. A travers ces résultats et surtout à travers l'amplitude des résultats qui différencie les deux indices, l'indice de Duncan possède des valeurs de ségrégation trop importantes pour pouvoir représenter au mieux ce phénomène. Le fait est qu'il ne prend en compte que deux communautés à chaque fois ce qui ne donne pas d'indication sur le rapport de ces deux communautés par rapport aux autres. Cette façon de calculer aurait peut-être fonctionné si nous n'avions que deux communautés comme c'était le cas pour Duncan (1955) entre les blancs et les non-blancs. Mais cette manière d'interpréter le phénomène ne tient pas compte des autres communautés, car elles sont toutes incluses dans les non-blancs. Ainsi, comme seuls deux critères sont à prendre en compte, la formule peut se présenter comme complète. Dans notre cas, deux communautés sont exclues de la formule à chaque fois et la moyenne de tous les indices dans lesquels est présente la communauté qui nous intéresse ne donne pas de valeur très différentes des résultats issus de la Table 2. L'indice de Gorard, lui, a pu tenir compte de chacune des communautés en comparant, un groupe par la population communale. Le résultat est par conséquent beaucoup plus représentatif de la réalité du phénomène de ségrégation.

4.5 Les premiers résultats spatialisés : l'indice de Morill

Jusqu'ici, les divers calculs montraient une différence entre les communautés sur la base de critères linguistiques et de rapport entre populations. Cependant, bien que ces indices soient une bonne entrée en matière en termes d'analyse de ségrégation spatiale, ils ne possèdent aucune variante spatialisée. Ainsi, peu importe où se situe la commune, la valeur de l'indice reste fixe. C'est pourquoi Morill (1991) propose d'introduire une matrice d'adjacence qui ne prend en compte que la présence ou non d'une commune contiguë. Cette manière de calculer introduit la spatialité et précise la localisation de la ségrégation (les communes non-adjacentes vaudront l'indice de Duncan, soit $ISM_a = IS_a$ si $a_{ij} = 0$). Une remarque est à faire concernant la formule de Morill. En temps normal, c'est sur la base de l'indice de ségrégation de Duncan que devrait se construire l'équation (Morill, 1991). Cependant, comme sa pertinence a été jugée insuffisante, nous avons préféré essayer avec l'indice de ségrégation de Gorard.

	G_a	IS_a	ID_a	ISG_a	$ISM_a (IS_a)$	$ISM_a (ISG_a)$
germanophones	0,92	0,80	0,79	0,55	0,70	0,45
italophones	0,82	0,69	0,65	0,30	0,59	0,21
ladinophones	0,93	0,85	0,87	0,82	0,82	0,79
étrangers	0,28	0,21	0,58	0,19	0,17	0,15

TABLE 5 – Valeurs des indices de Gini, de ségrégation de Duncan, de dissimilarité de Duncan (moyenne), de Gorard, de Morill avec l'indice de ségrégation de Duncan et de Morill avec l'indice de Gorard pour chacun des communautés linguistiques

C'est ainsi que les résultats donnent d'une manière générale des valeurs plus faibles que pour les autres indices. En effet, la Table 5 montre bien la différence entre les valeurs de l'indice de Gorard ISG_a et celui de Morill basé sur ce dernier $ISM_a (ISG_a)$. Comme la formule (8) fait la somme des comparaisons des proportions d'une communauté (nombres décimaux) entre communes adjacentes, la valeur attendue dans la partie de la formule située avant la soustraction sera plus grande que celle située après et par conséquent cette dernière fera chuter la valeur de l'indice de Gorard auquel elle se soustrait. C'est pourquoi $ISM_a (ISG_a) \leq ISG_a$ et $ISM_a (IS_a) \leq IS_a$.

Au niveau des résultats eux-mêmes, il est intéressant de remarquer que seuls les étrangers et surtout les ladinophones possèdent des valeurs sensiblement proches entre les indices. Pour les étrangers, seul l'indice de dissimilarité de Duncan (0, 58) brise la tendance à la dispersion de cette communauté. Pour les ladinophones, la valeur de l'indice ne passe jamais en dessous des 0, 82, nouvellement calculé à travers l'indice de Morill. Pour les italo-phones, les valeurs sont également proches entre elles sauf pour l'indice de Gorard (0, 30) qui sous-estime la concentration de cette communauté. Il en est de même pour les germanophones (0, 55).

Par conséquent, l'indice de Gorard nous semble beaucoup plus pertinent pour calculer l'indice de Morill que celui de Duncan du fait du rééquilibrage qu'il produit à travers la comparaison d'une communauté par le pourcentage de population totale dans cette commune vis-à-vis du Tyrol italien. Ainsi, parmi toutes les communautés, seuls les ladinophones sont considérés comme très ségrégués avec 0, 79 au lieu de 0, 82. Les étrangers ont eux également peu varié (de 0, 17 à 0, 15). Cependant, ils sont plus proches de la dispersion parfaite. Quant aux germanophones et aux italo-phones, ils ont également des valeurs plus faibles qu'attendues, mais les premiers passent pour la première fois en-dessous de la barre des 0, 50. La différence entre 0, 70 et 0, 45 montre tout de même une très grande incertitude dans l'interprétation des résultats. En effet, même l'indice de Gini G_a qui est considéré comme un exemple pour l'analyse des inégalités possède une valeur proche de la concentration parfaite pour les germanophones (0, 92). Cependant, étant donné que les germanophones sont majoritaires (ou présents) dans moins de la moitié des communes, ce résultat peut se justifier. En effet, il y a $i = 217$ communes pour le Trentin et $i = 116$ pour le Haut-Adige pour une population quasiment égale (respectivement de $N = 524832$ et $N = 511688$)! Ainsi, comme les germanophones sont plus présents dans le Haut-Adige et souvent très majoritaires dans des communes peu peuplées, ce résultat de 0, 45 possède une part d'explication. Pour les italo-phones (0, 21), il demeure que le résultat est un peu faible, mais il s'explique de la même manière que pour les germanophones. En effet, les italo-phones sont plus présents dans le Trentin et dans les villes du Haut-Adige. Ils sont donc là où la population communale est majoritaire et sont répartis sur 332 des 333 communes du Tyrol italien.

A titre de complément, si les indices de Gini et de ségrégation de Duncan sont proches (un basé sur l'autre) comme ceux de dissimilarité de Duncan et de Gorard (un basé sur l'autre), l'indice de Morill possède également sa spécificité. Etant donné que l'information de l'adjacence s'ajoute après l'indice de Gorard, le résultat aura tendance à «se fixer». Nous entendons par «se fixer» que plus la formule possèdera d'éléments (nombres décimaux) à la suite de la soustraction à Duncan ou Gorard, plus le résultat de l'indice de Morill se rapprochera de l'indice de Duncan ou Gorard. Ce sera certainement le cas pour l'indice de Wong (1993 ; Chong et Wong, 1998) pour lequel le périmètre et la surface sont ajoutés à la suite de la formule de Morill.

4.6 Un indice qui se veut complet (Wong)

Les caractéristiques de l'indice de Morill offraient la possibilité de mesurer la ségrégation spatiale avec un indice spatialisé. Ainsi, c'est l'adjacence qui permet à une commune de se soustraire à l'indice de Duncan (ou Gorard). Toutefois, un retour à ces indices est obligatoire le cas échéant. Wong (1993) (9) décida d'augmenter la caractérisation de l'indice de Morill en observant les distances des frontières intercommunales et en pondérant la formule avec un rapport périmètre P - surface S placé de telle manière à ce que l'unité (mètre, kilomètre) ne joue aucun rôle, puisque le résultat de cette fraction sera toujours le même

quelle que soit l'unité. Par ailleurs, comme il a été dit dans le chapitre précédent, les résultats de l'indice de Wong changeront peu en fonction de l'indice choisi à l'intérieur de la fonction comme le montre la Table 6. Dans notre cas, il s'agit des indices de ségrégation de Duncan IS_a et de Gorard ISG_a .

	G_a	IS_a	ID_a	ISG_a	$ISM_a(1)$	$ISM_a(2)$	$ISW_a(1)$	$ISW_a(2)$
germanophones	0,92	0,80	0,79	0,55	0,70	0,45	0,80	0,55
italophones	0,82	0,69	0,65	0,30	0,59	0,21	0,68	0,30
ladinophones	0,93	0,85	0,87	0,82	0,82	0,79	0,85	0,82
étrangers	0,28	0,21	0,58	0,19	0,17	0,15	0,20	0,18

TABLE 6 – Valeurs des indices de Gini, de ségrégation de Duncan, de dissimilarité de Duncan (moyenne), de Gorard, de Morill avec l'indice de ségrégation de Duncan (1), de Morill avec l'indice de Gorard (2), de Wong avec l'indice de ségrégation de Duncan (1) et de Wong avec l'indice de Gorard (2) pour chacune des communautés linguistiques

D'une manière générale, tous les indices de Wong sont proches de l'indice choisi pour les calculer avec une différence maximale de 0,01. Ainsi, la précision proposée par l'indice de Wong (1993) à travers l'ajout des distances frontalières, des périmètres et des surfaces ne change guère le résultat dans notre cas, c'est-à-dire que $ISW_a(2) \approx ISG_a$ et $ISW_a(1) \approx IS_a$. En effet, ce qui permet vraiment la fluctuation c'est le choix de l'indice comparant, dans notre cas Duncan ou Gorard. Au final, nous nous retrouvons avec un tableau (Table 6) dans lequel sont inscrits 8 indices, dont les résultats sont tous plus ou moins dissemblables. D'une manière générale, Gini (G_a) est incapable de montrer clairement les inégalités entre les ethnies. Les germanophones sont autant inégaux que les ladinophones alors que leur configuration spatiale est complètement différente. L'indice de dissimilarité de Duncan (ID_a) nécessite la comparaison de deux communautés, ce qui est problématique quand on en a quatre et cela n'a pas de sens de comparer une ethnie à toutes les autres du fait de la perte des caractéristiques (nombre, proportion par commune, proportion au Tyrol italien) que possède chaque groupe. Gorard justement permet de rééquilibrer la formule en comparant une ethnie à l'ensemble des habitants. Tous ces indices manquent cependant de contenu spatiale et c'est l'indice de Morill qui permet de l'inclure à travers la matrice d'adjacence, mais ne tient pas compte des échanges potentiels entre les communes, d'où les apports de Wong. Cependant, ils ne changent pas grand-chose au résultat final presque totalement dépendant de l'indice (de Duncan, de Gorard) se situant dans l'indice de Wong.

4.7 La prise en compte des altitudes

Tous les indices dans la lignée de Duncan ont été calculés et appliqués au Tyrol italien. Cependant, comme ces indices s'appliquaient initialement pour des quartiers de ville, personne n'a jugé nécessaire d'ajouter l'altitude. En effet, même dans une ville les altitudes, bien que peu différenciées en général, peuvent avoir leur utilité. Toutefois, ce n'est pas seulement dans le domaine de la ségrégation spatiale que cette variable n'est pas prise en compte, même si elle n'est pas primordiale. En effet, même dans la géographie physique, il n'est fait que très peu mention de l'altitude. Dans le cadre du Tyrol italien, la configuration géomorphologico-géologique du terrain oblige à prendre en considération cette variable. Toutefois, au regard des résultats, il semblerait que la variation est très faible comparaison avec l'indice de Wong.

Toutefois, le nouvel indice demeure plus pertinent. Plusieurs tentatives ont été faites pour vérifier les résultats finaux à la fois en intégrant les altitudes dans l'indice de compacité (périmètre/surface) et en les sortant (fraction isolée), à la fois avec l'indice de Gorard ou celui de Duncan et à la fois sans les

longueurs de frontières (la présence d'adjacence remplace les longueurs de frontières). Or, nous avons déjà déclaré précédemment que l'indice de Duncan n'était pas adapté. Par ailleurs, sans surprise, les résultats qui lui sont liés ne changent guère comparés à ceux de Wong, car la partie située à la droite de la soustraction de ce dernier indice contient une succession de multiplications de valeurs positives et inférieures à 1. C'est pourquoi l'indice de Duncan présent dans la formule de l'indice de Wong est remplacé par l'indice de Gorard. Nous obtenons ainsi quatre résultats présents dans la Table 7. Ces derniers sont très similaires entre eux. Ainsi, la prise en compte des longueurs de frontières, chères à Wong, ne sont pas nécessaires (ceci permet de faciliter les calculs, car ce sont souvent les logiciels coûteux qui sont capables de les traiter). De plus, s'il avait fallu tenir compte des longueurs de frontières, il aurait fallu privilégier avant tout les lieux de passage. Le Tyrol italien étant une région montagneuse, la communication d'une commune à l'autre peut être empêchée pour des raisons géomorphologiques. Ainsi, il faudrait prendre en compte les passages possibles qui peuvent être par exemple des routes, des cols ou des rails de chemins de fer. Cependant, en n'observant que les voies physiquement franchissables, la proportion de lieux traversables entre deux communes, divisée par le périmètre de la commune (comme l'indique l'indice de Wong) sera toujours très faible. Par ailleurs, il devient très compliqué de tenir compte seulement des passages possibles entre les 333 communes. Ainsi, la simple déclaration d'adjacence (comme montre l'indice de Morill) suffit à rendre compte de la réalité de la ségrégation spatiale sur le terrain. Elle a d'ailleurs tendance à augmenter le résultat pour le cas du Tyrol italien (0,19 au lieu de 0,18 pour les étrangers). Quant aux altitudes, leurs places ne jouent pas un rôle très déterminant dans l'équation au vu de sa configuration (assez longue) et du nombre de communes. Pour permettre une variation plus forte du résultat, placer les altitudes dans l'indice de compacité périmètre/surface de la forme $\frac{P*A}{S}$ est plus adapté. La place qu'occupe l'altitude dans le numérateur permet d'augmenter la contrainte (le résultat de l'indice aura tendance à augmenter) sur l'équation. L'altitude permet également d'une certaine manière de se passer des longueurs intercommunales dans un cadre montagneux. Sortir les altitudes de l'indice de compacité est également possible, mais le résultat sera beaucoup plus proche de l'indice de Wong. Cette façon d'utiliser les altitudes permet dans tous les cas d'éviter une division par zéro.

	ISKB _a (13)	ISKB _a (14)	ISKB _a (15)	ISKB _a (16)
germanophones	0,54	0,55	0,55	0,55
italophones	0,30	0,30	0,30	0,30
ladinophones	0,82	0,82	0,82	0,82
étrangers	0,18	0,19	0,19	0,19

TABLE 7 – Valeurs de l'indice de ségrégation de Wong avec les altitudes comprises dans l'indice de compacité (périmètre/surface) avec les longueurs de frontières (13), l'indice de ségrégation de Wong avec les altitudes non comprises dans l'indice de compacité et sans les longueurs de frontières (14), l'indice de Wong sans longueur de frontières et avec altitude comprises dans l'indice de compacité (15), l'indice de Wong avec les longueurs de frontières et avec les altitudes non comprises dans l'indice de compacité (16) pour chacune des communautés linguistiques. Tous ces indices contiennent l'indice de Gorard

D'une manière générale, le remplacement de l'indice de Wong par celui de Gorard, la non-nécessité des longueurs interfrontalières et l'ajout de l'altitude génèrent un indice ISKB_a capable de tenir compte d'une très grande partie des spécificités territoriales à la fois en termes de géographie physique, mais également humaine.

4.8 Les indices de Moran

A présent que l'on a évalué les résultats de différents indices de ségrégation spatiale, voyons comment réagissent les territoires avec les autocorrélations spatiales. Pour rappel, les valeurs des indices de Moran

locaux sont généralement situées entre -1 et 1. Pour pouvoir représenter les I de Moran locaux (Figure 5), nous avons décidé de sélectionner 4 à 5 classes à savoir pour des valeurs inférieures à -1, comprises entre -1 et -0.5, comprises entre -0.5 et 0.5, comprises entre 0.5 et 1, supérieures à 1. Pour ce faire, nous avons fixé la valeur de l'«âge du réseau» à 1 et 2, correspondant à la comparaison d'une commune respectivement avec ses voisines et les voisines de ses voisines.

Pour les germanophones (Table 8), l'indice d'autocorrélation global δ est de 0,88 au temps $t = 1$ et de 0,82 au temps $t = 2$, ce qui montre une autocorrélation très forte entre les communes. Ces résultats sont extrêmement significatifs, car les variables centrées réduites valent respectivement 29,28 et 37,73. Ainsi (Figure 5a et 5b), il semblerait que l'essentiel de la région du Haut-Adige, dans laquelle les germanophones sont majoritaires, possède des indices de Moran locaux supérieurs à 1 ainsi que quelques indices supérieurs à 0,5. Une partie de la zone ladine du Tyrol du Sud ainsi que les communes de Fortezza/Franzensfeste, de Merano/Meran, de Bolzano/Bozen et les communes situées à la pointe sud de la sous-région (contiguës à Bolzano) possèdent une autocorrélation spatiale locale nulle au temps $t = 1$. Ainsi, il est intéressant de voir qu'entre les deux régions, il y ait une limite claire de communes sans autocorrélation spatiale (comprises entre -0,5 et 0,5). La répartition des germanophones dans ces communes peut être considérée par conséquent comme aléatoire. Au temps $t = 2$, une augmentation du nombre de communes sans autocorrélation spatiale dans la zone entre le Trentin et le Haut-Adige est constatée, renforçant ainsi l'idée d'un «mur» entre les deux sous-régions. Dans le Trentin, pour les deux temps, l'autocorrélation spatiale est comprise entre 0,5 et 1. Malgré et à cause du nombre très important d'absence germanophone dans 183 des 217 communes, le Trentin possède une autocorrélation spatiale positive. Par ailleurs, des îlots d'autocorrélations spatiales négatives apparaissent dans les communes trentines de Castelfondo, Fierozzo, Frassilungo, Palù del Fersina et Luserna au temps $t = 1$. La première ne possède pas de germanophones, les trois suivantes montrent l'isolement de la population dite des «Mochènes» par rapport aux communes voisines, la dernière montre l'isolement de la population dite des «Cimbres». Au temps $t = 2$ s'ajoutent les communes voisines de Castelfondo, qui ont toutes la particularité d'avoir des territoires discontinus (entrecoupés entre elles) et la commune de Marebbe à l'Est du Tyrol du Sud (zone ladinophone) dans les autocorrélations spatiales négatives. Toutefois, l'autocorrélation spatiale ne descend pas plus bas que -1 pour les deux «âges du réseau». Ceci indique que dans l'ensemble sauf entre le Trentin et le Haut-Adige, la répartition des germanophones n'est pas due au hasard mais bien à une continuité territoriale.

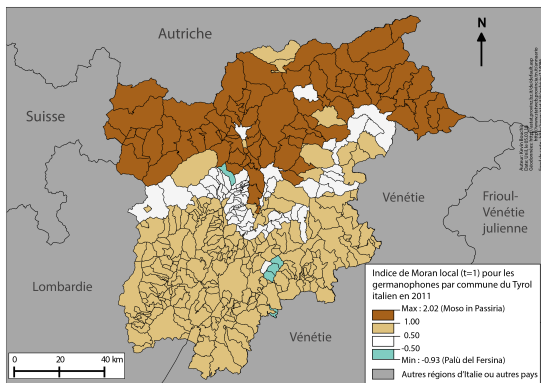
Pour les italophones (Table 8), les indices d'autocorrélation spatiale globaux sont comme pour les germanophones, très élevés. Ils se montent à 0,92 au temps $t = 1$ et à 0,85 au temps $t = 2$ et sont parfaitement acceptables, car les valeurs de la variable centrée réduite valent respectivement de 31,30 et de 41,14. D'une manière générale (Figure 5c et 5d), les autocorrélations spatiales locales pour les italophones sont similaires voire très similaires pour les deux temps. Cependant, les valeurs les plus élevées sont essentiellement dans le Tyrol du Sud, région où les italophones ne sont pas du tout majoritaires. D'ailleurs, tous les indices dans cette région pour les italophones sont positifs et même très positifs vu qu'ils dépassent 1. Il en est de même pour leur répartition dans les zones ladiniques de l'Est. De plus, le fameux «mur» de communes sans autocorrélation spatiale, présent chez les germanophones, existe aussi chez les italophones. Il s'est nettement renforcé au temps $t = 2$ s'étendant ainsi depuis l'Ouest vers l'Est du Tyrol italien et incluant la commune citadine de Merano/Meran. Enfin, les communes de Fierozzo, Frassilungo, Palù del Fersina et Luserna, contenant les minorités mochènes et cimbres, possèdent un indice

d'autocorrélation spatiale local négatif, sans surprise. Ceci fait bien ressortir la discontinuité très forte entre ces quatre communes entourées de voisins positifs voire très positifs. Seule Sant'Orsola Terme possède une autocorrélation spatiale nulle. Le fait que le Tyrol du Sud soit plus positif que le Trentin passe pour un paradoxe. Il peut être expliqué sans doute par la présence du «mur» et des quelques îlots italophones dans le Tyrol du Sud. Par ailleurs, quel que soit le temps ou la communauté (germanophone ou italoophone), le «mur» est constitué en grande majorité de communes du Trentin. L'essentiel des communes du Tyrol du Sud participant à cet obstacle se situe dans la pointe sud de cette sous-région. Le fait que Bolzano/Bozen fasse partie de ce «mur» devient logique à cause de sa plus forte composition italoophone.

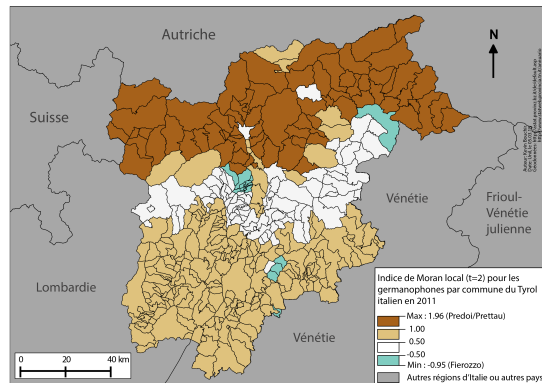
Pour les ladinophones (Table 8), l'indice d'autocorrélation spatiale global se retrouve un peu diminué comparé aux germanophones et aux italophones. Il vaut 0,74 au temps $t = 1$ et chute à 0,61 au temps $t = 2$ pour une variable centrée réduite respective de 23,01 et 25,75, ce qui rend compte également de la grande significativité du résultat. Pour les indices de Moran locaux au temps $t = 1$ (Figure 5e et 5f), la situation est très différente des deux autres communautés. D'une manière générale, deux zones ressortent clairement à savoir l'Est et le centre du Trentin Haut-Adige. En commençant par l'Est, zone où ils sont majoritaires, les communes possèdent un indice supérieur à 1. Il devient ensuite intéressant de voir que ces communes sont isolées, entourées par des unités spatiales ayant un indice inférieur à $-0,5$, voire à -1 . Ces territoires sont eux-mêmes entourés d'une couronne de communes d'autocorrélation spatiale essentiellement comprises entre $-0,5$ et $0,5$. De toute évidence, cette deuxième couronne renforce l'isolat des ladinophones dans leur zone d'habitat et laisse à penser qu'il n'y a pas ou peu de communication avec les régions voisines. En effet, l'essentiel du Tyrol italien possède une autocorrélation spatiale comprise entre $0,5$ et 1 pour ces communautés. Par ailleurs, la deuxième zone d'intérêt met en avant plusieurs communes avec des indices supérieurs à 1 . Ce sont des communes ayant une petite minorité de ladins. Elles sont entourées de communes ayant un indice supérieur à $0,5$, elles-mêmes entourées d'entités avec un indice d'autocorrélation spatiale nul. Au-delà, les indices des communes sont à nouveau supérieures à $0,5$ (essentiel du Tyrol italien). Au temps $t = 2$, les couronnes sont devenues plus étendues : la couronne négative de l'Est est devenue plus grande et plus négative, et les couronnes sans autocorrélation spatiale ont également grandi faisant ainsi une continuité spatiale entre les deux zones ladinophones. La couronne du centre a perdu bon nombre de communes très positives qui sont devenues positives. Par ailleurs, la commune de Bolzano/Bozen, située entre les deux couronnes, est restée positive. Son caractère urbain peut laisser penser que les ladinophones y sont bien implantés, ce qui n'est pas le cas. Même si elle est la 17e commune où ils sont le plus présent, Bolzano/Bozen possède un indice d'autocorrélation spatiale local de $0,51$ se situant ainsi dans la classe des positifs. Cependant, si elle avait été placée dans la classe sans autocorrélation spatiale, elle aurait eu la valeur la plus élevée.

	Moran $t = 1$	z-score $t = 1$	Moran $t = 2$	z-score $t = 2$
germanophones	0,88	29,28	0,82	37,73
italophones	0,92	31,30	0,85	41,14
ladinophones	0,74	23,01	0,61	25,75
étrangers	0,67	19,44	0,56	23,13

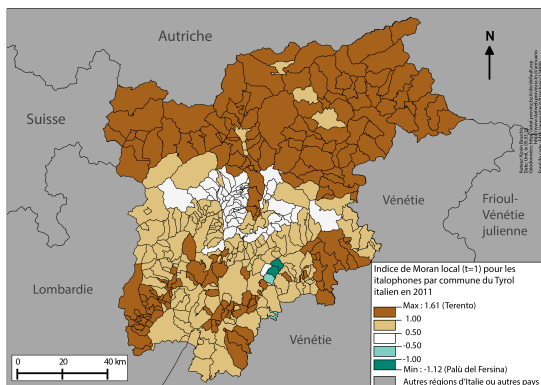
TABLE 8 – Valeurs de l'indice de Moran global et de la variable centrée réduite (z-score) aux temps $t = 1$ et $t = 2$



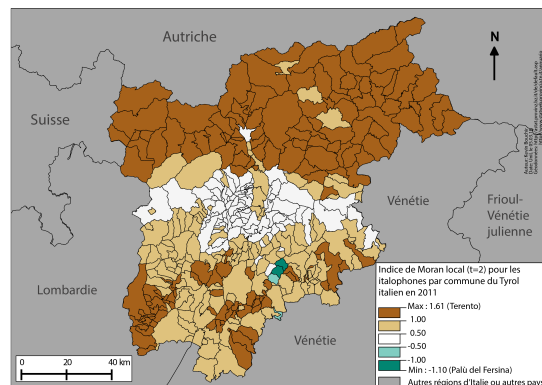
(a) Germanophones au temps t = 1



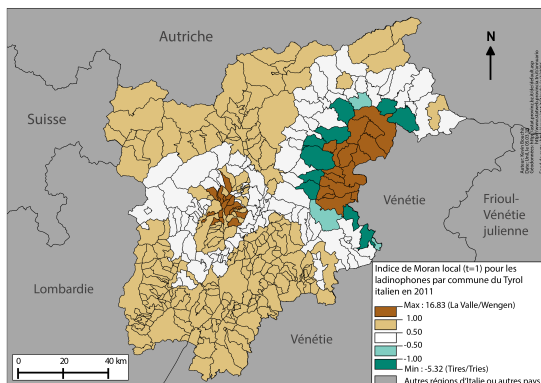
(b) Germanophones au temps t = 2



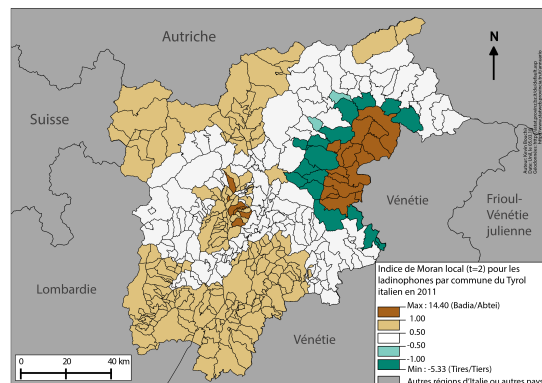
(c) Italophones au temps t = 1



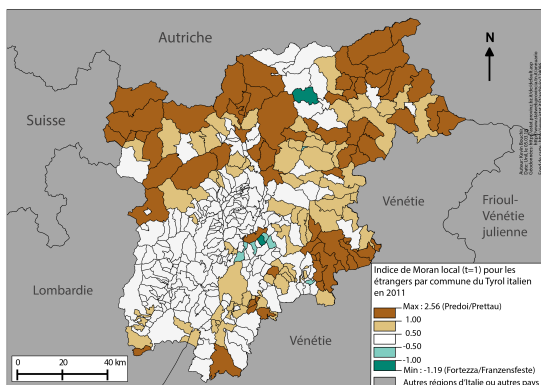
(d) Italophones au temps t = 2



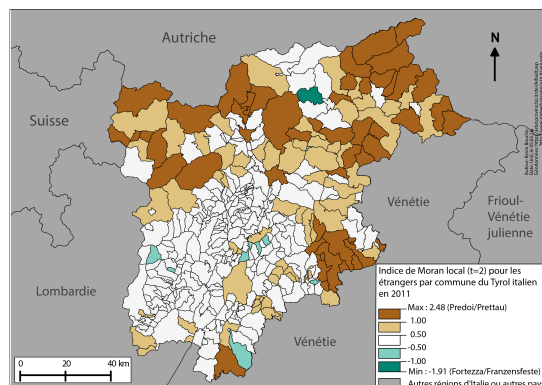
(e) Ladinophones au temps t = 1



(f) Ladinophones au temps t = 2



(g) Etrangers au temps t = 1



(h) Etrangers au temps t = 2

FIGURE 5 – Autocorrélation spatiale pour chaque communauté (aux temps t=1 et t=2)

Pour synthétiser ce sous-chapitre, il est possible de confirmer que plus l'autocorrélation est élevée et plus la variable centrée réduite augmente également, autrement dit plus le I de Moran global est grand et plus il est significatif. De plus, d'une manière générale, les indices locaux possèdent des valeurs extrêmes beaucoup plus grandes au temps $t = 2$ qu'au temps $t = 1$. Si pour $t = 1$, elles ne dépassent pas 2; pour $t = 2$ elles excèdent 10 pour les ladinophones par exemple. De plus, la continuité spatiale pouvait mieux s'apprécier en comparant les voisins des voisins au temps $t = 2$. Par ailleurs, pour les classes intermédiaires, il y a une tendance pour les communes isolées à devenir négatives ou du moins à une diminution de la valeur du I de Moran local. Il est également intéressant de voir que germanophones et italophones possèdent une représentation cartographique assez similaire avec un gradient plutôt Nord-Sud alors que les ladinophones possèdent plutôt un gradient Est-Ouest et plus précisément circonscrit aux communes dans lesquelles ils sont majoritaires. Quant aux étrangers, il est plus difficile de trouver une continuité spatiale bien que le Tyrol du Sud présente des zones d'autocorrélation spatiale plus forte que le Trentin, ce dernier ne possédant pas vraiment d'autocorrélation spatiale sauf dans les communes de l'Est où elles sont positives et au centre où elles sont négatives.

4.9 Les autocorrélations spatiales avec les cadrans de Moran

Les résultats des autocorrélations spatiales pour chaque communauté ont été analysés et décortiqués. Cependant, il serait intéressant de pouvoir pousser la recherche plus loin en essayant d'observer la visualisation des graphes lorsque les quatre variables suivantes sont associées : la variable x (vecteur contenant le pourcentage d'une communauté a), la variable Wx (vecteur des moyennes des voisines d'une commune i située dans le vecteur x), la variable R (issue de la différence entre Wx et x . $R = x - Wx$), δ_i (les I de Moran locaux). La première tentative est d'observer la relation entre Wx et x , soit la variante qui va permettre de générer les cadrans de Moran. Ce chapitre ne prendra en compte que les germanophones.

Ainsi, le graphe (Figure 6) montre une séparation assez stricte entre des communes totalement concentrées dans la zone $x = 0$ et d'autres placées proche de $x = 1$. Dans les cadrans de droite se trouvent les communes dans lesquelles la population germanophone est supérieure à la moyenne communale de 0,29. Il s'agit pour l'essentiel de communes situées dans le Tyrol du Sud pour le cadran en haut à droite (high-high) et d'îlots germanophones du Trentin pour celui en bas à droite (high-low) (6 communes dont 5 appartenant aux Cimbres et aux Mochènes). La commune de Salorno/Salurn apparaît également dans ce cadran, car son nombre de germanophones est d'environ 37% et est frontalière au Trentin. Le cadran en bas à gauche (low-low) contient pour l'essentiel des communes du Trentin. Il est intéressant de remarquer que Bolzano/Bozen, chef-lieu du Tyrol du Sud et îlot italophone dans le Tyrol du Sud fasse partie de ce cadran. Les communes ladines, telles qu'Ortisei/Sankt Ulrich, se situent également dans ce cadran. Le cadran en haut à gauche (low-high) quant à lui présente 32 communes (dont seulement 4 sont du Tyrol du Sud), toutes limitrophes ou proches de la limite entre le Trentin et le Tyrol du Sud.

Au niveau de la droite de Moran, il est intéressant de voir que les principales villes du Tyrol du Sud, zone essentiellement germanophone, sont toutes quasiment sur cette ligne. Il demeure dans cette représentation quelques valeurs aberrantes. Elles sont définies visuellement comme étant détachées des partitionnements environnants. La commune de Fortezza/Franzensfeste, par exemple, est située dans le cadran high-high. Bien qu'elle soit en majorité germanophone (près de 60%), la population italophone recouvre quasi à elle seule le reste, soit plus d'un tiers de la population. Étant donné que cette commune est

entourée de voisins ayant un plus fort taux de germanophones, elle apparaît comme une valeur aberrante sur le diagramme de Moran (Figure 6).

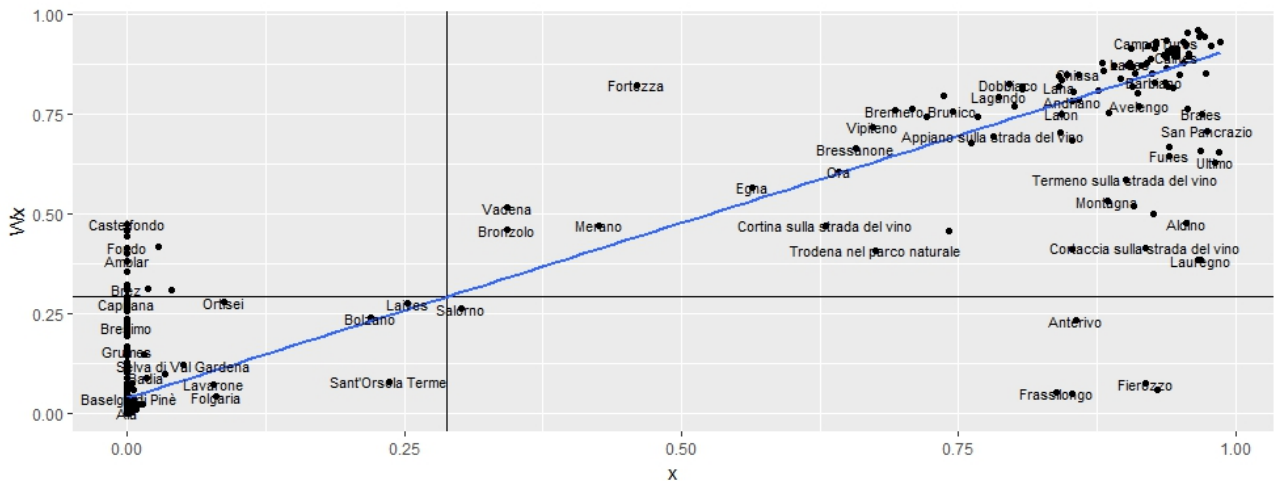


FIGURE 6 – Graphe montrant la relation la variable Wx = «moyenne des proportions de germanophones des voisins d'une commune i » et la variable x = «proportion de germanophones»

En analysant à présent l'évolution des I de Moran locaux δ_i par rapport à la proportion de germanophones dans les communes x (Figure 7), il est possible d'observer que la plupart des valeurs vis-à-vis de l'analyse précédente (Figure 6) n'ont que très peu changé de place. Il demeure toujours un contraste fort entre l'essentiel des communes du Trentin et celles du Tyrol du Sud, situées chacune aux extrêmes du graphe. Le plus surprenant vient des communes situées dans le partitionnement du Trentin (low-low). Il semblerait que dans le nouveau graphe, il y ait eu une forme d'inversion des places de ces entités. Ainsi, celles qui se situaient en bas du graphe se retrouvent en haut et vice versa. Pour le Tyrol du Sud, la situation n'a pas vraiment changé. Une différence notable concerne la commune de Fortezza/Franzensfeste qui est passablement descendue au niveau des ordonnées du nouveau graphe, son autocorrélation spatiale locale étant proche de 0. La commune de Folgaria, à l'inverse, est montée en termes d'ordonnées avec une autocorrélation spatiale locale proche de 0,7. De par la similarité partielle qui existe entre les figures 6 et 7, un lien certain demeure entre ces deux représentations qui peuvent être vues comme complémentaires.

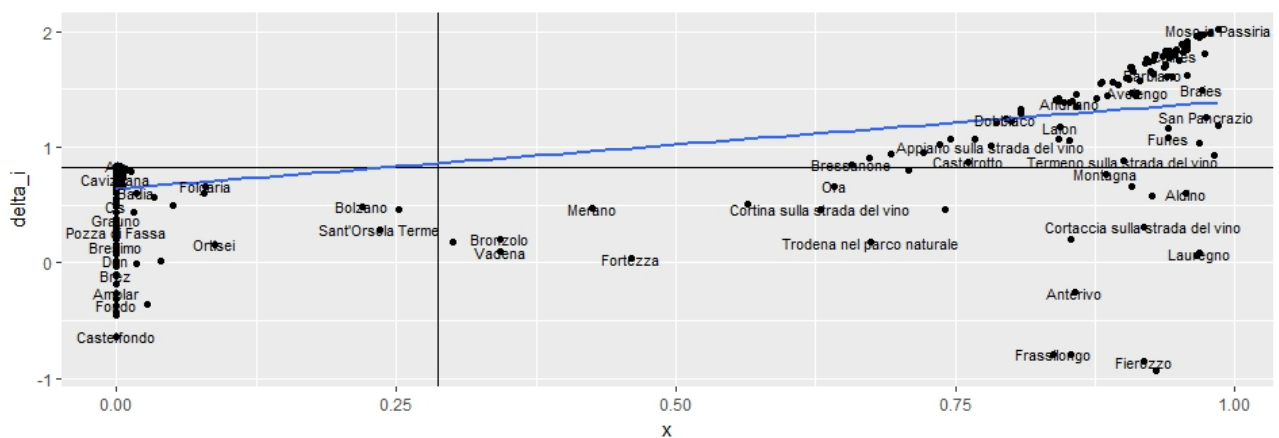


FIGURE 7 – Graphe montrant la relation entre les δ_i au temps $t = 1$ et la variable x = «proportion de germanophones»

Si, à présent, il fallait comparer les δ_i à la moyenne des voisins de chaque commune Wx , quelques transformations intéressantes apparaîtraient (Figure 8). Il semblerait que par rapport au graphe de la

figure 6, les communes aient conservé un certain ordre d'apparition en regardant l'axe des abscisses. Une petite exception peut être faite pour les communes qui étaient situées dans le cadran high-low de la Figure 6. Dans le nouveau graphe (Figure 8), elles se situent entre le partitionnement du Trentin et celui du Tyrol du Sud. Il est intéressant de voir qu'en défilant l'axe des abscisses, les communes du Trentin tendent à diminuer la valeur de leur I de Moran locale puis une augmentation continue se fait avec les communes du cadran high-low, qui correspondent à quelques communes frontalières entre le Trentin et le Tyrol du Sud) suivie par celles du Tyrol du Sud (high-high). Les îlots de communes germanophones dans le Trentin sont clairement séparés de cette tendance et forment un partitionnement entre elles (low-low). La commune de Fortezza/Franzensfeste, également, se retrouve un peu isolée de la tendance générale du graphe. En outre, ce dernier possède l'avantage de montrer, à quelques exceptions près, la répartition «géographique» des communes à travers des variables statistiques. En effet, en observant le graphe tourné à 90°, c'est-à-dire que la droite du graphique est observée comme si elle était le sommet de la représentation, les communes du Tyrol du Sud se trouvent au «Nord» (si le sommet du graphique est défini comme tel) et le Trentin au «Sud».

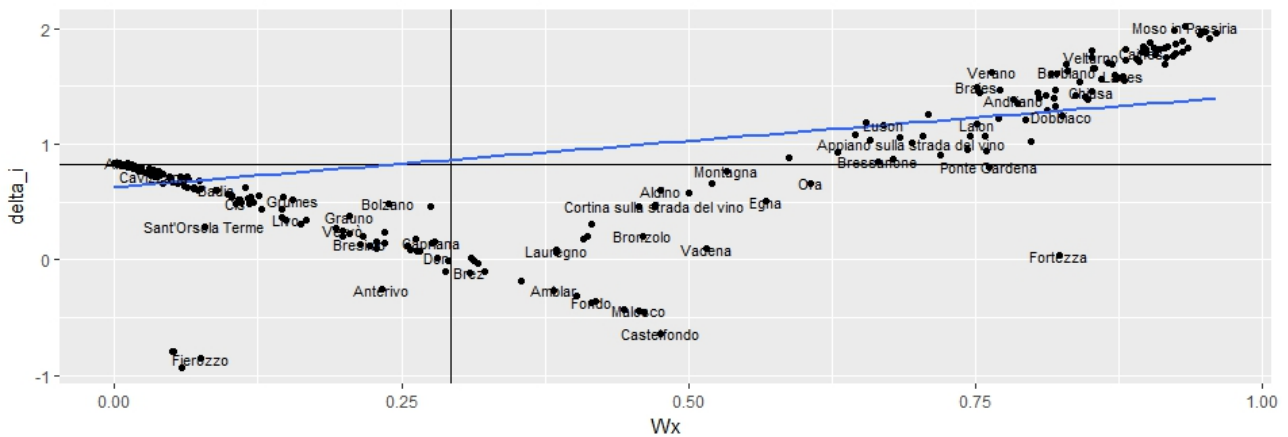


FIGURE 8 – Graphe montrant la relation entre les δ_i au temps $t = 1$ et la variable $Wx =$ «proportion de germanophones»

Finalement, il reste l'analyse du rapport entre les δ_i , et la différence entre la proportion de germanophones d'une commune et la moyenne des proportions de germanophones des communes voisines R (Figure 9). A première vue, il semblerait qu'il y ait une forme d'effet miroir entre la Figure 9 et la Figure 8. Dans la nouvelle figure, les valeurs des I de Moran des communes du Trentin sont en constante croissance en commençant par la commune de Castelfondo (dernière commune du Trentin dans le graphe précédent) (high-high). Pour les communes du Tyrol du Sud, le phénomène est sensiblement similaire. Dans l'ancien graphe, la commune de Moso in Passiria était située parmi les valeurs les plus élevées (high-low). Dans le nouveau graphe, elle est parmi les premières à apparaître pour le Tyrol du Sud. Quant aux îlots germanophones du Trentin, ils se retrouvent dans la zone des communes du Tyrol du Sud (high-low), mais complètement séparés du partitionnement principal. Autre fait marquant, la limite entre les communes du Trentin croissantes et celles du Tyrol du Sud décroissantes est proche de 0 au niveau de la variable R .

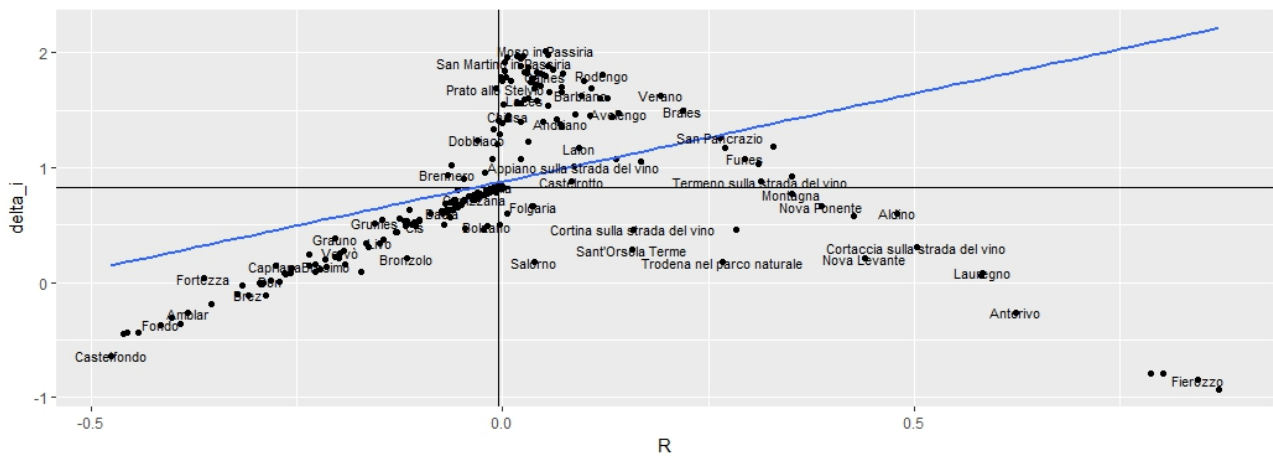


FIGURE 9 – Graphe montrant la relation entre les I de Moran locaux au temps $t = 1$ et la différence R entre la proportion de germanophones x et la moyenne des proportions de germanophones des communes voisines Wx

4.10 La visualisation comme outil éducatif

Revenant à présent sur la deuxième thématique de ce mémoire à savoir la diffusion d'un savoir complexe à des personnes lambda. Pour rappel, deux visualisations avaient été générées, une sur les autocorrélations spatiales et l'autre sur l'Indice de Herfindahl. Le but étant de vérifier si la méthode de visualisation utilisée pouvait aider à comprendre les autocorrélations spatiales. Pour cela, la deuxième visualisation sur l'indice de Herfindahl, jugée plus facile à comprendre, devait montrer si les outils mis en place pouvaient s'appliquer à toute forme de statistiques. Ainsi, 20 personnes (Table 9) ont été sondées, 10 femmes et 10 hommes et parmi eux 16 jeunes adultes et 4 seniors, répartis équitablement entre les visualisations. Les résultats sont présentés sous la forme de boîtes à moustache dans lesquelles la moyenne a été ajoutée (Figure 10).

Sexe	Femmes		Hommes		Total
	Jeunes	Seniors	Jeunes	Seniors	
Visualisation Herfindahl	4	1	4	1	10
Visualisation Autocorrélations	4	1	4	1	10
Total	8	2	8	2	20

TABLE 9 – Répartition du nombre de testeurs de chacune des applications (indice de Herfindahl et autocorrélations spatiales locales) par sexe et par âge

D'une manière générale, une différence majeure apparaît dans ces boîtes à moustache : la visualisation des autocorrélations spatiales (Figure 10b) s'étend de 3 à 10 au niveau des notes attribuées, tandis que la visualisation de l'indice de Herfindahl (Figure 10a) est comprise entre 6 et 10. Par ailleurs, la moyenne la plus basse s'élève à 7,9, soit proche de 8 et est attribuée à la compréhension du diagramme ternaire proposé dans l'application sur les autocorrélations spatiales, alors qu'elle monte à 8,5 pour l'application sur l'indice de Herfindahl. En outre, la grande majorité des notes étant comprises entre 7 et 10 et les médianes ne descendant jamais en-dessous de 8, les applications ne nécessiteront pas de modification majeure et par conséquent elles peuvent être considérées comme étant validées à première vue. Toutefois, en utilisant le test de Student (Table 10), nommé test t, entre les deux applications, il est intéressant de voir que seule la variable concernant la compréhension des statistiques possède une valeur p inférieure à $\alpha = 0,05$, seuil pour ne pas rejeter l'hypothèse H_0 . En partant du principe que l'hypothèse

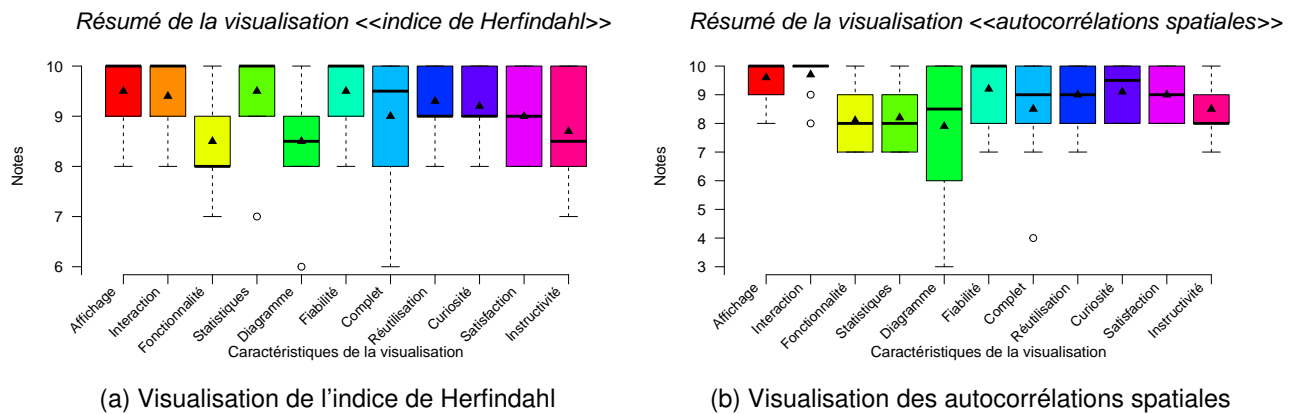


FIGURE 10 – Statistiques descriptives des expérimentations sur les visualisations. Chaque triangle représente la moyenne de la boîte à moustache

H_0 de cette variable se traduit comme «il y a autant de probabilité de comprendre les statistiques d’une visualisation que de l’autre», nous observons qu’avec 0,0037 l’hypothèse est à rejeter, car le test est «statistiquement significatif». Par conséquent, il demeure une forme d’inégalité dans la compréhension des statistiques sophistiquées dans leur calcul entre les deux applications. La différence de moyennes entre cette variable pour les deux applications est également la plus grande avec $\Delta\bar{x} \in | -1,3 |$ sur une différence allant de 0 à ± 9 . Ceci pose problème, car la compréhension des statistiques sophistiquées dans leur calcul représentent l’objectif principal des applications. Par ailleurs, la plus grande différence de variance entre les visualisations (4,27) est liée à la compréhension du diagramme ternaire. Ceci montre bien que les testeurs ont eu une appréhension assez différentes sur ce point quelle que soit la visualisation qu’ils ont obtenue. En outre, afin d’obtenir des résultats plus fins, il est nécessaire d’observer le phénomène, notamment de la compréhension des statistiques sophistiquées à travers les groupes, à travers la différence entre hommes et femmes, puis entre seniors et jeunes adultes pour mieux cibler laquelle de ces populations a d’une manière générale eu plus de peine. Comme l’hypothèse H_0 n’est pas à rejeter pour les diagrammes ternaires, les analyses suivantes resteront focalisées sur la compréhension des statistiques sophistiquées.

	Aff.	Int.	Fonc.	Stat.	Dia.	Fia.	Comp.	Réut.	Curi.	Sat.	Inst.
Test	0,56	1,15	-0,88	-3,88	-0,60	-0,71	-1,05	-0,82	-0,29	0	-0,56
V. p	0,59	0,28	0,40	0,0037	0,56	0,50	0,32	0,43	0,78	1	0,59
$\Delta\bar{x}$	0,1	0,3	-0,4	-1,3	-0,6	-0,3	-0,5	-0,3	-0,1	0	-0,2
$\Delta\text{var}(x)$	-0,23	-0,26	0,044	0,34	4,27	0,78	1,83	0,65	0,37	0	-0,18
Testeurs	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10

TABLE 10 – Test de Student pour chaque variable avec la valeur du test, la valeur p, la différence de moyenne $\Delta\bar{x}$, la différence de variance $\Delta\text{var}(x)$ et le nombre de testeurs pour chaque variable d’appréciation par visualisation (Herfindahl/autocorrélation spatiale)

Au niveau de la différence hommes-femmes (Figure 11), quatre boîtes à moustaches ont été générées montrant à chaque fois le sexe (homme, femme) et le type de visualisation (autocorrélations spatiales, indices de Herfindahl). Pour les femmes, le minimum ne descend jamais en-dessous de 6. Il est intéressant de remarquer entre les deux visualisations que l’indice de Herfindahl (Figure 11c) semble avoir été bien compris avec une moyenne de 9,2 et une médiane de 10 alors qu’elles tombent respectivement à 7,6 et 7 pour les autocorrélations spatiales (Figure 11a). Cependant, c’est la compréhension du diagramme ternaire qui montre une plus grande dispersion des avis allant 6 à 10 et dont les moyenne et médiane valent

8 toutes deux. Pour les hommes, l'interprétation est un peu différente. En effet, les deux variables vont de 8 à 10 pour la visualisation de l'indice de Herfindahl (Figure 11d) et de 3 à 10 pour celle des autocorrélations spatiales (Figure 11b). C'est à nouveau le diagramme ternaire qui possède la plus grande dispersion et en particulier pour l'application des autocorrélations spatiales. Il existe en effet un minimum à 3, mais la moyenne vaut 7,8 et la médiane 10. Cette dispersion s'explique par le fait que les testeurs ont mis des notes assez différentes, mais la majorité a inscrit 10 et une personne 3. Au niveau de la compréhension des statistiques, les autocorrélations spatiales ont 8,8 de moyenne et 8 de médiane et pour l'indice de Herfindahl respectivement 9,8 et 10, ce qui démontre un écart assez important respectivement de 1 et de 2. Cependant, ces écarts restent moins importants que pour les femmes.

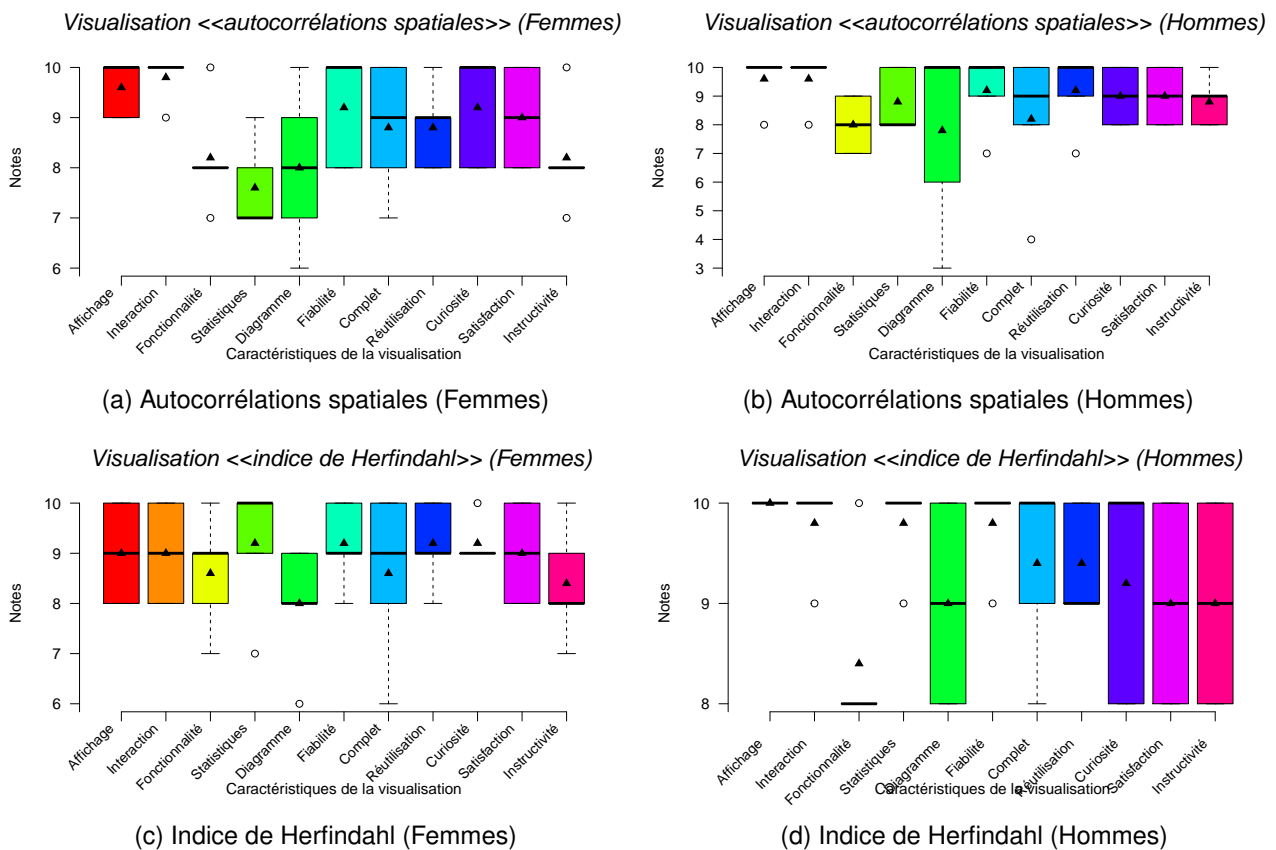


FIGURE 11 – Différences hommes-femmes dans les visualisations des autocorrélations spatiales et de l'indice de Herfindahl. Chaque triangle représente la moyenne de la boîte à moustache

En outre, des tests de Student (Table 11) ont également été effectués pour ces populations. Il est basé sur quatre comparaisons, à savoir entre hommes et hommes entre les deux visualisations, entre femmes et femmes entre les deux visualisations et entre hommes et femmes pour chacune des deux visualisations. La première chose qui est observable est qu'à aucun moment la valeur p ne descend en-dessous de 0,05. Le résultat le plus bas concerne la comparaison des deux visualisations entre les femmes à hauteur de 0,078, ce qui ne permet juste pas de rejeter l'hypothèse H_0 qui concerne la même probabilité de compréhension des statistiques sophistiquées dans leur calcul.

Il est vrai que la différence entre hommes et femmes a apporté certains éléments de réponse, mais la différence entre jeunes adultes et seniors (Figure 12) offre un complément. Pour rappel, les seniors sont les populations d'âge supérieur à 50 et les jeunes adultes d'âge inférieur à 35. Il est tout de même nécessaire de préciser que les représentants de ces populations ne sont pas équivalents. En effet, il

	Hommes H-A	Femmes H-A	Herf. H-F	Auto. H-F
Test	-1,58	-2,36	0,88	-1,5
V. p	0,19	0,078	0,43	0,21
$\Delta\mu$	-1	-1,6	0,6	-1,2
Δvar_x	1	-0,9	-1,5	-0,4
Testeurs	5/5	5/5	5/5	5/5

TABLE 11 – Test de Student pour la compréhension des statistiques avec la valeur du test, la valeur p, la différence de moyenne, la différence de variance et le nombre de testeur (selon soit la visualisation, soit le sexe en fonction de ce qui est comparé) pour la différence entre les hommes, entre les femmes, entre les sexes pour Herfindahl, entre les sexes pour les autocorrélations spatiales

existe sur les 20 testeurs, 16 jeunes adultes et seulement 4 seniors tous répartis équitablement entre les deux visualisations. Ces comparaisons présentent ainsi le risque que le nombre de seniors ne soit pas suffisamment élevé pour offrir une réponse au rejet de l'hypothèse H_0 . A première vue, pour les jeunes adultes (Figure 12a), les résultats s'étendent de 4 à 10 et pour les seniors (Figure 12b) de 3 à 10. Or, ce sont les résultats pour les autocorrélations spatiales. Pour l'indice de Herfindahl, l'intervalle est plus restreint allant respectivement de 7 à 10 (Figure 12c) et de 6 à 10 (Figure 12b). Il est ainsi déjà clair que les autocorrélations spatiales ont donné plus de fils à retordre que l'indice de Herfindahl. Cependant, ces intervalles sont importants, car il existe toujours un seul individu qui marque le minimum. Concernant la compréhension des statistiques sophistiquées dans leur calcul, la dispersion est plus grande pour les autocorrélations spatiales pour les jeunes adultes avec une moyenne de 8,3 et une médiane de 8 que pour l'indice de Herfindahl avec respectivement 9,5 et 10. Pour les seniors, l'application des autocorrélations spatiales possède une moyenne et une médiane de 8 et celle de l'indice de Herfindahl de 9,5. Le plus flagrant reste la compréhension du diagramme ternaire par les seniors. En effet, elle est à 4,5 en moyenne et médiane pour les autocorrélations spatiales et à 8 pour l'indice de Herfindahl.

En observant à présent la valeur p des jeunes adultes pour la compréhension des statistiques sophistiquées dans leur calcul (Table 12), nous remarquons qu'elle se situe à 0,049. En arrondissant, nous obtenons 0,05, soit le seuil $\alpha = 0,05$ nécessaire pour ne pas rejeter l'hypothèse H_0 . Comme la valeur p est en-dessous du seuil α (c'est-à-dire $p < \alpha$), nous considérons qu'il faut rejeter l'hypothèse de la même probabilité de compréhension des statistiques sophistiquées dans leur calcul. Pour les seniors, la valeur p se situe à 0,20, soit bien en-dessus de la limite de rejet.

	Jeunes H-A	Vieux H-A	Herf. J-V	Auto. J-V
Test	-2,38	-3,00	0	0,55
V. p	0,05	0,20	1	0,60
$\Delta\bar{x}$	-1,5	-1,5	0	0,25
Δvar_x	2,21	0,5	0,86	1,64
Testeurs	8/8	2/2	16/4	16/4

TABLE 12 – Test de Student pour la compréhension des statistiques avec la valeur du test, la valeur p, la différence de moyenne pour la différence entre les jeunes adultes entre les deux visualisations, entre les seniors pour les deux visualisations, entre les jeunes adultes et les seniors pour la visualisation de l'indice de Herfindahl, entre les jeunes adultes et les seniors pour la visualisation des autocorrélations spatiales. Les tests des deux dernières colonnes sont appariés

Pour les deux dernières colonnes de la Table 12, comme le nombre de participants est différent, le test t utilisé est apparié. Cependant, la valeur p n'étant pas en-dessous du seuil $\alpha = 0,05$, l'hypothèse H_0 est conservée.

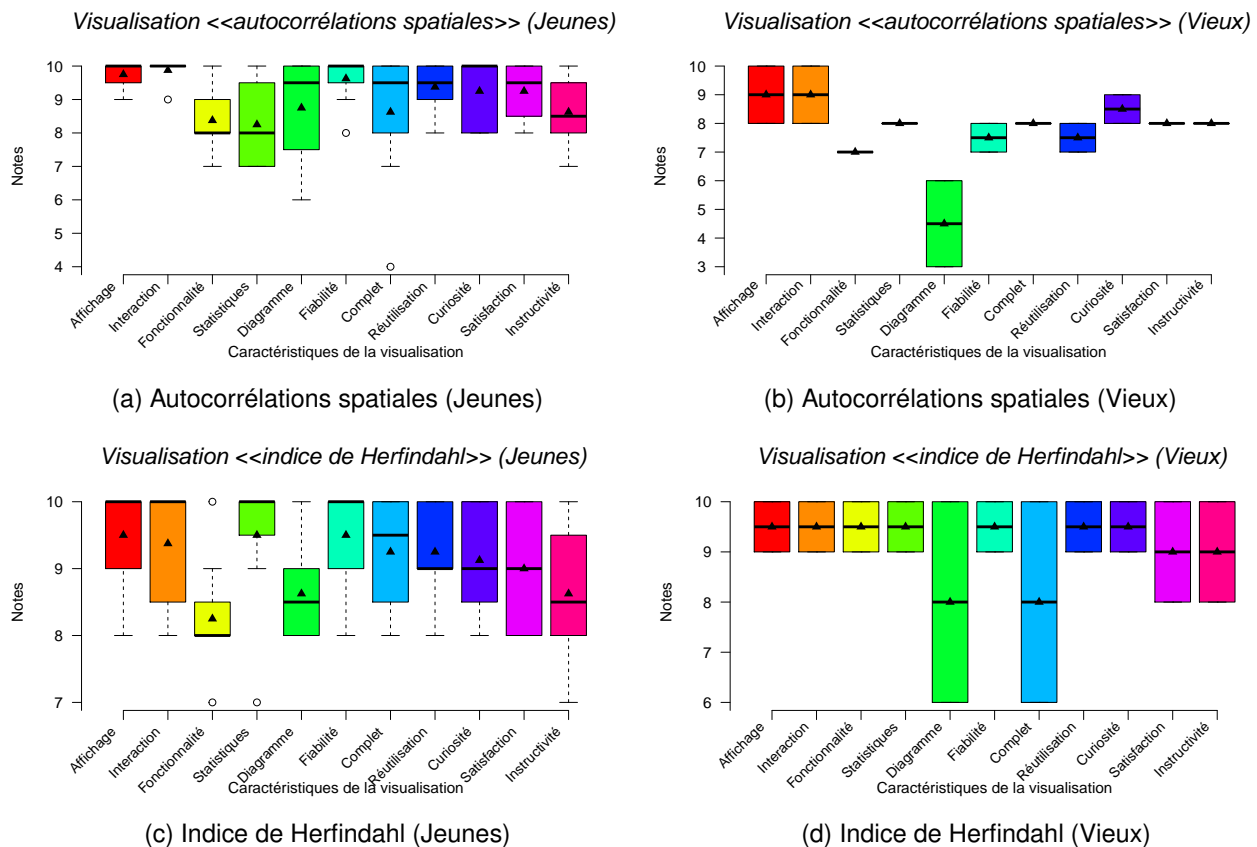


FIGURE 12 – Différences jeunes adultes-seniors dans les visualisations des autocorrélations spatiales et de l'indice de Herfindahl

Par ailleurs, il aurait été intéressant de voir s'il existe des facteurs principaux d'interaction entre le sexe et l'âge. Pour cela, nous aurions pu utiliser le test de l'analyse de la variance, connu sous le nom de test F ou ANOVA. Ce test peut être effectué avec un ou deux facteurs. Dans notre cas, c'est le test à deux facteurs qui nous aurait intéresser, car la question principale qui concerne les visualisations traite des éléments qui pourraient compliquer la compréhension de statistiques sophistiquées. Cependant, comme nous possédons un nombre d'individu faible pour la population senior (2 femmes et 2 hommes sur les 20 participants), nous pouvons considérer qu'il n'est pas nécessaire de poursuivre les calculs et les interprétations.

Au final, la combinaison de boîtes à moustache avec les tests de Student ont permis d'apprécier au mieux les aspects que les testeurs ont trouvés positifs ou négatifs. Les boîtes à moustache, complétées par la moyenne, ont l'avantage de montrer la plupart des statistiques descriptives et la dispersion des résultats. Les tests de Student confirment si les probabilités de chaque variable sont significativement différentes (H_1) et permettent de rejeter l'hypothèse H_0 . Dans notre cas, la compréhension des statistiques sophistiquées dans leur calcul représentait l'enjeu majeur des applications. Ce dernier a été vérifié par des questions indépendantes du questionnaire, nécessaires afin de s'assurer de l'état de compréhension de chaque candidat. Malgré le fait que l'hypothèse H_0 était rejetée d'une manière générale, les notes obtenues dans ce domaine étaient supérieures à 8 sur 10 en moyenne et pour la médiane. Sur les questionnaires cependant, la plupart des suggestions d'amélioration concernent une meilleure visibilité de certaines fonctionnalités comme le bouton d'information (7 personnes sur 20) ou l'illumination des points sur le graphique ternaire (7 personnes sur 20) ou encore plus d'informations sur les statistiques elles-mêmes et le diagramme ternaire ou l'affichage des communes lorsque les points du diagramme ter-

naire sont sélectionnés. Mais la grande majorité (15 personnes sur 20) a explicitement apprécié la rapidité d'exécution, l'interactivité, la simplicité et la praticité qu'offre l'application, et beaucoup (8 personnes sur 20) la compréhensibilité et l'intuitivité.

5 Conclusion

La ségrégation spatiale est un phénomène qui peut être caractérisé de plusieurs manières différentes. Les quelques indices utilisés en ont montré chacun une facette, que cela soit au niveau de l'égalité avec la loi de Zipf, l'indice de Herfindahl ou le coefficient de Gini ou de la concentration avec les dérivés de l'indice de Duncan ou encore la similarité de voisinage avec les autocorrélations spatiales. Ainsi, chacune de ces méthodes apporte sa réponse en complément des autres et c'est pourquoi il n'est pas possible d'affirmer qu'un seul de ces procédés suffirait pour comprendre toute la complexité de la répartition spatiale qu'induit la ségrégation. Dans l'idéal, il faudrait en utiliser au minimum trois, un de chaque catégorie (éviter l'indice de Duncan si possible), pour avoir la vue d'ensemble la plus générale. Le nouvel indice généré, $ISKB_a$, qui prend en compte à la fois un indice de ségrégation spatiale, l'existence d'une adjacence intercommunale (ou les longueurs des frontières intercommunales ou les infrastructures de passage intercommunal), le périmètre, la surface et les altitudes, trouve une application particulière et spécifique dans les régions montagneuses, les autres indices de ségrégation spatiale «classiques» ne s'utilisant essentiellement que dans les régions urbaines entre quartiers, voire entre arrondissements.

Par ailleurs, les autocorrélations spatiales, mesures idéales pour observer les ressemblances communales, se déclinent en indice global et local permettant d'avoir un résultat plus détaillé sur le phénomène social du Tyrol italien. Grâce aux diagrammes de Moran, la configuration spatiale statistique peut également être appréciée sans usage de la cartographie et complète les résultats des indices de Moran locaux. Il est d'ailleurs possible de combiner les valeurs des deux mesures dans un graphique.

En outre, la deuxième partie du mémoire sur l'accessibilité et la vulgarisation de ces mesures statistiques en vue d'une potentielle diffusion au sein de tout public confondu a révélé une méthode à la fois innovante et rapide en vue de remplir son but. L'agencement d'une carte dotée de boutons informatifs, pratiques et aux fonctionnalités rapides et d'un diagramme ternaire complémentaire et également interactif a permis d'instruire une vingtaine de personnes sur des statistiques sophistiquées dans leur calcul en l'espace de 15 à 25 minutes. Il demeure bien entendu des différences populationnelles au niveau de la cognition et l'appréhension des thématiques qu'offrent les applications, notamment entre la population âgée et jeune, mais également entre hommes et femmes. Toutefois, les différences n'étant pas significativement très grandes d'une manière générale, les visualisations ont rempli le rôle primaire qui leur était attribué.

6 Annexes

6.1 Lexique et données

a_{ij} = Matrice d'adjacence entre les communes i et j

A_{ij} = Différence d'altitude entre les communes i et j

e_{ii} = Composantes de la diagonale de la matrice d'échange

e_{ij} = Matrice d'échange pondérée avec $e_{i\bullet} = f_i$ (prenant en compte la pondération f_i soit la population, et la matrice d'adjacence a_{ij} entre les communes)

f_i = Pondération par la population de la commune i

F_{ia} = fréquence de la commune i pour la communauté a

G = Coefficient de Gini

H = Indice d'Herfindahl normal pour le Tyrol italien

H_i = Indice d'Herfindahl normal par commune

H^* = Indice d'Herfindahl normalisé pour le Tyrol italien

H_i^* = Indice d'Herfindahl normalisé par commune

I = I_\bullet = Indice de Moran global

$I_i = \delta_i$ = Indice de Moran local (pour chaque commune i)

ID_a = Indice de disimilarité de Duncan pour la communauté a

IS_a = Indice de ségrégation de Duncan pour la communauté a

ISG_a = Indice de Gorard pour la communauté a

ISM_a = Indice de Morill pour la communauté a

ISW_a = Indice de Wong pour la communauté a

i, j = Indices des communes

k = Constante de la loi de Zipf

l_{ij} = Longueur de la frontière entre deux communes

$L_{ij} = \frac{l_{ij}}{P_j}$ = Longueur de la frontière entre deux communes l_{ij} divisé par le périmètre de la commune P_j

m = Nombre de communautés linguistiques total (ici 4)

M = Nombre de participants à une expérience de visualisation

$N = N_{\bullet\bullet}$ = Population totale dans le Tyrol italien

$N_{\bullet a}$ = Population totale de la communauté a au sein du Tyrol italien.

$N_i = N_{i\bullet}$ = Population totale dans la commune i

n = Nombre de commune dans le Tyrol italien

$p_{ia} = \frac{N_{ia}}{N_{i\bullet}}$ = Proportion de la communauté linguistique a dans la commune i

$\bar{p}_a = \frac{N_{\bullet a}}{N_{\bullet\bullet}}$ = Proportion de la communauté linguistique a au niveau du Tyrol italien

R = Différence entre la proportion de germanophones d'une commune x et la moyenne des proportions de germanophones des communes voisines Wx

r_{ia} = Rang dans la loi de Zipf

$S = S_\bullet$ = Surface totale du Tyrol italien

S_i = Surface de la commune i

t = «Age» du réseau

Wx = Vecteur des valeurs moyennes des voisines d'une commune i située dans le vecteur x

x = Vecteur contenant le pourcentage d'une communauté a

Communes	Germanophones	Italophones	Ladinophones	Etrangers	Périmètres [m]	Surfaces [m2]	Altitude [m]
Ala	0	7480	41	1366	71780.71844	119856764.3	180
Albiano	6	1320	0	182	13195.45709	9964392.382	644
Aldeno	0	2850	6	180	15570.78059	8970562.496	209
Aldino	1595	28	3	43	41007.45197	62690821.24	1225
Amblar	0	175	48	6	22878.83591	14849694.64	980
Andalo	0	971	10	45	23711.48638	11378441.82	1042
Andriano	884	94	5	54	12706.26839	4891304.189	274
Anterivo	328	46	0	9	15100.76625	11064225	1209
Appiano sulla strada del vino	11234	1732	63	1335	49492.16215	59450994.2	416
Arco	8	15149	64	1650	55267.73401	63218432.23	91
Avelengo	687	17	0	49	30316.5888	27395595.17	1290
Avio	0	3742	7	366	45288.70688	68889351.78	131
Badia	58	138	3107	72	48112.17914	83071443.27	1315
Barbiano	1500	29	9	81	22541.92436	24511563.74	830
Baselga di Pinè	11	4563	0	354	41302.50235	41071542.82	964
Bedollo	0	1439	0	40	27496.48291	27459867.1	1059
Bersone	0	287	0	2	22709.95865	9780608.071	637
Besenello	10	2388	0	113	39915.86754	25937243.78	218
Bieno	0	392	0	38	17413.46429	11711127.64	815
Bleggio Superiore	0	1478	0	122	32971.13274	32674520.84	628
Bocenago	0	384	0	23	12133.73158	8452786.758	750
Bolbeno	0	340	0	22	19148.70648	12315351.24	575
Bolzano	23055	66673	614	14413	51404.47667	52291747.14	262
Bondo	0	680	0	18	27613.17557	10649467.19	823
Bondone	0	647	0	21	24732.36996	19185229.97	720
Borgo Valsugana	21	6087	6	712	37909.29965	52364031.19	380
Bosentino	0	800	0	29	19263.93478	4700362.173	688
Braies	649	4	1	15	52776.76127	90214253.73	1213
Breguzzo	0	554	0	18	42244.83575	34890088.27	798
Brennero	1465	338	9	302	63213.22243	114293502.9	1098
Brentonico	0	3574	5	303	43968.79308	57140555.53	692
Bresimo	0	201	52	1	30341.34353	41014498.86	1036
Bressanone	13720	4869	252	2016	57956.10316	84703342.35	559
Brez	0	419	247	40	26280.58877	19167396.97	792
Brione	0	131	0	9	18193.20595	9458014.792	893
Bronzolo	909	1510	16	214	14582.58976	7539647.779	238
Brunico	11650	2153	323	1504	46802.61835	44998062.23	838
Caderzone Terme	0	606	0	63	19911.06349	18607244.82	723
Cagnò	0	217	107	36	16888.87846	3407903.035	663
Caines	392	14	1	7	7291.811581	1634458.229	592
Calavino	0	1237	0	244	26593.35763	12875547.7	409
Calceranica al Lago	5	1254	0	46	9456.275855	3394568.778	465
Caldaro sulla strada del vino	6596	501	26	617	49972.02424	48044872.58	425
Caldes	0	883	124	83	29347.99494	20808931.73	697
Caldonazzo	17	3152	6	165	29228.42188	21413496.32	480
Calliano	0	1420	0	191	27218.20397	10198953.79	187
Campitello di Fassa	0	99	608	30	31121.83859	25023654.67	1448
Campo Tures	4916	116	18	243	72504.74727	163982348.2	864
Campo di Trens	2441	97	9	138	51547.50308	95386774.73	937
Campodenno	0	1152	266	80	30061.1356	25023328.05	534
Canal San Bovo	0	1538	5	49	77624.35001	125684137.1	757
Canazei	0	280	1524	103	49652.90438	67018669.03	1465
Capriana	0	558	0	47	17552.79755	12818983.07	1007
Carano	0	989	0	84	22704.69034	13570258.64	1086
Carisolo	0	893	0	91	34968.3195	25122757.5	808
Carzano	0	491	0	14	11772.17128	1820178.683	433

Castel Condino	0	234	0	4	24491.66997	11100553.28	811
Castelbello-Ciardes	2237	29	0	120	46179.19505	53791742.73	587
Castelfondo	0	418	149	65	31016.738	25871943.64	948
Castello Tesino	0	1278	0	37	83959.421	112835940.5	871
Castello-Molina di Fiemme	0	2105	30	132	54571.9088	54559669.1	1005
Castelnuovo	0	957	0	78	17939.24491	13144246.25	348
Castelrotto	4940	225	938	380	49862.80803	117898639.3	1060
Cavalese	0	3548	41	364	33724.74144	45376325.5	1000
Cavareno	0	762	144	131	22904.87268	9484368.93	973
Cavedago	0	476	33	21	19873.55494	10028185.95	864
Cavedine	0	2695	10	211	27715.75278	38227056.86	504
Cavizzana	0	236	13	10	7938.875524	3377093.242	710
Cembra	5	1591	5	247	20443.55343	16963032.09	667
Centa San Nicolò	6	596	0	15	18052.4777	11225275.84	830
Cermes	1277	65	3	96	14834.82202	6624037.48	292
Chienes	2517	60	25	114	43087.96671	33680567.55	784
Chiusa	4422	382	39	311	51741.52218	51289734	523
Cimego	0	382	0	27	22959.05015	10514078.15	557
Cimone	0	660	0	25	15598.70998	9807111.577	530
Cinte Tesino	0	365	0	7	45544.89194	30297088.9	851
Cis	0	250	53	6	10981.49162	5497637.971	732
Civezzano	30	3693	0	181	19195.77309	15666533.65	469
Cles	0	5013	931	825	52784.03586	39166206.61	658
Cloz	0	431	167	131	13642.72996	8211285.456	791
Comano Terme	0	2494	5	414	72071.93005	68114285.09	400
Commezzadura	0	877	12	104	24735.4944	22025256.95	850
Condino	0	1430	0	104	54461.58311	33741030.29	444
Coredo	0	948	497	180	49748.90872	32637925.76	831
Cornedo all'Isarco	2818	337	10	181	31716.51314	40611293.91	290
Cortaccia sulla strada del vino	2049	72	8	102	33999.05084	29268007.06	333
Cortina sulla strada del vino	412	187	1	54	6111.26282	1999057.664	212
Corvara in Badia	45	89	1171	28	31170.75542	38836321.97	1568
Croviana	0	605	18	70	12940.87379	4989316.967	721
Cunevo	0	446	83	42	14473.28222	5547797.358	572
Curon Venosta	2298	63	0	89	85707.64276	209650465.1	1520
Daiano	0	630	8	54	14916.89217	9510809.193	1160
Dambel	0	292	112	31	10469.72325	5145161.016	751
Daone	0	595	0	2	82863.15732	157456834.3	767
Darè	0	216	0	38	7476.351311	1157110.325	600
Denno	0	877	190	185	26903.09924	10644159.32	429
Dimaro	0	1099	32	119	28822.38243	27772687.4	766
Dobbiaco	2644	490	10	181	90823.79528	125303588.9	1256
Don	0	195	38	16	15157.39868	5114621.87	971
Dorsino	0	408	0	18	22264.24844	12507207.61	635
Drena	0	519	0	32	14260.91936	8336812.638	398
Dro	0	4093	13	460	33929.64991	27953829.02	123
Egna	2836	1669	19	504	33639.08012	23569653.26	214
Faedo	0	559	0	51	16440.16528	10680535.62	591
Fai della Paganella	0	866	0	32	27124.22569	12129501.74	957
Falzes	2510	62	36	74	29257.97504	33131947.85	1022
Faver	0	770	0	63	17545.61256	9426853.392	673
Fiavè	0	921	0	177	28943.16221	24282520.21	669
Fiera di Primiero	0	474	0	33	2116.659614	149482.5905	710
Fierozzo	442	39	0	0	20143.67215	17942802.52	1127
Fiè allo Sciliar	3045	143	20	267	39172.26374	43958453.56	880
Flavon	0	361	86	58	17048.55924	7576669.713	575
Folgaria	249	2706	0	175	44820.90279	71626810.97	1166

Fondo	0	923	333	179	32412.74947	30629248.99	987
Fornace	0	1115	0	206	11999.83823	7224195.149	740
Fortezza	451	292	14	224	41007.58425	61772902.84	749
Frassilongo	269	52	0	0	17870.89212	16680989.45	852
Funes	2430	49	8	97	52623.07234	81382202.03	1132
Gais	2993	82	9	104	40798.14192	60615400.69	841
Gargazzone	1230	318	15	108	9075.335464	4913702.432	267
Garniga Terme	0	348	0	33	17631.36745	13125325.83	810
Giovo	13	2372	6	74	33094.37297	20813233.91	496
Giustino	0	653	0	90	46676.01433	39387211.56	770
Glorenza	825	33	0	31	19756.02968	13219049.78	907
Grauno	0	132	0	10	12269.75282	7257920.538	976
Grigno	0	2117	9	145	33325.07647	46381534.28	263
Grumes	0	419	0	19	14178.88007	10725405.38	851
Imer	0	1136	6	41	33214.60803	27729798	670
Isera	7	2482	0	136	21031.02717	14092180.85	243
Ivano-Fracena	0	297	0	53	12108.34297	6078709.489	452
La Valle	20	10	1248	22	33942.30522	38918277.44	1353
Laces	4668	94	2	393	54371.88527	78705885.1	639
Lagundo	3842	658	11	372	36739.0909	23684042.8	350
Laion	2226	97	152	164	32613.79237	37525510.06	1093
Laives	4388	11208	80	1718	26068.6472	24108645.71	255
Lana	9544	821	27	957	43087.66541	36124315.96	310
Lardaro	0	196	0	9	20837.74846	10428464.84	732
Lasa	3719	63	9	158	55337.21256	110225026.3	868
Lasino	0	1254	0	48	20294.69587	16055128.79	463
Lauregno	334	11	0	0	23321.23917	13885018.22	1150
Lavarone	85	955	8	39	31159.57218	26312872.41	1170
Lavis	21	7666	20	929	21584.48668	12178279.48	232
Ledro	0	4980	20	448	75846.08474	156374627.6	660
Levico Terme	37	6538	14	927	39514.03893	62829075.33	506
Lisignago	0	448	0	36	14424.38945	7142917.373	582
Livo	0	568	234	85	19757.64122	15218018.48	741
Lona-Lases	0	668	0	208	26682.54863	11372793.08	639
Luserna	238	38	0	3	23533.26572	8197245.966	1333
Luson	1438	20	12	59	46174.37007	74413708.19	972
Magrè sulla strada del vino	961	173	5	158	23990.75889	13864310.73	241
Malles Venosta	4725	146	4	252	91510.00452	247428495.2	1051
Malosco	0	211	144	90	23242.46882	6728452.162	1041
Malè	0	1797	61	277	36894.59595	26528057.92	738
Marebbe	81	141	2578	144	90525.1139	160230189.3	1285
Marlengo	2060	320	4	189	18546.32279	12863285.75	363
Martello	851	0	0	28	59201.00941	142801238	1312
Massimeno	0	115	0	9	32533.92369	21032517.06	861
Mazzin	0	74	381	39	28156.73008	23628470.51	1395
Meltina	1548	58	5	42	39455.82801	36954383.3	1142
Merano	16513	16051	154	6143	38755.73907	26340223.77	325
Mezzana	0	811	13	60	29213.38595	27350149.84	940
Mezzano	0	1582	0	33	47991.08929	48846181.33	640
Mezzocorona	0	4619	20	522	29851.52937	25353409.26	219
Mezzolombardo	0	5806	66	945	19772.3667	13879185.86	227
Moena	0	408	2126	156	62206.03849	82595578.07	1184
Molveno	0	1027	29	54	37211.44682	34116110.91	865
Monclassico	0	722	10	150	16444.30915	8760623.116	770
Monguelfo-Tesido	2535	122	9	181	51386.11681	46437006.92	1087
Montagna	1459	109	8	73	34510.30521	19510498.02	497
Montagne	0	243	0	3	20826.78399	12207391.16	1004

Mori	0	8552	23	881	54978.50222	40084316.13	204
Moso in Passiria	2132	7	2	22	78951.54201	193531147.5	1007
Nago-Torbole	7	2355	15	351	31060.04833	28384663.07	222
Nalles	1581	153	11	97	21932.39711	12240041.59	321
Nanno	0	438	131	39	11148.77487	4304286.348	551
Naturno	5043	158	7	379	51471.83925	67113657.46	528
Nave San Rocco	0	1243	0	148	9669.627081	4891610.506	207
Naz-Sciaves	2481	146	26	252	27692.19306	15960130.44	772
Nogaredo	0	1842	0	75	10079.69972	3614746.568	216
Nomi	0	1339	0	64	16905.17325	6488781.583	179
Nova Levante	1630	96	8	177	34444.23975	51096368.62	1182
Nova Ponente	3577	86	9	268	53779.74306	112494870.3	1357
Novaledo	6	961	0	51	17450.93429	7965450.578	475
Ora	2290	972	22	285	22599.98482	11788204.22	242
Ortisei	411	288	3718	280	23323.41796	24156789.26	1234
Ospedaletto	0	734	0	86	18968.12496	16747605.77	360
Ossana	0	782	0	62	30587.67614	25248914.27	1003
Padergnone	0	666	0	66	10176.74612	3604757.515	286
Palù del Fersina	157	10	0	2	20118.59673	16648338.46	1360
Panchià	0	717	17	37	34939.72222	20211516.16	981
Parcines	3280	130	9	145	39510.2162	55398157.87	626
Peio	0	1799	30	62	80794.17134	162304709	1173
Pellizzano	0	742	12	57	61710.29509	48361306.8	925
Pelugo	0	314	0	65	32471.4149	22978567.01	652
Perca	1362	62	12	38	32483.52595	30364415.13	972
Pergine Valsugana	273	18194	36	1943	54153.94416	54329233.54	482
Pieve Tesino	0	669	0	12	74259.08762	69234498.03	843
Pieve di Bono	0	1253	0	78	22758.01808	20829248.55	514
Pinzolo	0	2891	16	210	50340.30339	69321992.57	770
Plaus	617	15	0	68	13273.06884	4866098.507	519
Pomarolo	0	2199	0	156	16543.90605	9228677.907	206
Ponte Gardena	143	24	9	26	9227.899561	2328223.737	470
Postal	1272	378	10	104	11453.1426	6692195.306	270
Pozza di Fassa	0	249	1765	124	56755.46313	72973226.46	1325
Praso	0	330	0	4	18055.58878	9847921.598	785
Prato allo Stelvio	3073	86	2	230	44140.54071	51002525.01	915
Predazzo	0	4185	75	271	60821.34532	109966390.3	1018
Predoi	585	16	0	3	45415.02957	86357905.53	1475
Preore	0	378	0	12	13648.874	4373302.414	530
Prezzo	0	199	0	11	14079.8798	3851985.523	664
Proves	258	6	0	3	24031.67628	18368764.06	1420
Rabbi	0	1288	70	42	61886.9048	132789267.7	1095
Racines	4164	91	4	150	82013.92784	203286680.1	976
Ragoli	0	727	0	33	52071.44104	64907271.34	556
Rasun Anterselva	2724	32	12	116	75631.88023	121574463.1	1030
Renon	6885	329	18	451	56809.83663	111361721.1	1154
Revò	0	817	326	122	24487.29236	13358008.18	724
Rifiano	1286	24	5	31	43167.11612	35943155.52	504
Rio di Pusteria	2508	103	19	354	49051.49391	83824454.39	777
Riva del Garda	12	13870	52	1904	44852.03974	40725383.67	73
Rodengo	1156	3	1	28	31643.72518	29618637.94	885
Romallo	0	258	300	46	7687.68583	2444932.341	733
Romeno	0	848	383	143	13846.52618	9128013.222	961
Roncegno Terme	16	2649	0	149	29228.98875	38083847.25	535
Ronchi Valsugana	0	414	0	7	16180.05659	10003192.8	776
Roncane	0	1403	0	40	39597.01579	29790427.01	842
Ronzo-Chienis	0	940	0	61	17940.94494	13208630.8	974

Ronzone	0	250	94	47	17188.28131	5297401.467	1085
Rovereto	49	32665	75	4965	49163.25707	50984963.04	204
Roverè della Luna	0	1370	5	198	25139.75366	10414525.33	251
Ruffrè-Mendola	0	289	117	10	14188.21697	6580122.098	1200
Rumo	0	649	92	81	28352.77251	30854456.51	944
Sagron Mis	0	181	0	2	19243.52824	11061841.18	1062
Salorno	1080	1768	11	732	38330.11346	33125517.05	224
Samone	0	511	0	33	9640.495984	4902115.622	673
San Candido	2603	448	9	161	55268.56485	79847897.45	1175
San Genesio Atesino	2805	89	4	75	45593.28015	68837747.13	1087
San Leonardo in Passiria	3427	36	4	74	60264.3614	89028515.58	689
San Lorenzo di Sebato	3437	95	74	182	43115.58562	51460668.69	810
San Lorenzo in Banale	0	1114	0	67	41595.13199	61406186.99	758
San Martino in Badia	31	25	1637	30	45892.95268	75943934.49	1135
San Martino in Passiria	3024	26	1	109	40169.39822	29991690.98	597
San Michele all'Adige	0	2449	24	438	15041.93678	5315922.312	228
San Pancrazio	1538	15	3	23	42884.15306	63173361.02	735
Sant'Orsola Terme	253	747	0	73	18686.79658	15364136.73	925
Santa Cristina Valgardena	75	79	1629	123	39633.94749	31923591.63	1428
Sanzeno	0	633	231	64	14040.49343	7874995.039	640
Sarentino	6624	123	7	165	89312.67084	302274552.3	961
Sarnonico	0	539	117	93	46246.74735	12189999.53	963
Scena	2668	45	4	129	36499.92807	48131006.43	600
Scurelle	0	1306	0	95	39420.73652	29996891.33	375
Segonzano	6	1339	0	186	21552.04028	20712414.84	660
Selva dei Molini	1434	13	3	17	56389.97385	104788555.5	1229
Selva di Val Gardena	133	132	2316	80	37809.6265	56240660.76	1563
Senale-San Felice	768	8	0	4	25152.02445	27631715.29	1279
Senales	1244	22	0	47	80999.73664	209839910.3	1327
Sesto	1762	81	5	94	46523.39985	80430151.22	1310
Sfruz	0	226	78	19	27321.04796	11808698.29	1015
Silandro	5304	291	8	422	59362.44044	115173196.2	721
Siror	0	1196	0	88	57937.16917	74920505.27	765
Sluderno	1731	21	0	65	22675.37575	20717996.8	921
Smarano	0	279	156	74	21090.81892	6283976.481	985
Soraga	0	52	629	55	32039.94835	19744944.05	1220
Sover	0	832	0	50	18613.97251	14822240.69	831
Spera	0	565	0	20	15226.93279	3252249.635	553
Spiazzo	0	1158	0	157	62057.00132	71061005.28	645
Spormaggiore	0	878	227	154	38656.67223	30200481.89	565
Sporminore	0	539	121	54	22806.54032	17471136.22	515
Stelvio	1097	17	0	85	58419.44569	141623576.6	1310
Stenico	0	1031	14	92	52135.38004	49146620	666
Storo	0	4346	0	309	45124.00657	62932749.48	409
Strembo	0	438	0	93	37266.68253	38331762.64	714
Strigno	0	1331	8	120	32504.87247	12175584.47	506
Taio	0	1989	594	375	27280.02007	11313708.16	515
Tassullo	0	1354	404	162	33935.91511	13540744.88	546
Telve	0	1877	0	118	55293.90781	64748287.43	550
Telve di Sopra	0	592	0	25	30564.78417	17716438.96	650
Tenna	0	919	0	36	8648.0404	3114753.633	569
Tenno	0	1800	5	162	24637.63919	28301483.12	428
Terento	1679	9	1	39	32579.52546	42161958.23	1210
Terlago	5	1812	0	66	31497.04365	37046882.21	456
Terlano	3226	620	12	344	33207.71785	18568753.18	248
Termeno sulla strada del vino	2983	106	6	217	35573.42204	19439209.27	276
Terragnolo	6	730	0	19	33493.82526	39566813.92	785

Terres	0	214	67	26	14076.44712	6344981.983	593
Terzolas	0	547	22	37	15881.44702	5588835.539	755
Tesero	0	2689	18	161	50093.7909	50545924.12	1000
Tesimo	1771	36	6	73	33184.03603	38125555.98	635
Tione di Trento	0	3099	18	491	57312.52406	33446880.71	565
Tires	902	11	6	55	36467.11682	42181616.09	1028
Tirolo	2277	68	5	112	33239.24711	25618525.23	594
Ton	0	851	373	95	24463.91367	26284875.28	482
Tonadico	0	1364	13	102	88781.35748	89399846.18	750
Torcegno	0	684	0	13	27084.32816	15192876.87	769
Trambileno	0	1299	0	56	51373.89218	50699594.96	525
Transacqua	0	2019	15	102	37568.55143	35581985.75	746
Trento	287	100016	368	13527	75408.0304	157877117.3	194
Tres	0	498	169	51	26438.08838	14605821.28	810
Trodena nel parco naturale	689	237	6	89	22513.20084	20561075.37	1127
Tubre	917	20	0	31	33237.55986	46267182.34	1240
Tuenno	0	1685	474	215	57835.62276	71288009.78	629
Ultimo	2865	15	2	36	65867.60604	208120413.7	1190
Vadena	350	566	4	102	32828.93631	13736017.15	243
Val di Vizze	2372	233	2	208	69759.63303	142122704.2	948
Valda	0	193	0	34	10817.68807	6137465.969	795
Valdaora	2837	94	10	172	38187.52814	49076691.57	1048
Valfloriana	0	502	0	28	35333.48793	39328936.16	853
Vallarsa	0	1307	0	36	52004.96734	77864718.32	724
Valle Aurina	5717	54	18	130	68501.00696	187890871.5	1054
Valle di Casies	2172	36	2	58	56543.10302	110135475.5	1206
Vandoies	3077	39	16	134	67141.69143	110819427.9	755
Varena	0	791	10	44	25565.41322	23068923.69	1180
Varna	3468	442	40	342	44579.67513	70339115.42	671
Vattaro	8	1109	0	50	14049.00957	8283678.55	689
Velturmo	2622	24	20	94	23669.97291	24584787.05	851
Verano	895	19	0	21	28400.28904	22009294.72	1204
Vermiglio	0	1749	27	93	53625.16453	95635909.48	1261
Vervò	0	426	222	64	20440.72987	15208244.02	886
Vezzano	0	2062	0	121	34753.34472	31804187.94	385
Vignola-Falesina	0	161	0	1	18514.00142	11951562.35	984
Vigo Rendena	0	471	0	30	14885.6147	4517756.722	612
Vigo di Fassa	0	35	1059	113	29287.80228	26850461.61	1382
Vigolo Vattaro	5	2081	0	103	21381.70283	20818411.25	725
Villa Agnedo	0	873	0	114	19672.92703	13314222.55	356
Villa Lagarina	0	3518	0	174	26890.08891	24129093.26	180
Villa Rendena	0	829	0	173	33685.09877	35039027.73	608
Villabassa	1256	106	2	118	28434.51543	18032745.95	1158
Villandro	1804	21	7	60	38996.38872	43950691.69	880
Vipiteno	4392	1548	24	566	46427.88545	32965274.16	948
Volano	11	2768	0	346	16278.04084	10740254.23	189
Zambana	0	1451	0	169	32320.53538	11686040.26	206
Ziano di Fiemme	0	1589	22	68	34542.31579	35747192.77	953
Zuclo	0	339	0	6	20975.19155	10304975.63	595

Variable/sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Sexe	F	H	F	F	F	H	H	F	H	H	H	H	F	H	H	F	F	F	F	H	F
Age	27	31	58	27	24	65	25	21	21	27	23	27	23	25	25	25	59	21	63	25	
Visu	F	F	F	F	F	F	F	F	F	F	P	P	P	P	P	P	P	P	P	P	
Conn. Tyrol italien	Oui	Oui	Oui	Non	Non	Oui	Oui	Oui	Non	Oui	Oui	Non	Oui	Oui	Non	Non	Non	Oui	Oui	Oui	
Local. Tyrol italien	Oui	Oui	Oui	Oui	Non	Oui	Oui	Oui	Non	Oui	Oui	Non	Oui	Oui	Non	Non	Non	Non	Oui	Oui	
Conn. comp. ling.	Non	Non	Oui	Non	Non	Oui	Non	Non	Non	Non	Oui	Non	Oui	Oui	Non	Non	Non	Non	Oui	Oui	
Conn. A. ou H.	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Oui	Non	Non	Non	Non	Non	Non	
Affichage	10	10	10	10	9	8	10	9	10	10	10	10	10	10	10	8	9	8	10	10	
Interaction	10	10	10	10	9	8	10	10	10	10	9	10	10	10	10	8	9	8	10	10	
Fonctionnalité	8	9	7	10	8	7	7	8	8	9	8	8	9	8	8	8	9	7	10	10	
Statistiques	7	10	8	9	7	8	8	7	8	10	9	10	10	10	10	10	9	7	10	10	
Diagramme	10	10	6	9	7	3	10	8	6	10	8	8	9	10	9	9	6	8	10	8	
Fiabilité	10	10	8	10	8	7	10	10	9	10	10	9	8	10	10	10	9	9	10	10	
Complet	7	10	8	10	9	8	4	10	9	10	9	8	10	10	10	9	6	8	10	10	
Réutilisation	9	10	8	10	9	7	10	8	9	10	10	9	10	9	9	9	9	8	10	10	
Curiosité	8	8	8	10	10	9	8	10	10	10	10	8	9	10	8	9	9	9	10	10	
Satisfaisante	8	9	8	10	9	8	8	10	10	10	8	8	10	10	9	9	8	8	10	10	
Instructif	8	9	8	10	7	8	8	8	9	10	8	10	7	8	9	9	8	8	10	10	

Résumé des réponses au questionnaire sur les visualisation. Pour l'attribution de la visualisation, F (face) correspond à l'autocorrélation spatiale et P (pile) à l'indice de Herfindahl

6.2 Liste des codes

«x.csv» correspond à un fichier dans lequel sont répertoriés les nombres d'habitants germanophones par commune

«ADJ.csv» correspond à la matrice d'adjacence entre communes

«Perimetres.csv» correspond à un fichier dans lequel sont répertoriés les périmètres pour chaque commune

«Surfaces.csv» correspond à un fichier dans lequel sont répertoriés les surfaces pour chaque commune

«Altitudes.csv» correspond à un fichier dans lequel sont répertoriés les altitudes pour chaque commune

Pour tous les codes sauf la loi de Zipf et les autocorrélations spatiales, source : (Tivadar, 2014) avec quelques modifications

«donnees_ger.csv» correspond à un fichier dans lequel sont répertoriés les proportions de germanophones pour chaque commune

Loi de Zipf

```
setwd("C:/Users/Kevin/Desktop/Tyrol")
donnees = read.csv2("Zipf_Nb_GER.csv", header = TRUE, sep = ",")
donnees = as.matrix(donnees)
donnees = as.numeric(donnees)
frequence <- donnees/sum(donnees)
lzipf <- function(s,N) -s*log(1:N)-log(sum(1/(1:N)^s))

opt.f <- function(s) sum((log(frequence)-lzipf(s,length(frequence)))^2)

opt <- optimize(opt.f,c(0.5,150)) #c(0.5,150) correspond l'intervalle du rang
opt
#Opt est le meilleur ajustement quadratique

library(stats4)

ll <- function(s) sum(donnees*(s*log(1:150)+log(sum(1/(1:150)^s))))

fit <- mle(ll, start=list(s=1))
summary(fit)

s.sq <- opt$minimum
s.ll <- coef(fit)

plot(1:150,frequence, log="xy")
lines(1:150,exp(lzipf(s.sq,150)), col=2)
lines(1:150,exp(lzipf(s.ll,150)), col=3)
# La ligne rouge correspond la somme de l'ajustement des
# carrs, la ligne verte est l'ajustement de la vraisemblance
# maximale
```

Indice d'Herfindahl

```
setwd("C:/Users/Kevin/Desktop/Tyrol")
Donnees = read.csv2("x.csv", header = TRUE, row.names = 1) #habitants par communes

#Herfindahl normal et normalisé pour chaque commune du Tyrol italien
x<-as.matrix(Donnees)
t<-rowSums(x)
p<-x/t
H<-rowSums(p^2)
H_Norm<-(H-1/ncol(x))/(1-1/ncol(x))

#Herfindahl normal et normalisé pour l'ensemble du Tyrol italien
```



```

Total<-sum(x)
pTotal<-colSums(x)/Total
Tot2<-pTotal^2
H_Tyr<-sum(Tot2)
H_T_Norm<-(H_Tyr-1/ncol(x))/(1-1/ncol(x))

H          #Herfindahl normal pour les communes du Tyrol italien
H_Norm     #Herfindahl normalisé pour les communes du Tyrol italien
H_Tyr      #Herfindahl normal pour l'ensemble du Tyrol italien
H_T_Norm   #Herfindahl normalisé pour l'ensemble du Tyrol italien

HER <- data.frame(H, H_Norm)
write.table(HER, file = 'Herfindahl.csv', sep="," , row.names=FALSE, col.names=FALSE)

```

Coefficient de Gini

```

setwd("C:/Users/Kevin/Desktop/Tyrol")
Donnees = read.csv2("x.csv",header = TRUE, row.names=1) #habitants par communes

```

```

Gini<-function(x) #calcul de Gini
{
  x<-as.matrix(Donnees)
  result<-vector(length=ncol(x))
  pij<-array(0,dim=c(ncol(x),nrow(x),nrow(x)))
  t<-rowSums(x)
  p<-x/t
  Total<-sum(x)
  pTotal<- colSums(x)/Total
  for (k in 1:ncol(x))
  {
    for (i in 1:nrow(x)) pij[k,,i]<-p[i,k]
    for (i in 1:nrow(x)) pij[k,i,]<-abs(pij[k,i,]-p[i,k])
    matprovi<-(t%*%t)*pij[k,,]
    result[k]<-sum(matprovi)/(2*Total*Total*pTotal[k]*(1-pTotal[k]))
    #print(pij) permet de vérifier que tous les matrices sont prises en compte
  }
  return(result)
}

```

Gini(x)

Indice de ségrégation de Duncan

```

setwd("C:/Users/Kevin/Desktop/Tyrol")
Donnees = read.csv2("x.csv",header = TRUE, row.names = 1) #habitants par communes

```

```

ISa<-function(x)
{
  x<-as.matrix(Donnees)
  result<-vector(length=ncol(x))
  Total<-sum(x) #Nombre d'habitant
  pTotal<-colSums(x)/Total #moyenne du taux de la communauté
  tx<-rowSums(x) #agir sur chaque ligne des habitants
  px<-x/tx #pourcentage de la communauté
  for (i in 1:ncol(x))
  {
    px[,i]<-tx*abs(px[,i]-pTotal[i])
    result[i]<-sum(px[,i])/(2*Total*pTotal[i]*(1-pTotal[i]))
  }
  return(result)
}

```

ISa(x)

Indice de dissimilarité de Duncan

```
setwd("C:/Users/Kevin/Desktop/Tyrol")
Donnees = read.csv2("x.csv", header = TRUE, row.names = 1) #habitants par communes
```

```
ID<-function(x)
{
  x<-as.matrix(Donnees)
  result<-matrix(data=0, nrow=ncol(x), ncol=ncol(x))
  pij<-array(0, dim=c(ncol(x), nrow(x), nrow(x)))
  varTotal <- colSums(x)
  for (k1 in 1:ncol(x))
    for (k2 in 1:ncol(x))
      result[k1, k2]<-0.5*sum(abs(x[, k1]/varTotal[k1]-x[, k2]/varTotal[k2]))
  return(result)
}
```

```
ID(x)
```

```
rowSums(ID(x)/3)
```

Indice de Gorard

```
setwd("C:/Users/Kevin/Desktop/Tyrol")
Donnees = read.csv2("x.csv", header = TRUE, row.names = 1) #habitants par communes
```

```
ISG<-function(x)
{
  x<-as.matrix(Donnees)
  result<-vector(length=ncol(x))
  tx<-rowSums(x)
  varTotal<-colSums(x)
  Total<-sum(x)
  for (k in 1:ncol(x)) result[k]<-0.5*sum(abs(x[, k]/varTotal[k]-tx/Total))
  return(result)
}
```

```
ISG(x)
```

Indice de Morill avec indice de ségrégation de Duncan

```
setwd("C:/Users/Kevin/Desktop/Tyrol")
Donnees = read.csv2("x.csv", header = TRUE, row.names = 1) #habitants par communes
ADJ = read.csv2("ADJ.csv", header = FALSE)
```

```
ISM<-function(x, c)
{
  x<-as.matrix(Donnees)
  c<-as.matrix(ADJ)
  IS<-vector(length=ncol(x))
  Total<-sum(x) #nombre d'habitants du Tyrol italien
  pTotal<-colSums(x)/Total #proportion de chaque communauté au Tyrol italien
  tx<-rowSums(x) #nombre d'habitants total par commune
  px<-x/tx #proportion d'habitants de chaque communauté sur l'ensemble de la commune
  for (i in 1:ncol(x))
  {
    px[, i]<-tx*abs(px[, i]-pTotal[i])
    IS[i]<-sum(px[, i])/(2*Total*pTotal[i]*(1-pTotal[i]))
  }
  result<-vector(length=ncol(x))
  pij<-array(0, dim=c(ncol(x), nrow(x), nrow(x)))
  t<-rowSums(x)
```

```

p<-x/t
for (k in 1:ncol(x))
{
  for (i in 1:nrow(x)) pij[k,,i]<-p[i,k]
  for (i in 1:nrow(x)) pij[k,i,]<-abs(pij[k,i,]-p[i,k])
  matprovi<-c%%pij[k,,]
  matprovi<-matprovi*diag(nrow(x))
  result[k]<-IS[k]-sum(matprovi)/sum(c)
}
return(result)
}

```

ISM(x)

Indice de Morill avec Gorard

```

setwd("C:/Users/Kevin/Desktop/Tyrol")
Donnees = read.csv2("x.csv",header = TRUE, row.names = 1) #habitants par communes
ADJ = read.csv2("ADJ.csv", header = FALSE)

```

```

ISM<-function(x,c)
{
  x<-as.matrix(Donnees)
  c<-as.matrix(ADJ)
  IS<-vector(length=ncol(x))
  tx<-rowSums(x)
  varTotal<-colSums(x)
  Total<-sum(x)
  for (k in 1:ncol(x)) IS[k]<-0.5*sum(abs(x[,k]/varTotal[k]-tx/Total))
  result<-vector(length=ncol(x))
  pij<-array(0,dim=c(ncol(x),nrow(x),nrow(x)))
  t<-rowSums(x)
  p<-x/t
  for (k in 1:ncol(x))
  {
    for (i in 1:nrow(x)) pij[k,,i]<-p[i,k]
    for (i in 1:nrow(x)) pij[k,i,]<-abs(pij[k,i,]-p[i,k])
    c<-as.matrix(c)
    matprovi<-c%%pij[k,,]
    matprovi<-matprovi*diag(nrow(x))
    result[k]<-IS[k]-sum(matprovi)/sum(c)
  }
  return(result)
}

```

ISM(x)

Indice de Wong sur la base de l'indice de ségrégation de Duncan

```

setwd("C:/Users/Kevin/Desktop/Tyrol")
Donnees = read.csv2("x.csv",header = TRUE, row.names = 1) #habitants par communes
ADJ = read.csv2("LO.csv", header = TRUE, dec=".", stringsAsFactors = FALSE)
PER = read.csv2("Perimetres.csv", header = TRUE, dec=".", sep=";", stringsAsFactors = FALSE)
SUR = read.csv2("Surfaces.csv", header = TRUE, dec=".", sep=";", stringsAsFactors = FALSE)

```

```

ISW<-function(x,f,perimeter,area)
{
  x<-as.matrix(Donnees)
  f<-as.matrix(ADJ)
  perimeter<-as.matrix(PER)
  perimeter<-as.vector(perimeter)
  area<-as.matrix(SUR)
  area<-as.vector(area)
  IS<-vector(length=ncol(x))

```

```

Total<-sum(x)
pTotal<-colSums(x) / Total
tx<-rowSums(x)
px<-x / tx
for (i in 1:ncol(x))
{
  px[,i]<-tx * abs(px[,i]-pTotal[i])
  IS[i]<-sum(px[,i]) / (2 * Total * pTotal[i] * (1-pTotal[i]))
}
result<-vector(length=ncol(x))
PerAij<-array(0,dim=c(ncol(x),nrow(x),nrow(x)))
pij<-array(0,dim=c(ncol(x),nrow(x),nrow(x)))
t<-rowSums(x)
p<-x / t
maxPerA<-max(perimeter / area)
w<-f / perimeter
for (k in 1:ncol(x))
{
  for (i in 1:nrow(x)) pij[k,,i]<-p[i,k]
  for (i in 1:nrow(x)) pij[k,i,]<-abs(pij[k,i,]-p[i,k])
  for (i in 1:nrow(x)) PerAij[k,,i]<-perimeter[i] / area[i]
  for (i in 1:nrow(x)) PerAij[k,i,]<-PerAij[k,i,]+perimeter[i] / area[i]
  matprovi<-(w/sum(w))%*(pij[k,]*PerAij[k,])
  matprovi<-matprovi * diag(nrow(x))
  result[k]<-IS[k]-0.25*sum(matprovi) / maxPerA
}
return(result)
}

```

ISW(x)

Indice de Wong sur la base de l'indice de Gorard

```

setwd("C:/Users/Kevin/Desktop/Tyrol")
rm(list = ls())
Donnees = read.csv2("x.csv", header = TRUE, row.names = 1) #habitants par communes
ADJ = read.csv2("LO.csv", header = TRUE, dec=".", stringsAsFactors = FALSE)
PER = read.csv2("Perimetres.csv", header = TRUE, dec=".", sep=";", stringsAsFactors = FALSE)
SUR = read.csv2("Surfaces.csv", header = TRUE, dec=".", sep=";", stringsAsFactors = FALSE)

```

```

ISW<-function(x, f, perimeter, area)
{
  x<-as.matrix(Donnees)
  f<-as.matrix(ADJ)
  perimeter<-as.matrix(PER)
  perimeter<-as.vector(perimeter)
  area<-as.matrix(SUR)
  area<-as.vector(area)
  IS<-vector(length=ncol(x))
  tx<-rowSums(x)
  varTotal<-colSums(x)
  Total<-sum(x)
  for (k in 1:ncol(x)) IS[k]<-0.5*sum(abs(x[,k] / varTotal[k]-tx / Total))
  result<-vector(length=ncol(x))
  PerAij<-array(0,dim=c(ncol(x),nrow(x),nrow(x)))
  pij<-array(0,dim=c(ncol(x),nrow(x),nrow(x)))
  t<-rowSums(x)
  p<-x / t
  maxPerA<-max(perimeter / area)
  w<-f / perimeter
  for (k in 1:ncol(x))
  {
    for (i in 1:nrow(x)) pij[k,,i]<-p[i,k]
    for (i in 1:nrow(x)) pij[k,i,]<-abs(pij[k,i,]-p[i,k])
    for (i in 1:nrow(x)) PerAij[k,,i]<-perimeter[i] / area[i]
  }
}

```

```

    for (i in 1:nrow(x)) PerAij[k,i,]←PerAij[k,i,]+perimeter[i]/area[i]
    matprovi←(w/sum(w))%%(pij[k,,]*PerAij[k,,])
    matprovi←matprovi*diag(nrow(x))
    result[k]←IS[k]-0.25*sum(matprovi)/maxPerA
  }
  return(result)
}
ISW(x)

```

Nouvel indice basé sur Wong, les longueurs intercommunales, l'indice de Gorard et les altitudes comprises dans le rapport périmètre/surface

```

setwd("C:/Users/Kevin/Desktop/Tyrol")
Donnees = read.csv2("x.csv",header = TRUE, row.names = 1) #habitants par communes
ADJ = read.csv2("LO.csv", header = TRUE, dec=".", stringsAsFactors = FALSE)
PER = read.csv2("Perimetres.csv", header = TRUE, dec=".", sep=";", stringsAsFactors = FALSE)
SUR = read.csv2("Surfaces.csv", header = TRUE, dec=".", sep=";", stringsAsFactors = FALSE)
ALT = read.csv2("Altitudes.csv", header = TRUE)

```

```

ISW←function(x,f,perimeter,area, altitude)
{
  x←as.matrix(Donnees)
  f←as.matrix(ADJ)
  perimeter←as.matrix(PER)
  perimeter←as.vector(perimeter)
  area←as.matrix(SUR)
  area←as.vector(area)
  altitude←as.matrix(ALT)
  altitude←as.vector(altitude)
  IS←vector(length=ncol(x))
  Total←sum(x)
  pTotal←colSums(x)/Total
  tx←rowSums(x)
  px←x/tx
  for (i in 1:ncol(x))
  {
    px[,i]←tx*abs(px[,i]-pTotal[i])
    IS[i]←sum(px[,i])/(2*Total*pTotal[i]*(1-pTotal[i]))
  }
  result←vector(length=ncol(x))
  PerAij←array(0,dim=c(ncol(x),nrow(x),nrow(x)))
  pij←array(0,dim=c(ncol(x),nrow(x),nrow(x)))
  t←rowSums(x)
  p←x/t
  maxPerA←max(perimeter*altitude/area)
  w←f/perimeter
  for (k in 1:ncol(x))
  {
    for (i in 1:nrow(x)) pij[k,,i]←p[i,k]
    for (i in 1:nrow(x)) pij[k,i,]←abs(pij[k,i,]-p[i,k])
    for (i in 1:nrow(x)) PerAij[k,,i]←perimeter[i]*altitude[i]/area[i]
    for (i in 1:nrow(x)) PerAij[k,i,]←PerAij[k,i,]+perimeter[i]*altitude[i]/area[i]
    matprovi←(w/sum(w))%%(pij[k,,]*PerAij[k,,])
    matprovi←matprovi*diag(nrow(x))
    result[k]←IS[k]-0.25*sum(matprovi)/maxPerA
  }
  return(result)
}
ISW(x)

```

Nouvel indice basé sur Wong, les adjacences, l'indice de Gorard et les altitudes non comprises dans le rapport périmètre/surface

```

setwd("C:/Users/Kevin/Desktop/Tyrol")
Donnees = read.csv2("x.csv", header = TRUE, row.names = 1) #habitants par communes
ADJ = read.csv2("ADJ.csv", header = FALSE)
PER = read.csv2("Perimetres.csv", header = TRUE, dec=".", sep=";", stringsAsFactors = FALSE)
SUR = read.csv2("Surfaces.csv", header = TRUE, dec=".", sep=";", stringsAsFactors = FALSE)
ALT = read.csv2("Altitudes.csv", header = TRUE)

ISW<-function(x,c,perimeter,area, altitude)
{
  x<-as.matrix(Donnees)
  c<-as.matrix(ADJ)
  perimeter<-as.matrix(PER)
  perimeter<-as.vector(perimeter)
  area<-as.matrix(SUR)
  area<-as.vector(area)
  altitude<-as.matrix(ALT)
  altitude<-as.vector(altitude)
  IS<-vector(length=ncol(x))
  Total<-sum(x)
  pTotal<-colSums(x)/Total
  tx<-rowSums(x)
  px<-x/tx
  for (i in 1:ncol(x))
  {
    px[,i]<-tx*abs(px[,i]-pTotal[i])
    IS[i]<-sum(px[,i])/(2*Total*pTotal[i]*(1-pTotal[i]))
  }
  result<-vector(length=ncol(x))
  PerAij<-array(0,dim=c(ncol(x),nrow(x),nrow(x)))
  pij<-array(0,dim=c(ncol(x),nrow(x),nrow(x)))
  aij<-array(0,dim=c(ncol(x),nrow(x),nrow(x)))
  t<-rowSums(x)
  p<-x/t
  maxPerA<-max(perimeter/area)
  for (k in 1:ncol(x))
  {
    for (i in 1:nrow(x)) pij[k,,i]<-p[i,k]
    for (i in 1:nrow(x)) pij[k,i,]<-abs(pij[k,i,]-p[i,k])
    for (i in 1:nrow(x)) PerAij[k,,i]<-perimeter[i]/area[i]
    for (i in 1:nrow(x)) PerAij[k,i,]<-PerAij[k,i,]+perimeter[i]/area[i]
    for (i in 1:nrow(x)) aij[k,,i]<-altitude[i]
    for (i in 1:nrow(x)) aij[k,i,]<-abs(aij[k,i,]-altitude[i])
    maxPerAlt<-max(altitude)
    aij[k,,]<-1-(aij[k,,]/maxPerAlt)
    matprovi<-c%*(pij[k,,]*PerAij[k,,]*aij[k,,])
    matprovi<-matprovi*diag(nrow(x))
    result[k]<-IS[k]-0.125*sum(matprovi)/((maxPerA)*sum(c))
  }
  return(result)
}

```

ISW(x)

Nouvel indice basé sur Wong, les adjacences, l'indice de Gorard et les altitudes comprises dans le rapport périmètre/surface

```

setwd("C:/Users/Kevin/Desktop/Tyrol")
rm(list = ls())
Donnees = read.csv2("x.csv", header = TRUE, row.names = 1) #habitants par communes
PER = read.csv2("Perimetres.csv", header = TRUE, dec=".", sep=";", stringsAsFactors = FALSE)
SUR = read.csv2("Surfaces.csv", header = TRUE, dec=".", sep=";", stringsAsFactors = FALSE)

```

```

ALT = read.csv2("Altitudes.csv", header = TRUE)
ADJ = read.csv2("ADJ.csv", header = FALSE)

ISW<-function(x, f, perimeter, area)
{
  x<-as.matrix(Donnees)
  c<-as.matrix(ADJ)
  perimeter<-as.matrix(PER)
  perimeter<-as.vector(perimeter)
  area<-as.matrix(SUR)
  area<-as.vector(area)
  altitude<-as.matrix(ALT)
  altitude<-as.vector(altitude)
  IS<-vector(length=ncol(x))
  tx<-rowSums(x)
  varTotal<-colSums(x)
  Total<-sum(x)
  for (k in 1:ncol(x)) IS[k]<-0.5*sum(abs(x[,k]/varTotal[k]-tx/Total))
  result<-vector(length=ncol(x))
  PerAij<-array(0,dim=c(ncol(x),nrow(x),nrow(x)))
  pij<-array(0,dim=c(ncol(x),nrow(x),nrow(x)))
  aij<-array(0,dim=c(ncol(x),nrow(x),nrow(x)))
  t<-rowSums(x)
  p<-x/t
  maxPerA<-max(perimeter*altitude/area)
  for (k in 1:ncol(x))
  {
    for (i in 1:nrow(x)) pij[k,,i]<-p[i,k]
    for (i in 1:nrow(x)) pij[k,i,]<-abs(pij[k,i,]-p[i,k])
    for (i in 1:nrow(x)) PerAij[k,,i]<-perimeter[i]*altitude[i]/area[i]
    for (i in 1:nrow(x)) PerAij[k,i,]<-PerAij[k,i,]+perimeter[i]*altitude[i]/area[i]
    for (i in 1:nrow(x)) aij[k,,i]<-altitude[i]
    for (i in 1:nrow(x)) aij[k,i,]<-aij[k,i,]-altitude[i]
    for (i in 1:nrow(x)) aij[k,,i]<-aij[k,,i]+altitude[i]
    aij[k,,]<-(aij[k,i,]/aij[k,,i])
    matprovi<-c%*(pij[k,,]*PerAij[k,,]*aij[k,,])
    matprovi<-matprovi*diag(nrow(x))
    result[k]<-IS[k]-0.25*sum(matprovi)/((maxPerA)*sum(c))
  }
  return(result)
}

ISW(x)

```

Nouvel indice basé sur Wong, les longueurs intercommunales, l'indice de Gorard et les altitudes non comprises dans le rapport périmètre/surface

```

setwd("C:/Users/Kevin/Desktop/Tyrol")
Donnees = read.csv2("x.csv", header = TRUE, row.names = 1) #habitants par communes
ADJ = read.csv2("LO.csv", header = TRUE, dec=".", stringsAsFactors = FALSE)
PER = read.csv2("Perimetres.csv", header = TRUE, dec=".", sep=";", stringsAsFactors = FALSE)
SUR = read.csv2("Surfaces.csv", header = TRUE, dec=".", sep=";", stringsAsFactors = FALSE)
ALT = read.csv2("Altitudes.csv", header = TRUE)

ISW<-function(x, f, perimeter, area, altitude)
{
  x<-as.matrix(Donnees)
  f<-as.matrix(ADJ)
  perimeter<-as.matrix(PER)
  perimeter<-as.vector(perimeter)
  area<-as.matrix(SUR)
  area<-as.vector(area)
  altitude<-as.matrix(ALT)
  altitude<-as.vector(altitude)
  IS<-vector(length=ncol(x))

```

```

Total<-sum(x)
tx<-rowSums(x)
varTotal<-colSums(x)
Total<-sum(x)
for (k in 1:ncol(x)) IS[k]<-0.5*sum(abs(x[,k]/varTotal[k]-tx/Total))
result<-vector(length=ncol(x))
PerAij<-array(0,dim=c(ncol(x),nrow(x),nrow(x)))
pij<-array(0,dim=c(ncol(x),nrow(x),nrow(x)))
aij<-array(0,dim=c(ncol(x),nrow(x),nrow(x)))
t<-rowSums(x)
p<-x/t
maxPerA<-max(perimeter/area)
w<-f/perimeter
for (k in 1:ncol(x))
{
  for (i in 1:nrow(x)) pij[k,,i]<-p[i,k]
  for (i in 1:nrow(x)) pij[k,i,]<-abs(pij[k,i,]-p[i,k])
  for (i in 1:nrow(x)) PerAij[k,,i]<-perimeter[i]/area[i]
  for (i in 1:nrow(x)) PerAij[k,i,]<-PerAij[k,i,]+perimeter[i]/area[i]
  for (i in 1:nrow(x)) aij[k,,i]<-altitude[i]
  for (i in 1:nrow(x)) aij[k,i,]<-abs(aij[k,i,]-altitude[i])
  maxPerAlt<-max(aij[k,i,])
  aij[k,,]<-(aij[k,i,]/maxPerAlt)
  matprovi<-(w/sum(w))%*(pij[k,,]*PerAij[k,,]*aij[k,,])
  matprovi<-matprovi*diag(nrow(x))
  result[k]<-IS[k]-0.25*sum(matprovi)/maxPerA
}
}
return(result)
}
ISW(x)

```

Matrice d'adjacence

```

# paquetages nécessaires
library("rgeos")
library("rgdal")

# répertoire de travail
setwd("C:/Users/Kevin/Desktop/Tyrol")
wd <- getwd()
regions <- readOGR(dsn = wd, layer = "Tyr")

# Détection des overlapping (superposition) des polygones indiquant une
# géométrie inexacte. Un tout petit décalage dans le tracé peut engendrer
# cette erreur. On doit donc vérifier que sum(gOverlaps.....) = 0
sum(gOverlaps(regions, byid=TRUE))

regionsAdjMatrix <- gIntersects(regions, byid = TRUE)
all(rowSums(regionsAdjMatrix) >= 2)

# Dataframes des matrices d'adjacence dont les noms de lignes et de colonnes
# correspondent au geocode
dfRegions <- data.frame(regionsAdjMatrix)
colnames(dfRegions) <- c(regions$geocode)
rownames(dfRegions) <- c(regions$geocode)

# Exportation des matrices d'adjacences (csv)
write.csv(dfRegions, file = 'adj_regions.csv')

```

matrice d'échange

```

# OUTPUT :
# E(f,A), through the four variants :

```



```

# variant 1: diffusive algorithm (with an extra parameter t>0)
rm(list = ls())
setwd("C:/Users/Kevin/Desktop/Tyrol")
A = read.csv2("ADJ.csv",header = FALSE) # coder A en utf8
A=as.matrix(A)
A=A-diag(diag(A)) # ote les valeurs de 1 sur la diagonale de A
names(A)
n=read.csv2("x.csv", header = TRUE, row.names = 1)
n=as.matrix(n)
n=rowSums(n)
t=1
f=rep(n,1)/sum(n) #petite modification de la formule. rep demande un vecteur puis un temps le
  tout pondéré par le total
sum(f) #doit être égale à un

E1=function(f,A,t){
  LapA=diag(rowSums(A))-A
  Psi=(diag(sqrt(1/f))%*%LapA%*%diag(sqrt(1/f)))/(sum(A)-sum(diag(A))) #normalised
  essentially positive generator
  U=eigen(Psi)$vectors
  Gamma=eigen(Psi)$values
  E=diag(sqrt(f))%*%U%*%diag(exp(-t*Gamma))%*%t(U)%*%diag(sqrt(f))
  return(E)
}

E=E1(f,A,1) # E

E[1:10,1:10]
sum(diag(E)) # proportion des stayers (par opposition a "movers")
min(E)

E=pmax(E,0) # remplace les valeurs negatives (imprecisions numeriques) par zero
sum(E) # doit etre egal a 1

rowSums(A) # degre

```

Indices de Moran locaux

```

#importation des tableaux de données (dataframes)
setwd("C:/Users/Kevin/Desktop/Tyrol/script20_2_17")
donnees_ger <- read.csv2("Donnees_ger.csv", row.names = 1, stringsAsFactors = FALSE)
# matrices avec uniquement les germanophones
library(magrittr)
library(dplyr)
# noms de lignes (Communes)

# paquetages nécessaires
library("stats")
# matrices des distances euclidiennes carrées
as.matrix(donnees_ger)
donnees_ger %>% dist(method = "euclidean") %>% .^2 -> D

# fonction: décomposition de la matrice Psi
decompo_psi <- fonction(D, A, f){
  # nb d'individus
  n <- length(f)
  # distances comme matrice
  D %>% as.matrix
  # matrice adjacence
  G <- as.matrix(A)
  # Laplacien de G LapG
  LapG <- diag(rowSums(G))-G
  # Psi

```

```

Psi <- (diag(sqrt(1/f)) %*% LapG %*% diag(sqrt(1/f))) / (sum(G)-sum(diag(G)))
# décomposition spectrale de Psi
U <- eigen(Psi)$vectors
Gamma <- eigen(Psi)$values
list(f=f, n=n, D=D, G=G, LapG=LapG, Psi=Psi, U=U, Gamma=Gamma) %>% return
}

x <- decompo_psi(D,A,f)

# x correspond à la liste de la fonction decompo_psi
# fonction : calcul delta et z-scores
delta_z <- function(t, x){
  # matrice d'échange
  E <- diag(sqrt(x$f)) %*% x$U %*% diag(exp(-t*x$Gamma)) %*% t(x$U) %*% diag(sqrt(x$f))
  # inertie globale pondérée
  Delta <- 0.5*t(x$f) %*% x$D %*% x$f
  # inertie locale
  DeltaLoc <- 0.5*sum(x$D+E)
  # autocorrelation relative
  delta <- (Delta-DeltaLoc) / Delta
  # matrice de transition de la chaîne de Markov
  W <- diag(1/x$f) %*% E
  # espérance et variance de delta sous HO
  Edelta <- ( sum(diag(W)) - 1) / (x$n-1)
  Vdelta <- 2 * (sum(diag(W %*% W)) - 1 - (sum(diag(W))-1)^2/(x$n-1) ) / (x$n^2 - 1)
  # z-score
  zScore <- (delta - Edelta) / sqrt(Vdelta)
  list(d=delta, z=zScore) %>% return
}

# on cherche le paramètre t pr lequel le z_score est maximal
data_frame(t = t_, z_score = z_) %>% filter(z_score == max(z_)) %>% print

# on cherche le paramètre t pr lequel le z_score est minimal
data_frame(t = t_, z_score = z_) %>% filter(z_score == min(z_)) %>% print

# on cherche le paramètre t pr lequel le z_score est minimal
data_frame(t = t_, z_score = z_) %>% filter(z_score == min(z_)) %>% print

# fonction: calcul des indices locaux d'autocorrélation spatiale
delta_i <- function(t, x){
  # matrice d'échange
  E <- diag(sqrt(x$f)) %*% x$U %*% diag(exp(-t*x$Gamma)) %*% t(x$U) %*% diag(sqrt(x$f))
  # inertie globale
  Delta <- 0.5*t(x$f) %*% x$D %*% x$f
  # matrice de transition de la chaîne de Markov
  W <- diag(1/x$f) %*% E
  # indices locaux d'autocorrélation spatiale
  delta_i <- ( as.vector(x$D %*% x$f) - diag(W %*% x$D) ) / (2*Delta)
  return(delta_i)
}

Ger_scores <- delta_i(t,x)
delta_z(t,x)
Ger_scores

#Permet d'explorer l'index des minima
which(Ger_scores %in% sort(unique(Ger_scores)))[1:3])

#Permet d'explorer les valeurs des minima
Ger_scores[ which(Ger_scores %in% sort(Ger_scores)[1:3])]

a<-data.frame(donnees_ger, Ger_scores)
write.csv(a, file = 'Delta_i_GER.csv')

```

nuage de points de Moran avec Wx et x

```
library(ggplot2)
setwd("C:/Users/Kevin/Desktop/Tyrol/script20_2_17")
v=read.csv2("Donnees_ger.csv", header = TRUE, row.names=1)
h=read.csv2("Noms.csv", header=TRUE, sep=",")
v=as.matrix(v)
v=rep(v,1) #petite modification de la formule. rep demande un vecteur puis un temps le tout
           pondéré par le total
v=as.numeric(v)
sum(v)
d<-data.frame(h,v)
t=1

moran_plot <- function(t, v, x){
  E <- diag(sqrt(x$f)) %*% x$U %*% diag(exp(-t*x$Gamma)) %*% t(x$U) %*% diag(sqrt(x$f))
  W <- diag(1/x$f) %*% E
  y <- W %*% v
  qplot(v,
        y,
        label=d$COMUNE,
        geom = c("point"),
        method = "lm",
        formula = "y~v",
        se = FALSE
  ) +
  labs(x = "x", y = "Wx", size=0.2)+geom_vline(xintercept = mean(v))+geom_hline(yintercept =
    mean(y))+ stat_smooth( method = "lm", fomula = "delta_i(t, _x)~v", se = FALSE, aes(
    weight=f))
}

moran_plot(t,v,x)+geom_text(size = 3, check_overlap = T)
```

Nuage de points de Moran avec les I de Moran et x

```
library(ggplot2)
setwd("C:/Users/Kevin/Desktop/Tyrol/script20_2_17")
v=read.csv2("Donnees_ger.csv", header = TRUE, row.names=1)
h=read.csv2("Noms.csv", header=TRUE, sep=",")
v=as.matrix(v)
v=rep(v,1) #petite modification de la formule. rep demande un vecteur puis un temps le tout
           pondéré par le total
v=as.numeric(v)
sum(v)
d<-data.frame(h,v)

moran_plot <- function(t, v, x){
  E <- diag(sqrt(x$f)) %*% x$U %*% diag(exp(-t*x$Gamma)) %*% t(x$U) %*% diag(sqrt(x$f))
  W <- diag(1/x$f) %*% E
  y <- W %*% v
  R<-v-y
  qplot(v,
        delta_i(t, x),
        label=d$COMUNE,
        geom = c("point"),
        method = "lm",
        fomula = "delta_i(t, _x)~v",
        se = FALSE
  ) +
  labs(x = "x", y = "delta_i", size=0.2)+geom_vline(xintercept = mean(v))+geom_hline(
    yintercept = mean(delta_i(t, x))+ stat_smooth( method = "lm", fomula = "delta_i(t, _x)
    ~v", se = FALSE, aes(weight=f))
}

moran_plot(t,v,x)+geom_text(size = 3, check_overlap = T)
```

Nuage de points de Moran avec les I de Moran et Wx

```
library(ggplot2)
setwd("C:/Users/Kevin/Desktop/Tyrol/script20_2_17")
v=read.csv2("Donnees_ger.csv", header = TRUE, row.names=1)
h=read.csv2("Noms.csv", header=TRUE, sep=",")
v=as.matrix(v)
v=rep(v,1) #petite modification de la formule. rep demande un vecteur puis un temps le tout
           pondéré par le total
v=as.numeric(v)
sum(v)
d<-data.frame(h,v)

moran_plot <- function(t, v, x){
  E <- diag(sqrt(x$f)) %*% x$U %*% diag(exp(-t*x$Gamma)) %*% t(x$U) %*% diag(sqrt(x$f))
  W <- diag(1/x$f) %*% E
  y <- W %*% v
  R<-v-y
  qplot(y,
        delta_i(t, x),
        label=d$COMUNE,
        geom = c("point"),
        method = "lm",
        fomula = "delta_i(t,_x)~y",
        se = FALSE
  ) +
  labs(x = "Wx", y = "delta_i", size=0.2)+geom_vline(xintercept = mean(y))+geom_hline(
    yintercept = mean(delta_i(t, x))+ stat_smooth( method = "lm", fomula = "delta_i(t,_x)
    ~y", se = FALSE, aes(weight=f))
  )
}

moran_plot(t,v,x)+geom_text(size = 3, check_overlap = T)
```

Nuage de points de Moran avec les I de Moran et R

```
library(ggplot2)
setwd("C:/Users/Kevin/Desktop/Tyrol/script20_2_17")
v=read.csv2("Donnees_ger.csv", header = TRUE, row.names=1)
h=read.csv2("Noms.csv", header=TRUE, sep=",")
v=as.matrix(v)
v=rep(v,1) #petite modification de la formule. rep demande un vecteur puis un temps le tout
           pondéré par le total
v=as.numeric(v)
sum(v)
d<-data.frame(h,v)

moran_plot <- function(t, v, x){
  E <- diag(sqrt(x$f)) %*% x$U %*% diag(exp(-t*x$Gamma)) %*% t(x$U) %*% diag(sqrt(x$f))
  W <- diag(1/x$f) %*% E
  y <- W %*% v
  R<-v-y
  qplot(R,
        delta_i(t, x),
        label=d$COMUNE,
        geom = c("point"),
        method = "lm",
        fomula = "delta_i(t,_x)~R",
        se = FALSE
  ) +
  labs(x = "R", y = "delta_i", size=0.2)+geom_vline(xintercept = mean(R))+geom_hline(
    yintercept = mean(delta_i(t, x))+ stat_smooth( method = "lm", fomula = "delta_i(t,_x)
    ~R", se = FALSE, aes(weight=f))
  )
}
```

```
moran_plot(t,v,x)+geom_text(size = 3, check_overlap = T)
```

Boîtes à moustache avec moyenne

```
a<-matrix(c(10, 10, 10, 10, 9, 8, 10, 9, 10, 10,
            10, 10, 10, 10, 9, 8, 10, 10, 10, 10,
            8, 9, 7, 10, 8, 7, 7, 8, 8, 9,
            7, 10, 8, 9, 7, 8, 8, 7, 8, 10,
            10, 10, 6, 9, 7, 3, 10, 8, 6, 10,
            10, 10, 8, 10, 8, 7, 10, 10, 9, 10,
            7, 10, 8, 10, 9, 8, 4, 10, 9, 10,
            9, 10, 8, 10, 9, 7, 10, 8, 9, 10,
            8, 8, 8, 10, 10, 9, 8, 10, 10, 10,
            8, 9, 8, 10, 9, 8, 8, 10, 10, 10,
            8, 9, 8, 10, 7, 8, 8, 8, 9, 10), nrow=10, ncol=11, dimnames = list(letters
            [1:10], c("Affichage", "Interaction", "Fonctionnalité", "Statistiques", "
            Diagramme", "Fiabilité", "Complet", "Réutilisation", "Curiosité", "
            Satisfaction", "Instructivité")))
c<-boxplot(a, col = rainbow(11), ylab = "Notes", las=2, cex.axis=0.2, axes=FALSE, ann=FALSE)
points(seq(c$n), colMeans(a), pch = 17)
#unnname(c)
title(main = list("Résumé_de_la_visualisation_<<autocorrélations_spatiales>>_", cex = 1.5,
                  col = "black", font = 3))
d<-list("Affichage", "Interaction", "Fonctionnalité", "Statistiques", "Diagramme", "Fiabilité",
        "Complet", "Réutilisation", "Curiosité", "Satisfaction", "Instructivité")
#axis(1, at=0:10, labels=d)
lablist<-as.vector(d)
axis(1, at=seq(1, 11, by=1), labels = FALSE)
text(seq(1, 11, by=1), par("usr")[3] - 0.4, labels = lablist, srt = 45, pos = 2, xpd = TRUE,
      cex=0.9, adj = 0.5)
axis(2, at=0:10, las=1)
mtext("Caractéristiques_de_la_visualisation", line = -16.5)
```

Test de Student

```
a<-matrix(c(10, 10, 10, 10, 9, 8, 10, 9, 10, 10,
            10, 10, 10, 10, 9, 8, 10, 10, 10, 10,
            8, 9, 7, 10, 8, 7, 7, 8, 8, 9,
            7, 10, 8, 9, 7, 8, 8, 7, 8, 10,
            10, 10, 6, 9, 7, 3, 10, 8, 6, 10,
            10, 10, 8, 10, 8, 7, 10, 10, 9, 10,
            7, 10, 8, 10, 9, 8, 4, 10, 9, 10,
            9, 10, 8, 10, 9, 7, 10, 8, 9, 10,
            8, 8, 8, 10,10,9, 8, 10, 10, 10,
            8, 9, 8, 10, 9, 8, 8, 10, 10, 10,
            8, 9, 8, 10, 7, 8, 8, 8, 9, 10), nrow=10, ncol=11, dimnames = list(letters
            [1:10], c("Affichage", "Interaction", "Fonctionnalité", "Statistiques", "
            Diagramme", "Fiabilité", "Complet", "Réutilisation", "Curiosité", "
            Satisfaction", "Instructivité")))

b<-matrix(c(10, 10, 10, 10, 10, 8, 9, 8, 10, 10,
            9, 10, 10, 10, 10, 8, 9, 8, 10, 10,
            8, 8, 9, 8, 8, 8, 9, 7, 10, 10,
            9, 10, 10, 10, 10,10,9, 7, 10, 10,
            8, 8, 9, 10, 9, 9, 6, 8, 10, 8,
            10, 9, 8, 10, 10,10,9, 9, 10, 10,
            9, 8, 10, 10, 10, 9, 6, 8, 10, 10,
            10, 9, 10, 9, 9, 9, 9, 8, 10, 10,
            10, 8, 9, 10, 8, 9, 9, 9, 10, 10,
            8, 8, 10, 10, 9, 9, 8, 8, 10, 10,
            8, 10, 7, 8, 9, 9, 8, 8, 10, 10), nrow=10, ncol=11, dimnames = list(letters
            [1:10], c("Affichage", "Interaction", "Fonctionnalité", "Statistiques", "
```

```
Diagramme", "Fiabilité", "Complet", "Réutilisation", "Curiosité", "Satisfaction", "Instructivité"))
```

```
f <- function(x,y){  
  test <- t.test(x,y, paired=TRUE)  
  out <- data.frame(stat = test$statistic ,  
                    df = test$parameter ,  
                    pval = test$p.value ,  
                    conl = test$conf.int[1],  
                    conh = test$conf.int[2],  
                    esti = test$estimate ,  
                    nulv = test$null.value  
  )  
  return(out)  
}  
  
sapply(seq(ncol(a)), function(x) f(a[,x], b[,x]))  
var(a[,1])-var(b[,1])  
var(a[,2])-var(b[,2])  
var(a[,3])-var(b[,3])  
var(a[,4])-var(b[,4])  
var(a[,5])-var(b[,5])  
var(a[,6])-var(b[,6])  
var(a[,7])-var(b[,7])  
var(a[,8])-var(b[,8])  
var(a[,9])-var(b[,9])  
var(a[,10])-var(b[,10])  
var(a[,11])-var(b[,11])
```

7 Bibliographie

Références

- [1] Agafonkin, V. (2017). Interactive ChoroplethMap. *Leaflet*. Disponible sur <https://leafletjs.com/examples/choropleth/> (consulté le 16.08.2017)
- [2] Apparicio, P. (2000). Les indices de ségrégation résidentielle : un outil intégré dans un système d'information géographique. *Cybergeo : european journal of geography*. DOI : 10.4000/cybergeo.12063
- [3] Anselin, L. (1995). Local Indicators of Spatial Association – LISA. *Geographical Analysis*. 27(2). 93-115. DOI : 10.1111/j.1538-4632.1995.tb00338.x
- [4] Aubry, P., Piégay, H. (2001). Pratique de l'analyse de l'autocorrélation spatiale en géomorphologie : définitions opératoires et tests. *Géographie physique et Quaternaire*. 55(2). 111. DOI : 10.7202/008297ar
- [5] Autonome Provinz Bozen SüdTirol/Provincia autonoma di Bolzano Alto Adige-Landesinstitut für Statistik ASTAT (2015). *1-Sprachgruppen nach Gemeinden (%) – Volkszählungen 1971-2011*. Disponible sur <http://qlikview.services.sdiag.it/QvAJAXZfc/opensdoc.htm?document=gemeindedatenblatt.qvw&host=QVS%40titan-a&anonymous=true>
- [6] Autonome Provinz Bozen SüdTirol/Provincia autonoma di Bolzano Alto Adige-Landesinstitut für Statistik ASTAT (2015). *Bevölkerung 3/Popolazione 3. Bolzano/Bozen* : Auteur. Disponible sur http://www.provinz.bz.it/astat/download/JB2015_K3.pdf
- [7] Bavaud, F. (2014). Spatial weights : constructing weight-compatible exchange matrices from proximity matrices. Dans : M. Duckham, E. Pebesma, K. Stewart, A.U. Frank (Eds.) : *Geographic Information Science. GIScience 2014* (pp.81-96), Lecture Notes in Computer Science, vol 8728. Springer, Cham. DOI : https://doi.org/10.1007/978-3-319-11593-1_6
- [8] Bavaud, F. (2015). *Auto-corrélation spatiale* [Support de cours]. Lausanne : Université, Institut de géographie et durabilité.
- [9] Bretagnolle, A. (1996). Étude des indices de concentration d'une population. *Espace géographique*, 25(2), 145-157.
- [10] Chesneau, C. (2017). Sur l'Estimateur du Maximum de Vraisemblance (emv) [Document de cours], Université de Caen, Unité de formation et de recherche (URF) des sciences, Laboratoire de Mathématiques Nicolas Oresme (LMNO), Caen, France.
- [11] Chessel, D., Thioulouse, J. (2003). Analyse de données spatialisées. *Fiche de biostatistique, Biométrie et Biologie Evolutive*, Université Lyon1.
- [12] Dreamstime (s.d.). Noir De Globe de la Terre 3d Illustration Stock. *Dreamstime*. Disponible sur <https://thumbs.dreamstime.com/b/noir-de-globe-de-la-terre-d-39174055.jpg> (consulté le 17 juillet 2018)

- [13] Duncan, O.D, Duncan, B. (1955). A methodological analysis of segregation indexes. *American sociological review*, 20(2), 210-217.
- [14] Essennsee (s.d.). Ternary Plot. *JsFiddle*. Disponible sur <http://jsfiddle.net/essennsee/y9yask5n/5/> (consulté le 18.09.2017)
- [15] FernOfTheAndes (2014, 4 avril). Adding tooltip elements after updating data using d3 enter method. *Stackoverflow*. Disponible sur <https://stackoverflow.com/questions/22867573/adding-tooltip-elements-after-updating-data-using-d3-enter-method> (consulté le 25.10.2017)
- [16] Gobbi, G., Thurner, B. (2013). *Statistisches Jahrbuch 2012 / Annuario statistico 2012* :Autonome Provinz Bozen Südtirol/Provincia autonoma di Bolzano Alto Adige-Landesinstitut für Statistik (ASTAT). Disponible sur : http://astat.provinz.bz.it/downloads/jahrbuch_2012.pdf (consulté le 07.06.2017)
- [17] Gorard, S., Taylor, C. (2002). What is segregation ? A comparison of measures in terms of 'strong'and 'weak'compositional invariance. *Sociology*, 36(4), 875-895.
- [18] Greenwald, A.G.(1976). Within-subjects designs : To use or not to use ? *Psychological Bulletin*, 83, 314–320.
- [19] Hijmans, R. (2009). *Download. Country : Italy. Format : Shape* [en ligne]. Disponible sur <http://www.gadm.org/download>. (Consulté le 18.06.2016)
- [20] Hobsbawm, E. (1993). Qu'est-ce qu'un conflit ethnique ?. *Actes de la recherche en sciences sociales*, 100(1), 51-57.
- [21] Houdebine, M. (1999). Concentration géographique des activités et spécialisation des départements français. *Economie et statistique*, 326(1), 189-204.
- [22] Istituto nazionale di statistica (ISTAT) (2017). *Principali statistiche geografiche sui comuni* [Page Web]. Disponible sur <https://www.istat.it/it/archivio/156224> (Consulté le 07.06.2017)
- [23] Istituto nazionale di statistica (ISTAT) (2017, 11 janvier). *Archivio dei confini delle unità amministrative a fini statistici* [Page Web]. Disponible sur <https://www.istat.it/it/archivio/124086> (Consulté le 30.06.2017)
- [24] Jouglà, E. (1997). Tests statistiques relatifs aux indicateurs de mortalité en population. *Revue d'épidémiologie et de santé publique*, 45(1), 78-84.
- [25] Kirby, G. (1985). Zipf's law. *UK Journal of Naval Science*, 10(3), 180-185.
- [26] Lange, M. (2014). *Widged/index.html*. *Github*. Repéré à <https://gist.github.com/widged/5780720> (consulté le 18.09.2017)

- [27] Légal, J.-B.(s.d.). *Les plans expérimentaux*(fichier électronique). Université Paris X-Nanterre, Faculté de psychologie, EA 4386 - Laboratoire parisien de psychologie sociale (LAPPS), Paris, France. Disponible sur <http://j.b.legal.free.fr/Blog/share/Concepts/S%E9ance%20gr%E8ve%202.pdf>
- [28] Le Gallo, J. (2002). Économétrie spatiale : l'autocorrélation spatiale dans les modèles de régression linéaire. *Economie & prévision*, 155(4), 139-157.
- [29] Martin, D. W. (2008). How to Decide on a Between-Subjects vers Within-Subject Design. In *Doing psychology experiments* (7th ed., pp. 148-170). Belmont, Calif : Wadsworth Cengage Learning. DOI :10.13140/RG.2.1.1565.0000
- [30] Morill R. (1991). On the measure of geographic segregation, *Geography research forum*, 11, pp. 25-36.
- [31] Mpiktas (2017, 22 mai). Cross. How to calculate Zipf's law coefficient from a set of top frequencies ?. *Stack Exchange*. <https://stats.stackexchange.com/questions/6780/how-to-calculate-zipfs-law-coefficient-from-a-set-of-top-frequencies> (consulté le 23.05.2017)
- [32] Niwetsch (2018, 1 juin). segregation-spatiale. *Github*. Disponible sur : <https://github.com/Niwetsch/segregation-spatiale.git> (consulté le 1 juin 2018)
- [33] Office fédéral de la statistique (2016). *Intégration - Indicateurs : Logement - Indice de ségrégation spatiale* [Page Web]. Disponible sur <http://www.bfs.admin.ch/bfs/portal/fr/index/themen/01/07/blank/ind43.indicator.43045.430107.html?open=4301,439,434,104,438&close=438> (consulté le 15.05.16).
- [34] Oliveau, S. (2005). *Les indicateurs locaux d'association spatiale (LISA) comme méthode de régionalisation : Une application en Inde*. In 7ème rencontres ThéoQuant (pp. 1-10).
- [35] Oliveau, S. (2010). Autocorrélation spatiale : leçons du changement d'échelle. *L'Espace géographique*, 39(1), 51-64.
- [36] Oliveau, S. (2014). Autocorrélation spatiale. *Hypergéométrie*. Disponible sur <http://www.hypergeo.eu/spip.php?article696#> (Consulté le 23.06.2018)
- [37] Perron, J.-Y. (2003). *Tarif de cubage général : volume marchand brut*. Québec : Direction des inventaires forestiers. Disponible sur <https://www.mffp.gouv.qc.ca/forets/inventaire/pdf/tarif-de-cubage-64.pdf> (Consulté le 08.05.2018)
- [38] Poulton, E.C.(1973). Unwanted range effects from using within-subject experimental designs. *Psychological Bulletin*, 80, 113–121.
- [39] Poulton, E.C., Freeman, P.R.(1966). Unwanted asymmetrical transfer effects with balanced experimental designs. *Psychological Bulletin*, 66, 1–8.

- [40] Servizio statistica della Provincia Autonoma di Trento (2012). *Annuario statistico 2011* : Auteur. Disponible sur [http://www.statweb.provincia.tn.it/annuario/\(S\(cp0qe445dyujvy55ubqz14m2\)\)/Default.aspx](http://www.statweb.provincia.tn.it/annuario/(S(cp0qe445dyujvy55ubqz14m2))/Default.aspx) (Consulté le 07.06.2017)
- [41] Servizio statistica della Provincia Autonoma di Trento (2013). *15^e Censimento generale della popolazione e delle abitazioni. Popolazione residente al 31 dicembre 2011 (dati definitivi)* :Auteur. Disponible sur http://www.statistica.provincia.tn.it/binary/pat_statistica_new/popolazione/PopResidenteAl31122011.1364206667.pdf
- [42] Servizio statistica della Provincia Autonoma di Trento (2014). *Rilevazione sulla consistenza e la dislocazione territoriale degli appartenenti alle popolazioni di lingua ladina, mòchena e cimbra (15^e Censimento generale della popolazione e delle abitazioni - dati definitivi)* : Auteur. Disponible sur http://www.statistica.provincia.tn.it/binary/pat_statistica_new/popolazione/RilevazionePopolazioniLadinaMochenaCimbra.1394031752.pdf
- [43] Steinicke, E., Ladins, R. R. (2001). Potential for conflicts in areas of ethno-linguistic minorities of the Eastern Alps. *Annales*, 11(2), 26.
- [44] Tivadar, M.(2014, juin). Fonctions R : calcul des indices OASIS. *Irstea, DTGR Grenoble*. Disponible sur <https://oasis.irstea.fr/wp-content/uploads/2014/06/RFunctions1.html>. Consulté le 30.05.2017
- [45] Wong, D. W. (1993). Spatial indices of segregation. *Urban studies*, 30(3), 559-572.
- [46] Wong, D. W., Chong, W. K. (1998). Using spatial segregation measures in GIS and statistical modeling packages. *Urban Geography*, 19(5), 477-485.
- [47] W3Schools (2018). How – TO Popup. *W3Schools*. Disponible sur https://www.w3schools.com/howto/howto_js_popup.asp (consulté le 23.03.2018)

Références consultées

- [48] Diener, F., Diener, M. (2007-2008) *Chaînes de Markov : compléments* [Support de cours]. Nice : Université, Laboratoire J.-A. Dieudonné.
- [49] Anselin, L., Bera, A. K., Florax, R., Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional science and urban economics*, 26(1), 77-104.
- [50] Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Dormann, C.F., Hirzel, A., Jetz, W., Kissiling, W.D., Kühn, I., McPherson, J.M., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M., Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data : a review. *Ecography*, 30(5), 609-628.
- [51] Dörrenhaus, F. (1953). Deutsche und Italiener in Südtirol. *ERDKUNDE*, 7(3). DOI : [10.3112/erdkunde.1953.03.02](https://doi.org/10.3112/erdkunde.1953.03.02)
- [52] Getis, A., Ord, J. K. (1995). Local spatial autocorrelation statistics : distributional issues and an application. *Geographical analysis*, 27(4), 286-306.

- [53] Griffith, D. A. (1987). *Spatial autocorrelation. A Primer. Washington DC : Association of American Geographers.*
- [54] Meraner, R. (2004). Sprache, Sprachunterricht und Sprachenpolitik in Südtirol. *Bildung und Erziehung*, 57(1), 53-76.
- [55] Öhler, H. (2007). 14. *Allgemeine Volkszählung 2001 - Band 6 - Pendlerströme aus Berufs- und Studiengründen/14. Censimento generale della popolazione 2001 - Tomo 6 - Flussi pendolari per motivi di lavoro e studio. Bolzano/Bozen : Autonome Provinz Bozen Südtirol/Provincia autonoma di Bolzano Alto Adige-Landesinstitut für Statistik ASTAT.* Disponible sur https://www.google.ch/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwjYbM9fnMAhUD1hoKHUmQA1UQFggcMAA&url=http%3A%2F%2Fwww.provincia.bz.it%2Ffastat%2Fit%2Fmobilita-turismo%2F464.asp%3F%26VerkehrTransport_action%3D300%26VerkehrTransport_image_id%3D85028&usg=AFQjCNG4z1XpYezWTNP2wXZZ37RwUuZk0Q&cad=rja
- [56] Steinicke, E. (2007). *Sprachen und Kulturen : zur Zukunft von ethnischer Identität und demographischer Entwicklung in den Alpen.* Consulté à l'adresse http://www.uibk.ac.at/alpinerraum/publications/vol1/09_steinicke.pdf
- [57] Beismann, M., Löffler, R., Steinicke, E., Walder, J. (2011). Minorités linguistiques autochtones des Alpes italiennes. *Revue de Géographie Alpine | Journal of Alpine Research* [En ligne], 99-2. Disponible sur DOI : [10.4000/rga.1469](https://doi.org/10.4000/rga.1469)
- [58] Detruy, M. (2013, 28 octobre). Le Tyrol du Sud rêve d'Autriche. *Le Figaro* [en ligne]. Disponible sur <http://www.lefigaro.fr/international/2013/10/28/01003-20131028ARTFIG00279-le-tyrol-du-sud-reve-d-autriche.php>
- [59] Reboul, L. (2015). *II-Autocorrélation spatiale* [Support de cours]. Marseille : Université Aix-Marseille II, Institut de Mathématiques de Luminy.