

Rapport de stage licence professionnelle

# Rôle du contexte paysager des établissements scolaires sur les performances académiques des enfants

Sydol ADOMADE

*Université de Caen Normandie, UFR SEGGAT,*

*Licence pro systèmes d'information géographique, diagnostic et aménagement du territoire*

*Année 2022-2023*

## Encadrants :

Thierry FEUILLET

Gladys BARRAGAN-JASON

Maxime CAUCHOIX

## Remerciements

Je tiens à exprimer ma profonde gratitude envers toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce travail.

Tout d'abord, ma mère, Rachel AGOUGOU, mérite une reconnaissance spéciale pour son rôle inestimable dans mon parcours académique et personnel. Son amour inconditionnel, sa patience et son encouragement ont été des sources de motivation lorsque j'affrontais des défis apparemment insurmontables. Sa foi en moi a été un moteur puissant qui m'a incité à donner le meilleur de moi-même. Les sacrifices qu'elle a consentis pour mon éducation ne passent pas inaperçus, et je lui suis profondément reconnaissant pour tout ce qu'elle a fait pour moi.

J'adresse également mes sincères remerciements au professeur Thierry Feuillet pour sa précieuse contribution à la réalisation de mon mémoire. Son soutien et son expertise ont joué un rôle essentiel tout au long de ce projet, et je tiens à exprimer ma gratitude profonde et inestimable envers lui.

Madame Gladys Barragan-Jason, qui a généreusement partagé son expertise et son temps pour guider mon travail, mérite également toute ma reconnaissance. Son engagement envers mon projet a largement dépassé mes attentes. Ses conseils éclairés, ses discussions stimulantes et son dévouement à mon éducation ont été essentiels à sa réussite. Son impact sur ma formation académique est indéniable, et j'ai été honoré de travailler sous sa direction.

Monsieur Maxime Cauchoux, qui a mis à ma disposition son expérience et ses connaissances sans réserve, mérite une mention spéciale. Ses conseils critiques et ses encouragements ont été précieux, et sa disponibilité constante pour me guider tout au long de ce stage a été déterminante pour mon progrès. Je lui suis grandement reconnaissant pour sa contribution significative à mon parcours.

Je tiens également à exprimer ma gratitude envers tous mes enseignants de la licence professionnelle SIG-DAT qui ont contribué, de près ou de loin, à la réalisation de ce mémoire. Leur soutien, leur enseignement, leurs conseils et leurs ressources ont joué un rôle déterminant dans la réussite de ce projet ambitieux. Chaque personne ayant apporté sa contribution mérite ma sincère reconnaissance.

Ensemble, vous avez été une source d'inspiration et de soutien tout au long de mon parcours académique. Vos encouragements et votre expertise ont joué un rôle essentiel dans la réussite de mon mémoire, et je vous en suis infiniment reconnaissant. Votre impact sur ma vie restera gravé dans ma mémoire, et je m'efforcerai de continuer à honorer vos enseignements et votre soutien à l'avenir.

# Sommaire

<b>Remerciements</b> .....	1
<b>Introduction</b> .....	3
<b>I. STRUCTURE D’ACCEUIL</b> .....	4
<b>II. NOTRE TRAVAIL</b> .....	8
<b>III. CONSTRUCTION DES DONNEES</b> .....	9
<b>IV. ANALYSE</b> .....	21
<b>Conclusion</b> .....	49
<b>Bibliographie</b> .....	51
<b>Table des matières</b> .....	64
<b>Liste des figures</b> .....	52
<b>Liste des tableaux</b> .....	53
<b>Annexes</b> .....	54

# Introduction

Les adolescents, âgés de 13 à 14 ans, traversent une période critique de leur développement, à la fois sur le plan physique et psychologique. Ces faits ont été mis en évidence par des études menées par les professeurs d'université à Moncton sur « Les relations que les adolescents entretiennent avec leur environnement » (Annexe 6). C'est une phase où l'identité individuelle se forme, où l'apprentissage devient de plus en plus complexe, et où la prise de conscience de l'environnement qui les entoure évolue. C'est aussi une période où ils sont exposés à diverses influences, y compris technologiques, sociales et éducatives, qui peuvent parfois les éloigner de la nature. De nombreuses études ont mis en lumière les effets positifs de l'exposition à un environnement naturel sur la santé humaine, notamment sur les capacités cognitives (mémoire, attention, apprentissage)<sup>9</sup>. Cela signifie que le simple fait de passer du temps en contact avec la nature peut avoir un impact sur la manière dont les adolescents pensent, apprennent et résolvent des problèmes. Un environnement naturel favorise la réduction du stress, améliore la concentration, stimule la créativité et encourage la réflexion. En conséquence, l'amélioration du contact des adolescents avec la nature peut avoir des avantages significatifs. Au-delà des résultats scolaires, ces avantages incluent une meilleure gestion du stress, une capacité accrue à se concentrer en classe, une créativité accrue pour résoudre des problèmes, et un bien-être général. Ils peuvent également développer un respect et une compréhension plus profonde de l'environnement, ce qui peut avoir des implications à long terme pour leur engagement dans la préservation de la nature (Annexe 1 et Annexe 6). Il est essentiel de reconnaître l'importance du lien entre les adolescents et la nature, non seulement pour améliorer leurs résultats académiques, mais aussi pour favoriser leur épanouissement holistique en tant qu'individus. Cette recherche soulève des questions essentielles sur la manière dont l'éducation peut tirer parti de ces découvertes pour créer des environnements d'apprentissage plus enrichissants et équilibrés pour les jeunes générations.

Le contexte paysager d'un établissement scolaire considère l'ensemble des éléments naturels et artificiels qui le composent, à savoir les espaces verts, les cours de récréation, les bâtiments environnants, les rues adjacentes, et bien plus encore. De plus en plus, il est largement reconnu que ces éléments exercent une influence significative sur l'apprentissage, le bien-être et la santé des élèves. Une meilleure compréhension de l'impact du contexte paysager sur l'apprentissage des élèves est important à étudier. En analysant attentivement les caractéristiques du paysage environnant, les décideurs peuvent identifier les aspects qui favorisent ou entravent le développement des élèves. Cela ouvre la voie à des mesures spécifiques visant à créer des espaces éducatifs plus attrayants, inspirants et propices à l'apprentissage. L'évaluation du contexte paysager des établissements scolaires comprend plusieurs dimensions essentielles. Tout d'abord, il est primordial d'analyser la présence et la qualité des espaces verts. Les espaces verts, qu'il s'agisse d'arbres majestueux, de parcs paisibles ou de jardins éducatifs, offrent de nombreux avantages, tels que la réduction du stress, l'amélioration de la concentration et la promotion de l'activité physique<sup>5</sup>. En les intégrant de manière réfléchie dans les espaces scolaires, on peut créer un environnement qui favorise non seulement l'apprentissage, mais également le bien-être global des élèves.

L'objectif principal que nous poursuivons dans ce notre travail consiste à étudier le lien entre les éléments naturels, principalement la végétation aux abords des établissements scolaires, et certaines caractéristiques structurelles, et les résultats scolaires des enfants. Cette recherche vise à éclairer les relations subtiles entre l'environnement naturel et l'éducation, et à comprendre comment la présence de la nature peut influencer le développement cognitif des jeunes apprenants.

# **I. STRUCTURE D'ACCEUIL**

## **1. Le laboratoire UMR CNRS IDEES**

Le laboratoire IDEES Caen, spécialisé dans l'Identification et la Différenciation de l'Espace, de l'Environnement et des Sociétés, est un pôle de recherche affilié à l'UMR CNRS IDEES, qui opère dans trois villes françaises : Caen, Rouen et Le Havre. Cette unité mixte de recherche est associée au Centre National de Recherche Scientifique (CNRS) depuis 1984 et résulte de la fusion de plusieurs équipes de recherche, notamment le MTG et la LEDRA de l'université de Rouen, le CIRTAI de l'université du Havre et le GEOSYSCOM de l'université de Caen.

L'UMR CNRS IDEES réunit des enseignants-chercheurs, des chercheurs, des ingénieurs de recherche et des doctorants qui collaborent pour analyser une variété de phénomènes sociaux, environnementaux et spatiaux. Avec environ 120 chercheurs transdisciplinaires, le laboratoire joue un rôle central dans les sciences humaines et sociales en Normandie en raison de la diversité de ses projets de recherche et de ses collaborations interdisciplinaires.

Les objectifs de recherche du laboratoire UMR CNRS IDEES comprennent la compréhension des territoires et la prise en compte des vulnérabilités et des influences sociales qui les caractérisent. Le diagnostic des territoires est réalisé à l'aide de Systèmes d'Information Géographique (SIG), ce qui permet d'effectuer une analyse géographique basée sur des données factuelles et élaborées. Les enjeux liés à l'environnement et à l'aménagement du territoire sont également explorés en utilisant des méthodes géomatiques variées et précisément définies.

Les chercheurs de l'IDEES Caen abordent une multitude de thèmes de recherche, allant des risques auxquels les territoires sont exposés à l'aménagement des territoires. Ils travaillent sur des contextes géographiques diversifiés, qu'il s'agisse de la France, de l'Afrique, du Proche-Orient, et explorent des pistes innovantes, notamment dans le domaine de la géographie de la santé en utilisant des analyses spatiales, l'épidémiologie et l'étude des systèmes de santé pour comprendre les inégalités en matière de santé.

Globalement, le laboratoire UMR CNRS IDEES à Caen est un centre de recherche interdisciplinaire qui se consacre à l'étude des phénomènes sociaux, environnementaux et spatiaux à travers l'utilisation de méthodes géomatiques et de SIG, en abordant une variété de thèmes et de contextes géographiques, tout en favorisant la collaboration entre chercheurs.

## a) Les Axes de recherche IDEES Caen

Les quatre axes de recherche du laboratoire IDEES Caen offrent une vision complète et diversifiée des domaines d'intérêt et de spécialisation de ce laboratoire en sciences humaines et sociales :

**Sources, Modèles, Simulations (SMS) :** Cet axe se concentre sur le développement de projets de recherche utilisant des méthodes novatrices et encourage l'exploration de nouvelles méthodologies et théories. Il met particulièrement l'accent sur la modélisation des dynamiques socio-environnementales et la géoarchéologie. L'utilisation de modèles de simulation, de modélisation multi-modèles, et d'exploration HPC permet d'aborder des questions complexes liées à la gestion de l'eau et aux risques liés à l'eau.

**Interactions Environnements-Sociétés :** Cet axe adopte une perspective diachronique, systémique et multiscalaire pour étudier les relations entre la société et l'environnement. Il analyse l'évolution des phénomènes dans le temps, examine les interactions entre les éléments constitutifs d'un système, et prend en compte différentes échelles spatiales pour comprendre la complexité des phénomènes étudiés.

**Recompositions territoriales, Réseaux, Espaces Maritimes (ReTREM) :** Cet axe se penche sur les transformations territoriales en mettant l'accent sur les interactions entre les acteurs et les dynamiques spatiales. Il examine les changements économiques, sociaux, environnementaux et politiques qui influencent les territoires, en mettant en évidence le rôle des réseaux et des espaces maritimes. La collaboration avec des acteurs territoriaux et la production de connaissances scientifiques contribuent aux politiques publiques et au développement territorial durable.

**Inégalités, Parcours de vie et Action Publique :** Cet axe se focalise sur les dimensions sociales et vise à explorer de nouvelles pistes de recherche, notamment en lien avec les inégalités socio-économiques et les parcours de vie. Il favorise la collaboration interdisciplinaire et cherche à éclairer l'action publique en confrontant les recherches aux réalités sociales et spatiales. Il encourage le dialogue avec d'autres axes de recherche et cherche à saisir les réalités sociales sous différents angles d'approche.

Ces différents axes de recherche démontrent la diversité des domaines couverts par le laboratoire IDEES Caen, allant de la modélisation et la simulation à l'étude des interactions sociétés-environnements, en passant par l'analyse des transformations territoriales et la compréhension des inégalités sociales. Ils illustrent également l'engagement du laboratoire dans des approches interdisciplinaires et sa volonté de contribuer à la production de connaissances utiles à la société et aux politiques publiques.

### **b) Le Tuteur de Stage**

Le Professeur Thierry Feuillet, sous la direction duquel s'est déroulé notre stage, est un universitaire ayant une solide expérience dans le domaine de la géographie, en particulier dans l'analyse spatiale. Voici quelques éléments clés de son parcours académique et de ses activités de recherche :

- Il est titulaire d'un doctorat en géographie portant sur les "formes périglaciaires dans les Pyrénées centrales - analyse spatiale, chronologique et valorisation." Cette thèse montre son expertise dans la géographie physique et la compréhension des processus géomorphologiques.
- Le Professeur Feuillet a obtenu une habilitation à diriger des recherches en 2021, ce qui témoigne de sa capacité à superviser des recherches avancées et à encadrer des doctorants.
- Il est maître de conférences depuis 2015 à l'Université Paris 8 Vincennes-Saint-Denis, une institution renommée en France.
- Actuellement, il occupe le poste de professeur des universités à l'Université de Caen Normandie, ce qui représente un haut niveau de reconnaissance académique.
- Le Professeur Feuillet est un membre actif de l'UMR CNRS 6266 IDEES, démontrant ainsi son engagement dans les activités de recherche de ce laboratoire.
- En tant qu'enseignant de géomatique et d'analyse spatiale, il contribue à la formation des étudiants et des chercheurs dans ces domaines spécialisés.
- Le Professeur Feuillet est impliqué dans plusieurs projets de recherche, dont certains sont financés par l'Agence Nationale de la Recherche (ANR), ce qui témoigne de son implication dans la recherche de pointe. Les projets qu'il mène abordent des questions cruciales liées au changement climatique, à la santé publique, à la transition écologique et à d'autres domaines d'importance sociétale.

Le Professeur Thierry Feuillet est un géographe de renom avec une expertise diversifiée en géomatique, en analyse spatiale et en géographie physique. Son engagement dans des projets de recherche de haut niveau démontre son impact dans le domaine de la recherche en géographie et sa contribution à la compréhension des enjeux contemporains liés à l'environnement, à la santé et à la société.

## **2. Le Laboratoire SETE**

La Station d'Écologie Théorique et Expérimentale (SETE) est une unité de recherche du CNRS en France, qui a évolué au fil des années pour devenir un centre d'écologie expérimentale. La SETE a été créée en 1948 avec un focus initial sur l'étude des systèmes de grottes souterraines sous un angle géomorphologique et biologique.

- En 2007, la SETE a élargi son champ de recherche pour devenir une station d'écologie expérimentale.
- Depuis 2016, elle est connue sous le nom de "Station d'Écologie Théorique et Expérimentale".
- La SETE est actuellement dirigée par Camille Parmesan (depuis 2022).
- Elle compte environ 70 chercheurs, techniciens, personnel administratif et doctorants.
- La mission principale de la SETE est la création de nouvelles connaissances sur la biodiversité, les écosystèmes et leur interaction avec les sociétés en utilisant des approches théoriques et expérimentales.

- La SETE combine la théorie et l'expérimentation sur le terrain pour mieux comprendre les dynamiques des écosystèmes et contribuer à la gestion et à la conservation des écosystèmes dans leur ensemble.
- La SETE contribue à la formation de jeunes chercheurs en offrant des opportunités de stage, de doctorat et de recherche postdoctorale.

Deux chercheurs associés à la SETE ont participé à mon encadrement :

**Gladys Barragan-Jason** : Chargée de recherche en psychologie de la conservation au CNRS ; elle est spécialisée dans l'étude de la perception du changement climatique chez les enfants. Elle a collaboré à divers projets de recherche en France et à l'étranger, y compris à l'Université d'Ottawa au Canada.

**Maxime Cauchoix** : Directeur de recherche en écologie au CNRS SETE, il apporte son expertise en écologie à la recherche.

La SETE est donc un centre de recherche dédié à l'écologie qui mène des recherches approfondies sur la biodiversité, les écosystèmes et leur interaction avec les sociétés. Elle joue un rôle clé dans la formation de jeunes chercheurs et contribue à la compréhension des enjeux environnementaux contemporains. Les chercheurs associés à la SETE travaillent sur des sujets variés, de la perception du changement climatique à la biologie, en passant par la participation à des projets éducatifs.



## **II. NOTRE TRAVAIL**

Notre projet de recherche vise à explorer les liens entre l'environnement naturel autour des écoles en France et les résultats académiques des élèves. Il combine des aspects de la géographie, de l'environnement, de l'éducation et de l'analyse de données pour mieux comprendre l'impact de la nature sur l'apprentissage des enfants.

L'objectif fondamental est d'établir s'il existe un lien direct entre l'environnement naturel entourant les écoles en France et les performances académiques de ces écoles. Pour atteindre cet objectif, notre travail suit une approche méthodologique qui se détaille comme suit :

### **1. Collecte de données sur l'environnement des écoles**

Dans cette phase, nous collectons des données sur les écoles en France, en mettant l'accent sur les caractéristiques naturelles de leur environnement. Nous prenons en compte des métriques telles que la végétation, les températures moyennes et les précipitations moyennes. Nous utilisons des données géolocalisées pour obtenir des informations précises sur chaque école et son environnement naturel.

### **2. Établissement de liens**

Dans cette phase, nous explorons les relations entre les caractéristiques naturelles de l'environnement des écoles et diverses caractéristiques des écoles elles-mêmes, telles que le type d'établissement, le contexte urbain ou rural, etc. Cette analyse vous permettra de comprendre comment l'environnement naturel est associé aux caractéristiques des écoles.

### **3. Analyse des résultats académiques**

Nous établissons des corrélations entre les caractéristiques naturelles de l'environnement et les résultats académiques des élèves. Nous examinons comment la proximité d'espaces verts, la densité de la végétation et les conditions météorologiques peuvent influencer les performances académiques.

### **4. Conclusion et implications**

Nous tirerons des conclusions basées sur nos analyses et examinerons les implications de ces résultats.

Comment l'environnement naturel peut-il être optimisé pour favoriser l'apprentissage des élèves ? Comment cela peut-il informer les politiques éducatives et les aménagements environnementaux autour des écoles ?

Ce travail multidisciplinaire a le potentiel de fournir des informations précieuses sur l'interaction entre l'environnement naturel, l'éducation et les résultats académiques. Il pourrait également contribuer à éclairer les décisions politiques visant à créer des environnements scolaires plus propices à l'apprentissage et au bien-être des élèves.

### III. CONSTRUCTION DES DONNEES

#### 1. Les données utilisées

Lors de la réalisation de cette étude, nous avons mobilisé un éventail de données diversifié, adapté aux objectifs et aux résultats escomptés. Les données recueillies incluent une gamme d'éléments allant des caractéristiques naturelles aux indicateurs de performance des écoles mais aussi de facteurs sociaux et économiques. Les sources et les types de données utilisés sont donc variés et représentent une grande diversité.

- **Adresse et géolocalisation des établissements d'enseignement du premier et second degré**

Les données sur la géolocalisation des établissements d'enseignement en France sont disponibles au format GeoJSON. Ces données sont mises à disposition par le ministère de l'enseignement Français. Elles fournissent la localisation précise de plus de 65 000 écoles à travers le territoire français, y compris la France métropolitaine et les Communautés d'Outre-Mer. Ces données sont accompagnées de nombreuses informations détaillées sur les écoles, notamment leurs domaines d'enseignement spécialisés. Chaque établissement est identifié par un numéro d'identification unique appelé "numero\_uai", qui jouera un rôle important dans nos traitements ultérieurs.

Ces données sont représentées sous forme d'une couche de points géolocalisés sur toute la France. Le système de coordonnées utilisé est le Lambert 93 (epsg:2154), une projection cartographique spécifiquement conçue en France pour assurer une représentation précise et cohérente des coordonnées géographiques sur une carte. Cette projection se base sur la projection conique conforme de Lambert, qui permet de minimiser les distorsions dues à la représentation plane d'une surface courbée comme celle de la Terre. Dans le système Lambert 93, les coordonnées sont exprimées en mètres, ce qui permet de définir avec précision la position d'un point sur la carte.

La donnée a été créée le 14 novembre 2016, et nous avons utilisé la version la plus récente disponible au 15 avril 2023. Pour notre utilisation, nous avons respecté les termes de la licence ouverte v2.0, qui nous autorise à exploiter ces données librement.

- **Zones de végétation BD topo**

La couche "Zone de végétation" de la Base de Données Topographiques (BD Topo) en France fournit des informations détaillées sur la végétation et les espaces verts présents sur le territoire français. Elle couvre une large gamme de types de végétation, tels que les parcs, les jardins, les forêts, les espaces naturels, les haies et d'autres formes de couverture végétale.

La BD Topo représente les zones de végétation sous forme de polygones en délimitant avec précision les contours extérieurs de chaque zone. La couche utilise le système de coordonnées géographiques WGS84 (World Geodetic System 1984) comme référence de base. Le WGS84 est un système de coordonnées géodésiques mondial largement utilisé, qui approxime la forme de la Terre en utilisant un ellipsoïde de référence. Il s'agit d'un système de coordonnées géographiques développé par le département de la défense des Etats-Unis en collaboration avec l'Institut national des normes et de la technologie (NIST) et d'autres organisations internationales. Cette couche sera par la suite reprojetée en Lambert 93.

Dans le cadre de notre étude, vous avez utilisé la version de la BD Topo disponible au 21 avril 2023. Cette mise à jour nous donne accès aux informations les plus récentes sur la végétation et les espaces verts à cette date. Nous avons utilisé ces données conformément à la licence v3.0, qui autorise l'utilisation dans le cadre de notre étude.

- **Les précipitations et températures.**

Les données de températures et de précipitations que nous avons exploitées proviennent de la base de données Worldclim ([www.worldclim.org](http://www.worldclim.org)). Worldclim est une base de données qui offre gratuitement des données climatiques à haute résolution pour l'ensemble de la planète. Ces données sont issues de sources variées, telles que des stations météorologiques, des satellites et des modèles climatiques. Les informations recueillies à partir de ces différentes sources ont été croisées pour obtenir des moyennes de températures et de précipitations. Ainsi, les données fournies par Worldclim représentent une synthèse des informations climatiques provenant de diverses sources, garantissant une couverture étendue pour l'ensemble de la planète.

Les couches de températures et de précipitations historiques sont des fichiers raster avec une résolution de 1 kilomètre, ce qui représente la résolution la plus fine mise à disposition par la base de données. Chaque couche correspond à une moyenne mensuelle des variables climatiques pour un mois spécifique de l'année. En tout, le fichier téléchargé contient 12 couches raster différentes, chacune représentant un mois de l'année. Ces couches sont géoréférencées dans le système de coordonnées WGS 84. Les données exploitées sont la version la plus récente disponible au moment de leur utilisation, soit celle de janvier 2020. Ces informations climatiques nous permettent d'obtenir une vue d'ensemble assez globale du climat environnant de chaque établissement, offrant ainsi une perspective précieuse pour nos analyses et nos traitements ultérieurs.

- **Diplôme national du brevet par établissement**

Ce jeu de données est mis à disposition par le ministère de l'enseignement français et contient des informations détaillées sur les résultats des établissements aux examens du brevet. Les données couvrent la période allant de 2006 à 2021 et concernent toutes les régions de France métropolitaine, ainsi que les communautés d'outre-mer.

La création de ce jeu de données remonte au 10 janvier 2019, mais la version que nous avons exploitée a été mise à jour le 2 juin 2023. Cette mise à jour récente nous permet de bénéficier des données les plus actuelles concernant les résultats du brevet dans les établissements.

Le jeu de données est distribué sous la licence v2.0, ce qui nous autorise à l'exploiter dans notre cadre académique sans enfreindre les droits d'auteur ou les restrictions légales. La donnée est téléchargeable au format Excel sur le site « [data.education.gouv.fr](http://data.education.gouv.fr) ».

## 2. Tutoriel d'utilisation de la BD topo SQL France entière dans R.

La base de données topographique exhaustive couvrant l'intégralité du territoire français étant d'une envergure considérable, elle est mise à disposition sous forme de fichier SQL pour faciliter son téléchargement. Malheureusement, le package "sf" utilisé pour importer les couches de données dans l'environnement de programmation R ne dispose pas des pilotes requis pour interpréter directement une couche au format SQL. Afin de contourner cette limitation technique, une série de manipulations a été entreprise pour convertir la couche de données SQL en un format compatible avec R, à savoir le format géopackage, qui est plus couramment utilisé dans les applications géospatiales. Par le biais de ces opérations, les données issues de la base de données SQL ont été transposées dans un format adapté à l'analyse géospatiale en R, permettant ainsi leur intégration au sein de cet environnement de programmation.

Le processus débute par le téléchargement de la base de données depuis le site de l'Institut National de l'Information Géographique et Forestière (IGN). Cette base de données est un fichier énorme, compressé et divisé en huit parties, que nous téléchargeons successivement et stockons dans un même dossier. Une fois que les huit fichiers ont été téléchargés, nous procédons à leur décompression afin d'accéder au contenu. Les logiciels de décompression reconnaissent automatiquement les partitions, ce qui signifie qu'une fois qu'un fichier est décompressé, toutes les autres parties sont également prises en compte dans le processus de décompression. À l'issue de cette étape, nous obtenons un fichier unique d'une taille d'environ une centaine de giga-octets, qui regroupe l'ensemble des données de la base de données topographique, au format SQL. Ce fichier représente une mine d'informations géospatiales prêtes à être exploitées dans nos projets et analyses scientifiques.

La deuxième étape de notre démarche exige l'installation préalable du système de gestion de base de données relationnelles PostgreSQL. Il est donc crucial de vérifier que PostgreSQL est bien installé sur notre système pour pouvoir suivre cette méthode. Une fois que PostgreSQL est installé et que la base de données SQL est décompressée, nous procéderons à l'ajout des données qui nous intéressent, en l'occurrence les zones de végétation, à notre base de données PSQL. Cette opération se déroulera en ligne de commande sur notre ordinateur. Pour importer la couche sans altérer la colonne de géométrie, nous aurons besoin de certains paquets qui seront installés en ajoutant l'extension PostGIS à PostgreSQL au préalable.

La première étape à effectuer en ligne de commande consiste à changer le répertoire de travail pour accéder au fichier de ressources de PostgreSQL. En effet, cette manipulation permettra de se positionner dans l'environnement adéquat pour exécuter les commandes nécessaires à l'ajout des données dans notre base de données.

```
C:\Users\adosy>cd "C:\Program Files\PostgreSQL\15\bin\"
```

Avec la commande `cd`, nous nous déplaçons vers le répertoire "bin" de PostgreSQL, où sont situées les ressources essentielles de notre Système de Gestion de Base de Données Relationnelles (SGBDR).

L'étape suivante consiste à installer l'extension "postgis" qui est nécessaire pour le traitement des données spatiales. Pour cela, nous utilisons l'outil en ligne de commande "psql" qui permet d'interagir avec les bases de données PostgreSQL de manière interactive. En utilisant cette interface, nous exécutons la commande SQL "CREATE EXTENSION IF NOT EXISTS postgis" pour installer l'extension dans notre base de données.

```
C:\Users\adosy>psql -h localhost -p 5432 -U postgres -d bd_topo -c "CREATE EXTENSION IF NOT EXISTS postgis;"
```

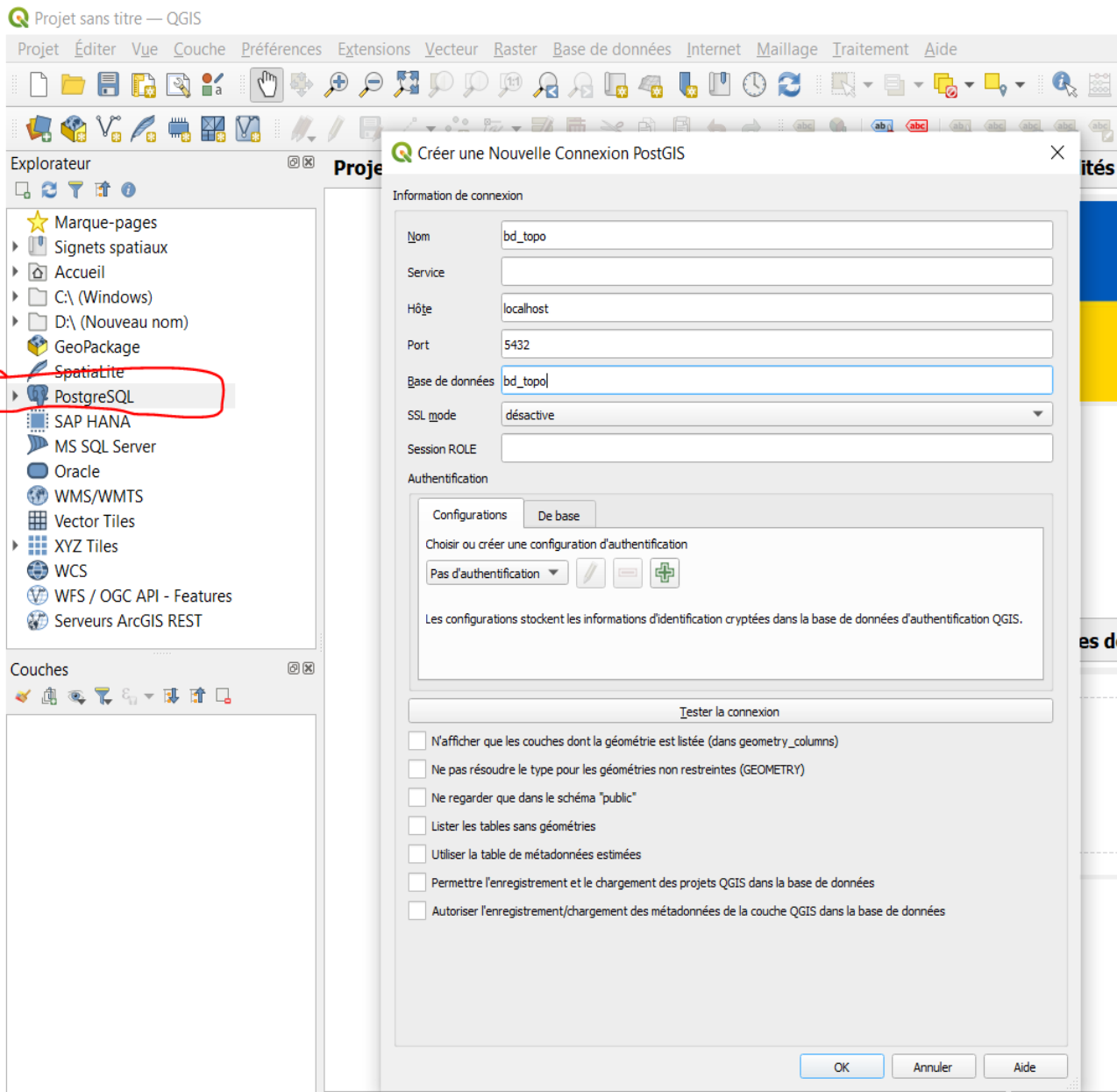
Préalablement, nous spécifions certains paramètres pour pouvoir interagir avec la partie adéquate de notre base de données :

- L'option `-h` permet de spécifier l'hôte ; dans notre cas, puisque tout se passe sur la même machine, nous utilisons "localhost".
- L'option `-p` indique le port 5432, qui est le port d'écoute utilisé par la base de données pour répondre à nos différentes requêtes et interagir avec les différents utilisateurs.
- L'option `-u` est utilisée pour spécifier l'utilisateur, qui par défaut est "postgres" dès l'installation de PostgreSQL.
- L'option `-d` permet de spécifier la base de données avec laquelle nous souhaitons interagir.
- L'option `-c` qui permet prend en charge la commande que nous souhaitons exécuter.

Pour ajouter les données à PostgreSQL, nous utilisons le même outil de manière presque identique. Cependant, une différence majeure réside dans le fait que contrairement à la procédure précédente, nous ne cherchons pas à exécuter une commande, mais plutôt à ajouter un fichier à la base de données. Ainsi, au lieu d'utiliser l'option `-c`, nous adoptons l'option `-f`, qui nous permet de spécifier le chemin d'accès au fichier contenant les données relatives aux zones de végétation que nous souhaitons incorporer. En faisant appel à l'option `-f`, nous sommes en mesure d'indiquer précisément le chemin d'accès complet menant au fichier renfermant les informations relatives aux zones de végétation. Cela nous permettra d'importer ces données à partir du fichier spécifié et de les intégrer de manière adéquate dans notre base de données PostgreSQL.

```
C:\Program Files\PostgreSQL\15\bin>psql -h localhost -p 5432 -U postgres -d bd_topo -f D:\bd_topo_sql\Bd_topo\BDTOPO\1_DONNEES_LIVRAISON_2023-03-08\BDT_3-3_SQL_WGS84G_FRA-ED2  
023-03-15\OCCUPATION_DU_SOL\zone_de_vegetation.sql
```

Une fois que les données ont été intégrées à notre base de données PostgreSQL, nous procédons à la connexion de cette dernière à notre logiciel de Système d'Information Géographique (SIG), QGIS. Pour cela, après avoir ouvert QGIS, nous repérons dans la barre latérale supérieure gauche l'icône du driver PostgreSQL. En effectuant un clic droit sur cette icône, nous sélectionnons l'option "Nouvelle connexion". Ensuite, nous remplissons les informations requises pour établir la connexion avec la base de données, comme illustré dans l'image ci-dessous. Grâce à cette configuration, nous sommes alors en mesure d'accéder aux données stockées dans la base de données PostgreSQL à travers QGIS.



**FIGURE 1 : CONNEXION D'UNE BASE DE DONNEES POSTGRESQL A QGIS**

Grâce à cette connexion établie entre notre logiciel SIG QGIS et la base de données PostgreSQL, nous avons la possibilité de télécharger et exporter les données au format geopackage. Une fois que nous avons exporté les données au format à partir de QGIS, nous pouvons les importer dans notre logiciel RStudio. Cela nous permet de travailler avec ces données géospatiales dans RStudio pour effectuer nos diverses analyses, traitements et visualisations.

Nous avons utilisé divers formats de données compatibles avec notre environnement logiciel RStudio. Parmi ces formats, nous avons traité des données telles que les géopackages, les shapefiles, les geojsons et les rasters, qui ont été gérées grâce aux packages spécialisés dans la manipulation de données spatiales. Ces packages fournissent toutes les fonctionnalités nécessaires pour exploiter directement et efficacement les autres types de données géospatiales auquel nous avons eu recours.

Nous avons travaillé avec des formats tabulaires tels qu'Excel ou le format CSV, qui sont couramment utilisés pour stocker des données structurées. RStudio dispose nativement de fonctions dédiées qui permettent d'importer, lire et manipuler ces formats tabulaires de manière flexible.

### **3. Densité de végétation autour des écoles**

Pour parvenir à une appréhension plus approfondie de l'influence de la végétation sur les performances scolaires, il est essentiel de mieux connaître la répartition de la végétation aux abords des écoles. Pour atteindre cet objectif, nous devons générer des données spécifiques en procédant à une étape cruciale de création de données. Cette étape consistera à évaluer avec précision la superficie des espaces végétalisés situés à proximité des établissements scolaires dans l'ensemble du territoire français. Ce calcul de superficie, nous permettra d'obtenir une vision globale des zones végétales entourant les écoles, ce qui nous permettra de mieux comprendre l'impact potentiel de l'environnement végétal sur les performances des élèves. Nous serons dès lors en mesure d'approfondir notre analyse et de dégager des tendances significatives, contribuant ainsi à une meilleure compréhension des liens entre la végétation et les résultats éducatifs. Cette démarche revêt une importance capitale dans le cadre de nos investigations sur les facteurs environnementaux susceptibles d'influencer le système éducatif en France.

Comme indiqué précédemment, la majeure partie de notre travail a été réalisée sous R. Grâce à un code spécialement conçu et progressivement amélioré, nous avons automatisé le calcul des superficies de zones végétalisées à l'intérieur de zones tampons de rayons variables. Les distances choisies sont de 100, 200, 500 et 1000 mètres autour des écoles.

Concernant les données des écoles, nous avons utilisé celles fournies par le ministère de l'Enseignement, qui représentent les établissements sous forme de points géolocalisés dans le système de coordonnées Lambert 93.

Quant aux données de végétation, nous avons utilisé la base de données BD topo de l'IGN, que nous avons jugée plus complète et appropriée pour atteindre nos objectifs. Dans cette base, les zones de végétation sont représentées par des polygones et le système de coordonnées de base est le WGS84. Nous avons donc dû reprojecter ces données en Lambert 93 afin de les harmoniser avec le système de projection des données des écoles.

Pour mener à bien cette tâche, nous nous sommes inspirés de la fonction développée par le professeur Thierry Feuillet. Nous avons adapté cette fonction pour répondre spécifiquement à nos besoins, en effectuant les calculs de superficie souhaités pour les différentes distances préalablement fournies.

Dans la suite de notre démarche, nous allons fournir une description complète du code utilisé, en expliquant les différentes étapes qui nous ont permis de parvenir à une table de données contenant les valeurs de végétation pour chaque école, et ce, pour les différentes distances spécifiées. Cette table constituera une ressource essentielle pour notre analyse ultérieure des liens entre la végétation environnante des écoles et les performances scolaires des élèves.

- **Fonction de calcul des superficies**

La fonction que nous avons nommée ``dens_feature`` réalise un traitement spatial visant à calculer la densité des attributs d'une couche d'observations (``obs``) par rapport à une autre couche géométrique (``feat``). Cette fonction présente une approche générique qui permet son application à d'autres mesures de type polygone autour d'une couche de points, en utilisant des distances préalablement spécifiées. En utilisant ``dens_feature``, nous pouvons évaluer la densité des attributs de la couche ``feat`` à différentes distances autour de chaque point de la couche ``obs``. L'aspect générique de cette fonction offre une grande flexibilité, car nous pourrions l'appliquer à diverses métriques et mesures géographiques pour analyser la répartition spatiale des attributs autour des points d'intérêt.

- **La fonction prend quatre arguments en entrée :**

``obs`` : La variable ``obs`` désigne la couche géométrique des établissements scolaires, une couche de points servant de référence pour évaluer les métriques liées à la végétation environnante. Dans notre cas, il s'agit de la couche des établissements.

``feat`` : Une couche d'observations géométriques, représentée sous forme de polygones, contenant les données que nous souhaitons évaluer. Dans notre étude, nous nous intéressons particulièrement à l'étendue de la végétation environnante des établissements scolaires, et c'est cette couche de polygones qui nous permettra de quantifier cette ampleur.

``span`` : Un vecteur contenant les distances à utiliser pour créer des tampons (buffers) autour des observations. Il s'agit d'un paramètre qui définit les distances à tester pour calculer la densité.

``id`` : Une clé d'identification des observations. Il s'agit du nom de l'attribut qui servira de clé pour joindre les résultats agrégés avec les observations.

- La fonction commence par transformer les coordonnées des observations (``obs``) dans le système de coordonnées de la couche ``feat`` à l'aide de la fonction ``st_transform``.
- Ensuite, elle extrait l'attribut spécifié (``id``) de la couche ``obs`` et le stocke dans un objet de données (``obsId``) en vue d'une utilisation ultérieure pour les agrégations.
- Une liste vide (``result_list``) est créée pour stocker les résultats finaux pour différentes distances (``span``).
- La fonction utilise ensuite une boucle ``for`` pour parcourir toutes les distances spécifiées dans ``span``.
- À l'intérieur de la boucle, un tampon (``buff``) est créé autour des observations à la distance ``d`` actuelle.
- Ensuite, l'intersection entre la couche ``feat`` et le tampon est calculée, et l'aire des zones intersectées est également calculée.
- Les valeurs de l'attribut ``id`` sont agrégées par somme des aires pour chaque distance ``d``, et les résultats agrégés sont stockés dans ``tmpAgr``.
- Les résultats agrégés sont ensuite joints avec les observations (``obsId``) en utilisant l'attribut ``id``.
- Les valeurs manquantes dans les résultats agrégés sont remplacées par zéro.



- L'objet de sortie est nommé avec le préfixe "dens\_" et la distance `d`, puis il est ajouté à la liste `result\_list`.
- La boucle continue pour chaque distance `d` dans `span`.
- Enfin, la fonction retourne le dernier résultat de la liste, c'est-à-dire les résultats agrégés pour la distance ayant la valeur la plus élevée dans `span`.

La fonction calcule ainsi la densité des attributs de la couche `obs` par rapport à la couche `feat` pour différentes distances spécifiées dans `span`. Les résultats sont renvoyés sous forme de tableau avec l'attribut `id` et les colonnes représentant la densité pour chaque distance `d`.

### • Calcul des densités de végétation

Des suites de ces explications on comprend d'emblée que pour utiliser cette fonction, il est nécessaire de préparer certains paramètres au préalable. Parmi eux, nous devons créer un vecteur contenant les différentes distances (100 mètres, 200 mètres, 500 mètres et 1000 mètres) que nous souhaitons utiliser pour évaluer la végétation autour des établissements scolaires.

Un autre paramètre important est l'identifiant, dont il est essentiel de connaître la syntaxe exacte de la colonne correspondante dans les données de géolocalisation des établissements. Cette colonne doit contenir des identifiants uniques pour chaque école, permettant d'associer correctement les informations de géolocalisation à chaque point d'établissement. La colonne 'numero\_uai' qui renseigne les numéros d'identification unique de chaque établissement répond parfaitement à cette condition.

Les données de géolocalisation des établissements scolaires et des zones de végétation revêtent une importance capitale pour notre étude. Cette importance réside non seulement dans les informations qu'elles nous apportent, mais également dans leur ampleur et leur poids considérables. En particulier, les données de végétation peuvent être de grande envergure à l'échelle de toute la France, ce qui complexifie notablement les opérations de croisement entre ces deux vastes ensembles de données.

Le traitement de ces données volumineuses peut engendrer des problèmes de performance, tels que des temps de calcul excessivement longs et des risques de dysfonctionnements, même sur des machines disposant de capacités substantielles. Ces contraintes peuvent entraîner des ralentissements dans le processus de calcul et provoquer des problèmes de gestion de la mémoire vive, notamment dans RStudio et même d'autres outils de Systèmes d'Information Géographique (SIG).

Face à cette contrainte, nous avons dû trouver des solutions. Après exploration de nombreuse piste nous avons fait le choix de morceler les calculs. En effet cela consiste à diviser en plusieurs parties les calculs et les effectuer à tour de rôle pour ensuite les rassembler dans une même table.

Dans cette perspective, la décision a été prise de réaliser les calculs individuellement pour chaque établissement. Pour ce faire, nous avons mis en place une boucle qui parcourt séquentiellement chaque établissement, de 1 jusqu'au nombre total d'établissements. À chaque itération, elle sélectionne l'établissement en cours en filtrant la table d'établissements ligne par ligne. Ensuite, elle importe les données de la végétation située dans un rayon de 1500 mètres autour de cet établissement. Par la suite, elle applique la fonction 'dens\_feature' à ces données, puis ajoute les résultats à une table de compilation préalablement préparée.

#### 4. Calcul des moyennes de températures et de précipitation.

Étant donné que nous souhaitons approfondir notre connaissance de l'impact de l'environnement naturel sur les performances des élèves nous avons effectué le choix d'inclure des données climatiques dans notre étude. Se limiter uniquement aux éléments de végétation pourrait être assez réducteur. C'est pourquoi nous avons pris la décision de nous intéresser également aux données de températures et de précipitations.

L'objectif est d'obtenir une valeur moyenne de température et de précipitation autour des écoles afin d'avoir une idée de l'état du climat environnant ces établissements scolaires. En incluant ces données climatiques, nous pourrions prendre en compte un aspect supplémentaire de l'environnement qui pourrait potentiellement influencer les performances scolaires des élèves.

En rassemblant les informations sur la végétation, la température et les précipitations, nous pourrions disposer d'un ensemble plus complet de variables environnementales pour approfondir notre analyse et mieux comprendre l'interaction entre la nature et les résultats éducatifs.

Les données climatiques historiques fournies par la base de données WorldClim sont une ressource précieuse pour notre étude, car elles couvrent une période étendue et fournissent des informations sur les métriques climatiques mondiales à haute résolution. Nous avons accès à des données de climat moyen pour chaque mois de l'année, ce qui nous permet d'obtenir une vision détaillée des variations saisonnières du climat dans chaque région de la France.

La résolution la plus fine disponible, à savoir 1 kilomètre, nous offre une granularité élevée dans les données climatiques, ce qui signifie que nous disposons d'une valeur de température ou de précipitation distincte pour chaque carré de 1 kilomètre sur l'ensemble du territoire français. En utilisant cette résolution, nous avons pu extraire les valeurs de température correspondant à la localisation géographique exacte de chaque établissement scolaire dans notre base de données. Ainsi, pour chaque école, nous avons pu obtenir une mesure précise de la moyenne dans son voisinage immédiat.

Pour réaliser cette opération, nous avons fait usage de la fonction "extract" du package "terra", qui est spécialement conçu pour manipuler les données raster. Grâce à cette fonction, nous avons pu extraire les valeurs pour chaque école en fonction de sa position géographique, ce qui nous permettra d'intégrer cette dimension climatique dans notre analyse.

##### 1. La fonction extract

La fonction `extract()` est une fonction clé du package R "terra" qui permet d'extraire les valeurs des cellules d'une grille raster pour un ensemble de points, lignes ou polygones. Elle offre une flexibilité élevée en termes de types de géométries en entrée et de types de cellules de la grille raster. Cela signifie que vous pouvez obtenir les valeurs des cellules de la grille raster à des emplacements spécifiques définis par les géométries fournies. La fonction renvoie un vecteur de valeurs correspondant à chaque géométrie d'entrée. Une colonne d'identification de la position de chaque élément de la table d'emplacement est fournie pour permettre de repérer chaque établissement. Cette fonction prend en entrée deux paramètres principaux. Le paramètre X est objet de type 'SpatRaster' ou 'SpatExtent' représentant la grille raster à partir de laquelle extraire les valeurs des cellules. Le paramètre Y quant à lui est un objet de type "SpatVector" (géométrie de points, lignes ou polygones) représentant les emplacements pour lesquels vous souhaitez extraire les valeurs des cellules de la grille raster. La possibilité est donnée de rajouter des paramètres supplémentaires et facultatifs pour spécifier des

options spécifiques comme la méthode d'interpolation ou la gestion de valeurs manquantes ou autres besoins spécifiques.

Lorsque nous utilisons cette fonction pour déterminer une valeur de température ou de précipitation représentative de la réalité climatique environnante des établissements, nous commençons par identifier la géolocalisation de chaque établissement. En utilisant ces coordonnées, nous déterminons le pixel correspondant dans les données climatiques. Cependant, étant donné que le rayon d'un kilomètre autour de chaque établissement peut englober plusieurs pixels, il est nécessaire de prendre en compte une zone plus large pour obtenir une mesure représentative de la réalité climatique. Pour ce faire, nous effectuons un travail préliminaire. Nous calculons une valeur moyenne par pas de 1 kilomètre (1 pixel) autour de chaque pixel, en prenant en compte toutes les valeurs de température ou de précipitation disponibles dans cette zone. En procédant ainsi, nous recréons de nouveaux pixels dont la valeur représente la moyenne des températures ou des précipitations à un kilomètre autour de chacun d'eux. Cela nous permet d'obtenir une estimation plus précise de la valeur climatique dans la zone environnante de chaque établissement. En fin de compte, pour tous les établissements, la valeur de température ou de précipitation obtenue sera au minimum une moyenne basée sur les données des pixels situés dans un rayon d'un kilomètre autour de chaque établissement. Cela nous offre une mesure plus fiable et représentative de la réalité climatique locale pour chaque établissement. Cette approche est essentielle pour fournir des données climatiques pertinentes et adaptées à chaque emplacement spécifique.

## 2. Description du code

Dans le cadre de notre travail, nous avons développé un code qui exécute successivement les différentes étapes de la méthodologie décrite précédemment. Ce code permet de réaliser les opérations suivantes :

- **Importation des rasters de température ou de précipitation pour chaque mois de l'année :**

Le code utilise une boucle ``for`` pour importer douze rasters, un pour chaque mois de l'année. Les noms des fichiers sont générés dynamiquement à l'aide de la fonction ``sprintf`` en utilisant les numéros de mois (``i``). Chaque raster, pour le cas des températures par exemple, est attribué à une variable avec un nom correspondant (``tavg_1`` pour janvier, ``tavg_2`` pour février, etc.). Les rasters sont chargés à l'aide de la fonction ``rast()``.

- **Empilement des rasters dans un vecteur :**

Tous les rasters de température sont ensuite empilés dans un même vecteur appelé ``tavg``. Cela se fait en utilisant l'opérateur de concaténation ``c()`` pour rassembler les rasters ``tavg_1`` à ``tavg_12`` dans un seul vecteur.

- **Calcul de la moyenne des températures et intersection avec l'emprise de la France :**

La moyenne des températures est calculée à partir du vecteur ``tavg`` à l'aide de la fonction ``terra:mean()``. Cette opération permet d'obtenir une moyenne annuelle de température ou de précipitation. Un seul raster résulte donc de cette opération.

Ensuite, la fonction ``crop()`` est utilisée pour découper le raster résultant en utilisant les limites de l'emprise de la France définies par l'objet ``fr``. Enfin, la projection du raster est modifiée en utilisant ``project("epsg:2154")``.

- **Définition du type de la matrice pour le calcul de la moyenne dans un rayon donné :**

Le code définit le type de matrice (``typ``) qui sera utilisé pour le calcul de la moyenne dans un rayon donné. Dans ce cas, le type de matrice est défini comme "circle".

- **Création d'un nouveau raster de dimensions 1000m :**

Un nouveau raster appelé ``rst_tavg`` est créé avec des dimensions de 1000 mètres. Il est basé sur l'objet ``tavg``, en utilisant la même référence spatiale (système de coordonnées) que le raster "tavg". Cela permettra d'utiliser ce raster pour calculer les moyennes dans un rayon spécifié autour de chaque cellule.

- **Calcul de la matrice focale (focalMat) pour les températures moyennes :**

Une matrice focale (``mat``) est calculée à partir du raster de température moyenne ``tavg``. Cette matrice sera utilisée pour spécifier le voisinage dans lequel le calcul de la moyenne sera effectué pour chaque cellule.

- **Calcul des moyennes dans un rayon de 1000m autour de chaque cellule :**

Le code utilise la fonction ``focal()`` pour calculer les moyennes des températures (``dens_tavg``) dans un rayon de 1000 mètres autour de chaque cellule du raster ``tavg``. La matrice focale ``mat`` est utilisée pour définir le voisinage dans lequel le calcul est effectué.

- **Extraction des valeurs de température moyenne pour les emplacements des écoles :**

Enfin, les valeurs de température moyenne sont extraites pour les emplacements des écoles, qui sont définis par l'objet ``etab``. Les valeurs sont stockées dans le data frame ``tavg_dens``, qui contient les informations des emplacements des écoles ainsi que les valeurs de température moyenne correspondantes.

De l'application de toute ces étapes, résulte une table contenant les valeurs de températures ou de précipitation liés aux établissements par leur numéro d'identification unique. Ceci va nous servir donc pour mieux appréhender le climat au tour des écoles.

Nous avons entrepris de comprendre l'impact de l'environnement naturel sur les performances des élèves des établissements scolaires en France. Pour ce faire, nous avons réalisé des calculs de densités de végétation et des moyennes de températures et de précipitations autour des écoles. Ces différentes mesures environnementales ont été intégrées dans notre analyse pour fournir un aperçu plus complet des facteurs qui pourraient influencer les résultats éducatifs.

Afin d'atteindre notre objectif, il est indispensable d'obtenir une géolocalisation exhaustive de tous les établissements scolaires présents sur l'ensemble du territoire français. Heureusement, le ministère de l'Enseignement a grandement facilité notre tâche en mettant à disposition en libre téléchargement les données de géolocalisation des établissements scolaires en France. Le numéro d'identification unique des établissements fourni dans cette base de données nous a été d'une grande utilité pour nous repérer dans l'ampleur conséquente des calculs que nous avons dû effectuer. Grâce à ces informations précieuses, nous avons pu avancer efficacement dans notre démarche et obtenir des résultats pertinents pour notre travail.

Les calculs de densités de végétation ont été effectués en préparant préalablement des paramètres essentiels, tels que la création d'un vecteur contenant les distances à évaluer (100 mètres, 200 mètres, 500 mètres et 1000 mètres) et l'identification des écoles à partir de leur colonne 'numero\_uai'. Cette approche nous a permis de quantifier les surfaces végétales dans les zones environnantes de chaque

établissement scolaire. En divisant les calculs par département, nous avons pu gérer efficacement le traitement de données volumineuses, évitant ainsi les problèmes de performance.

Outre la densité de végétation, nous avons également pris en compte les données climatiques historiques fournies par la base de données WorldClim. Ces données nous ont permis d'obtenir les moyennes de températures et de précipitations autour des écoles. L'utilisation de la fonction `extract()` du package "terra" a été essentielle pour extraire les valeurs de température et de précipitation correspondant aux emplacements géographiques exacts de chaque établissement. En agrégeant ces données, nous avons pu obtenir une mesure plus fiable de la réalité climatique environnante de chaque établissement.

En rassemblant les résultats des calculs de densité de végétation et des moyennes climatiques, nous avons constitué un ensemble complet de variables environnementales pour notre analyse. Cela nous permettra d'approfondir notre compréhension des relations entre l'environnement naturel et les performances éducatives.

Cette étude revêt une importance particulière car elle nous permet d'explorer des aspects de l'environnement qui n'ont peut-être pas été pleinement pris en compte dans les analyses précédentes sur l'éducation. En intégrant les facteurs environnementaux, tels que la végétation et le climat, nous pouvons mieux appréhender l'influence de l'environnement sur les performances des élèves. Cependant, il est important de noter que cette étude présente certaines limitations. Tout d'abord, les données géospatiales et climatiques sont susceptibles de contenir des incertitudes et des biais inhérents aux méthodes de collecte et de traitement. De plus, d'autres facteurs socio-économiques et culturels pourraient également jouer un rôle significatif dans les performances éducatives, mais n'ont pas été pris en compte dans cette analyse. En dépit de ces limitations, cette étude constitue une première étape importante pour mieux comprendre les interactions entre l'environnement naturel et l'éducation. Elle ouvre la voie à de futures recherches approfondies dans ce domaine et à l'exploration de nouvelles pistes pour améliorer la qualité de l'éducation et le bien-être des élèves.

En conclusion, l'intégration des données de densité de végétation et des moyennes climatiques dans notre analyse nous a permis de construire les informations qui nous permettront d'explorer l'influence de l'environnement naturel sur les performances éducatives en France. Cette approche multidimensionnelle contribue à une meilleure compréhension des facteurs environnementaux qui peuvent influencer l'apprentissage des élèves. Nous espérons que les résultats de ce travail inciteront à davantage de recherches dans ce domaine et contribueront à la mise en place de politiques éducatives plus éclairées et durables.

## IV. ANALYSES ET RESULTATS

Après notre phase de calcul de densités de végétation et de moyennes de températures et de précipitations autour des établissements scolaires en France, nous entamons désormais la phase cruciale d'analyse des données obtenues. Cette étape permettra de tirer des enseignements pertinents et de mieux comprendre l'interaction entre l'environnement naturel et les performances éducatives.

Au cours de cette analyse, nous examinerons les résultats obtenus à partir des mesures de densités de végétation pour différentes distances autour des écoles. Nous chercherons à identifier les tendances et les variations géographiques de la densité de végétation, afin de déterminer si cette variable exerce une influence significative sur les performances scolaires.

De même, nous étudierons les données climatiques relatives aux températures et aux précipitations moyennes autour des écoles. Nous chercherons à comprendre comment les variations climatiques locales peuvent influencer les résultats éducatifs et si certaines régions présentent des caractéristiques climatiques favorables ou défavorables pour l'apprentissage.

En combinant ces données environnementales avec les informations sur les performances scolaires des élèves, nous tenterons d'établir des corrélations potentielles entre l'environnement naturel et les résultats éducatifs. Cette approche nous permettra de découvrir des associations significatives qui pourraient être utiles pour informer les politiques éducatives et les pratiques pédagogiques.

D'autres données fournies par les organismes gouvernementaux comme la catégorisation des communes ou encore les indices de position sociale seront pris en compte dans nos diverses analyses. Cela nous permettra de mettre en évidence d'autres variables pouvant influencer les performances scolaires.

Par ailleurs, cette phase d'analyse impliquera également l'utilisation d'outils statistiques et de méthodes d'analyse des données. Nous réaliserons des tests statistiques pour évaluer la force des associations entre les variables environnementales et les performances scolaires. Les facteurs socio-économiques et culturels, bien qu'ils n'aient pas été inclus dans notre modèle initial, pourraient jouer un rôle significatif dans les performances éducatives et seront dans la mesure du nécessaire pris en compte avec prudence.

Cette phase d'analyse des données sera cruciale pour donner du sens à nos résultats et pour éclairer les implications de notre étude sur la relation entre l'environnement naturel et l'éducation.

## 1. Méthodologie

Avant de nous attaquer aux résultats scolaires, nous allons nous intéresser à certaines caractéristiques propres aux écoles elles-mêmes. Notre premier objectif sera de vérifier si certaines des caractéristiques structurelles liées aux établissements scolaires exercent une influence sur le taux de végétalisation de leur environnement proche. Pour ce faire, nous allons examiner attentivement des facteurs tels que la dénomination, la localisation (urbaine, rurale, etc.) et d'autres variables pertinentes. Cette analyse préliminaire nous permettra de mieux comprendre les relations potentielles entre les caractéristiques des écoles et leur environnement végétalisé, ce qui pourrait être utile pour interpréter les résultats scolaires par la suite.

Pour chacune des variables que nous allons étudier, nous commencerons par effectuer une analyse descriptive basique. Cette étape nous permettra d'obtenir une vue rapide des moyennes et des écarts-types en fonction des catégories d'établissements que nous considérons. En analysant ces statistiques descriptives, nous pourrons observer les premières différences potentielles entre les différentes catégories d'établissements en ce qui concerne la variable étudiée.

Cependant, pour valider ces différences observées, nous devons passer à une étape de vérification de la fiabilité de notre comparaison. Pour ce faire, nous aurons besoin d'effectuer un test statistique approprié. Ce test nous permettra de déterminer si les différences que nous avons constatées entre les différents établissements ne sont pas simplement dues au hasard, mais sont bien influencées par le facteur que nous prenons en compte (par exemple, le secteur d'activité public ou privé).

L'objectif de cette étape est de s'assurer que nos conclusions sont statistiquement significatives et robustes. En utilisant un test statistique adéquat, nous pourrons évaluer de manière objective si les différences observées sont réellement liées au facteur que nous étudions ou si elles pourraient être dues à des variations aléatoires dans les données.

## 2. Test statistique

Une fois nos diverses comparaisons effectuées, nous devons d'abord vérifier s'il existe une différence significative entre la végétation et la variable prise en compte. Pour cela, nous allons réaliser un test statistique pour évaluer si les moyennes de végétalisation des deux groupes sont différentes.

L'hypothèse de départ est que les établissements quel que soit leur catégorie ont des taux de végétalisation équivalents. Cependant, nous souhaitons établir si cette hypothèse est valide ou si nous devons la rejeter au profit de l'idée qu'il y a une réelle différence entre les deux.

Pour garantir la validité de notre comparaison, nous choisissons un niveau de signification de 5 %. Ce niveau de signification ( $\alpha = 0,05$ ) représente le seuil au-delà duquel nous rejetterons l'hypothèse nulle (égalité des moyennes) si la probabilité d'erreur est trop élevée.

En effectuant le test statistique avec ce niveau de signification, nous nous assurons que les chances de conclure à une différence significative lorsque celle-ci n'existe pas réellement sont limitées à 5 %, ce qui rend notre analyse plus fiable.

Pour effectuer notre test, nous faisons le choix d'utiliser le t-test. Le t-test est un test statistique utilisé pour comparer les moyennes de deux groupes indépendants. Il évalue s'il existe une différence significative entre ces moyennes en prenant en compte la taille des échantillons, les écarts-types et les moyennes. Le t-test calcule une statistique t, qui mesure l'écart normalisé entre les groupes. Ensuite, il génère une valeur p, qui indique la probabilité que la différence observée soit due au hasard. Si la

valeur p est inférieure à un seuil prédéfini (0,05 dans notre cas et en général), on considère qu'il y a une différence significative entre les moyennes des groupes. Nous pouvons effectuer ce test à l'aide de la fonction 't.test' dans R.

Dans toutes nos analyses, nous nous concentrons spécifiquement sur la végétation située dans un rayon de 1000 mètres autour des établissements scolaires. Cette fenêtre de 1000 mètres est considérée comme assez large pour appréhender de manière significative l'environnement immédiat des écoles.

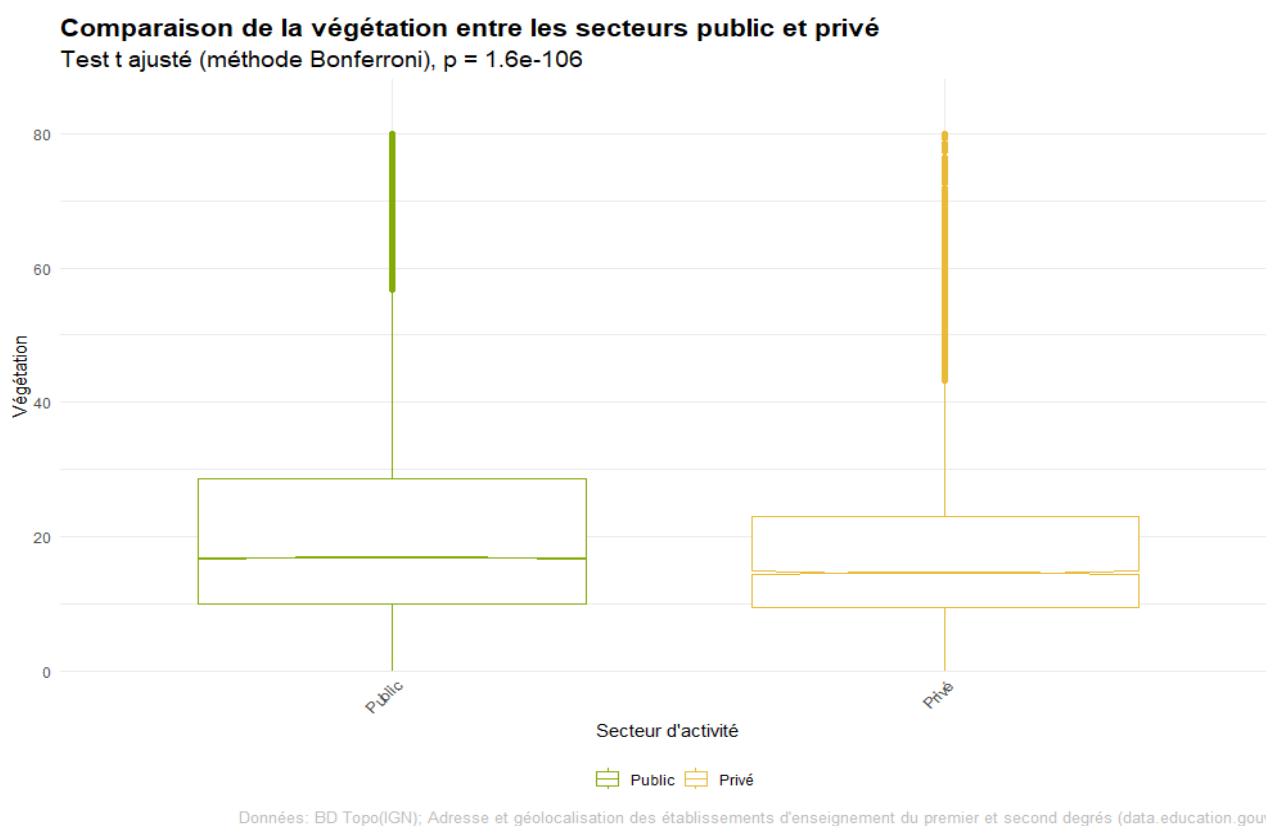
### 3. Caractéristiques des établissements et végétation

#### a) Secteur public/privé et végétation

Pour cette étude, nous allons nous intéresser à une première caractéristique des établissements scolaires : le secteur d'activité, c'est-à-dire s'ils sont publics ou privés. La base de données des établissements scolaires en France nous offre des informations détaillées sur cette distinction essentielle.

Notre objectif principal est de comparer la végétation environnante des établissements publics avec celle des établissements privés. Pour ce faire, nous allons diviser l'ensemble des établissements en deux catégories distinctes en fonction de leur secteur d'activité. Nous procéderons ensuite à une analyse approfondie en calculant les moyennes du taux de végétalisation pour chacune de ces catégories. Cela nous permettra de déterminer s'il existe une différence entre les établissements publics et privés en termes de végétation présente dans leur environnement proche.

Le résultat de cette analyse présente comme suit :



**FIGURE 2 : COMPARAISON ENTRE VEGETATION AUTOUR DES ETABLISSEMENTS PRIVES ET PUBLICS**

La végétation est ici exprimée en pourcentage.



D'après notre analyse, il ressort que dans le secteur public, les alentours des établissements sont globalement plus végétalisés que dans le domaine privé.

Dans le diagramme en boîte, l'écart interquartile entre les deux groupes est plus important du côté des établissements publics que du côté des établissements privés. Cela indique une plus grande dispersion des valeurs de végétation au sein du groupe des établissements publics.

La médiane, représentée par la ligne au milieu de la boîte, est supérieure dans le secteur public par rapport à celle obtenue dans le secteur privé. Cela signifie que la tendance centrale des valeurs de végétation est plus élevée dans le secteur public que dans le secteur privé. Les valeurs extrêmes, représentées par les points au-delà des moustaches, sont également plus élevées pour les écoles publiques que pour les écoles privées. En résumé, les écoles publiques semblent être, dans l'ensemble, plus végétalisées que les écoles privées dans leur environnement proche. Les résultats indiquent une plus grande présence de végétation dans les alentours des établissements du secteur public par rapport à ceux du secteur privé. La table suivante fait un résumé des valeurs importantes de cette analyse statistique.

	Privé	Public
<b>secteur_public_privé_libe</b>	Privé	Public
<b>Moyenne</b>	<b>18.35179</b>	21.54496
<b>premier_quartile</b>	<b>9.476198</b>	9.996006
<b>troisieme_quartile</b>	<b>22.98068</b>	28.67487
<b>ecart_interquartile</b>	<b>13.50448</b>	18.67886
<b>mediane</b>	<b>14.64972</b>	16.81031
<b>longueur_encoche</b>	<b>27.00896</b>	37.35773
<b>valeur_max</b>	<b>79.91369</b>	79.98026

**TABLEAU 1 : STATISTIQUES UNIVARIEES DE LA VEGETATION EN FONCTION DU SECTEUR D'ACTIVITE**

Il est, à présent, essentiel de vérifier si la différence nette observée entre les deux groupes n'est pas simplement due au hasard. Pour cela, nous allons utiliser le test t, également appelé t-test, qui nous permettra de déterminer la valeur p associée à notre analyse comparative.

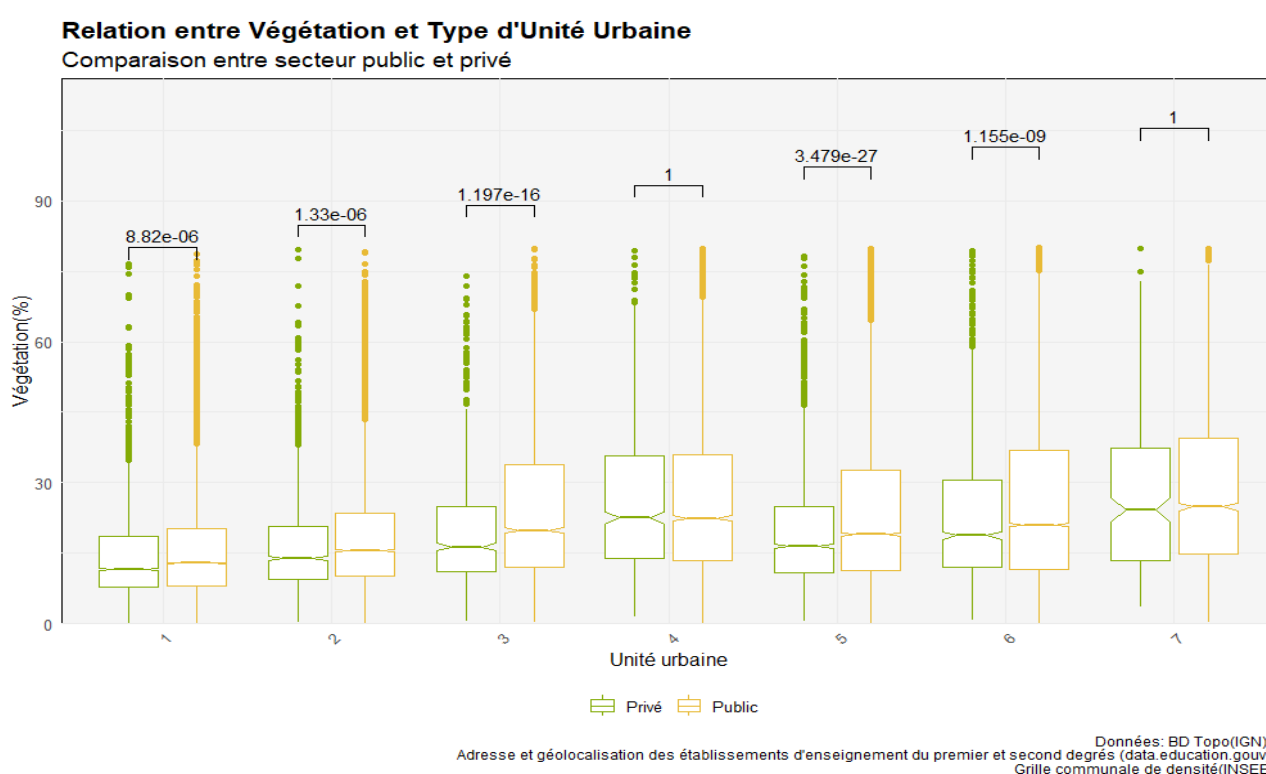
En effectuant ce test sur notre jeu de données, la p-value que nous obtenons est inférieure à  $1,6e-106$ . Une p-value inférieure à 0,05, qui indique une extrême significativité statistique dans le résultat de notre test. Cela suggère qu'il y a une différence très significative entre les moyennes ou distributions de la variable végétation pour les différentes catégories de secteur d'activité (public ou privé). On rejette alors l'hypothèse nulle ( $H_0$ ) selon laquelle il n'y a pas de différence significative entre les groupes. Les différences que nous observerons n'étant donc pas dues au hasard, on peut conclure que la végétation varie en effet en fonction du secteur d'activité.

## b) Relation entre végétation et type urbain des communes

Il est possible que la différence de végétalisation entre les établissements publics et privés soit influencée par le contexte rural ou urbain des communes où ils se trouvent. Selon l'Insee, une commune urbaine est caractérisée par un bâti continu (pas de coupure de plus de 200 mètres entre deux constructions) avec au moins 2 000 habitants, tandis qu'une commune rurale aura moins de constructions présentes.

Il est envisageable que les établissements privés préfèrent s'implanter dans les communes urbaines. Les établissements publics seraient par conséquent plus présents dans les zones rurales, qui offrent généralement plus d'espaces verts.

Cependant, ces suppositions doivent être confirmées par des tests statistiques appropriés. Nous devons mener des analyses supplémentaires pour contrôler l'influence du contexte rural ou urbain sur la végétalisation des établissements.



**FIGURE 3 : COMPARAISON DE LA VEGETATION AUTOUR DES ETABLISSEMENTS PRIVES ET PUBLICS EN FONCTION DES TYPES D'UNITES URBAINES**

Le graphique que nous avons, offre une représentation de la végétation en fonction de la classification des communes. Cette classification des communes est établie en prenant en compte le nombre d'habitants et la répartition de ces habitants sur le territoire. Elle tient également compte de la dispersion des bâtiments sur le territoire communal.

En utilisant cette classification, nous pouvons distinguer les communes urbaines des communes rurales. Les communes urbaines sont caractérisées par une densité de population élevée, un bâti continu avec peu de coupures entre les constructions et un nombre minimum d'habitants (au moins 2 000 habitants selon l'Insee). En revanche, les communes rurales présentent une densité de population plus faible, un bâti moins continu avec des espaces plus étendus et moins de constructions.

En effet, l'observation du graphique met en évidence une tendance progressive d'augmentation du taux de végétation autour des établissements scolaires à mesure que l'on se dirige vers le milieu rural. Cette évolution peut être expliquée par une densité de bâtiments moins importante dans les zones rurales par rapport aux zones urbaines. En milieu rural, il y a plus d'espaces non bâtis, ce qui favorise la présence de végétation.

Par ailleurs, on remarque également que la longueur de la moustache est d'autant plus grande pour les établissements situés dans le milieu rural. Cela signifie que les valeurs extrêmes, c'est-à-dire les zones avec le taux de végétation le plus élevé, sont principalement concentrées dans les communes rurales et autour des écoles publiques. Ces constats renforcent l'idée que la végétalisation est plus marquée dans les zones rurales, où la présence de bâtiments est moins dense, permettant ainsi à la végétation de prospérer davantage.

Les intervalles interquartiles, qui représentent la dispersion des données, sont toujours plus importants du côté des établissements publics que du côté des établissements privés. Cela indique que les zones avec le taux de végétation le plus élevé, sont plus fréquentes et plus élevées pour les établissements publics que pour les établissements privés.

Dans la majorité des cas, les p-values sont inférieures à 0.05, ce qui signifie que le risque d'erreur est limité à 5 % ou moins. Cela nous permet de conclure de manière significative que la différence de végétalisation entre les établissements publics et privés est probablement réelle et ne résulte pas simplement du hasard. Cependant, une exception se présente dans le cas des ceintures urbaines, où la p-value est égale à 1. Cela signifie que dans ce contexte géographique, la répartition de la végétation est purement hasardeuse, et il n'y a pas de différence significative entre les établissements publics et privés en termes de végétalisation. Cette observation est cohérente avec la position des ceintures urbaines dans la hiérarchie des communes, étant donné qu'elles se situent à cheval entre différentes classes de communes (urbaines et rurales). La végétalisation dans les ceintures urbaines est donc plus hétérogène et dépend davantage de facteurs autres que le statut public ou privé des écoles.

Ces observations renforcent l'idée que les établissements publics bénéficient en moyenne d'une végétalisation plus importante dans leur environnement proche par rapport aux établissements privés. La présence plus fréquente et plus marquée de végétation autour des établissements publics pourrait être liée à leur localisation prédominante dans des zones rurales, où la végétation naturelle est plus répandue et où la densité de bâtiments est moins importante.

Pour renforcer cette idée, une dernière étape consiste à comparer les effectifs des établissements publics et privés dans les différents contextes géographiques. Cette démarche vise à vérifier si les établissements publics sont effectivement plus nombreux que les écoles privées, et si les établissements privés ont tendance à privilégier les espaces plus urbains plutôt que les communes rurales. En réalisant cette comparaison, nous pourrions mieux comprendre la répartition des établissements publics et privés selon leur localisation géographique. Si nous constatons une prédominance des établissements publics dans les zones rurales, cela renforcerait l'idée que les établissements publics sont plus fréquemment végétalisés en raison de leur prédominance dans les espaces ruraux. D'autre part, si les établissements privés sont principalement concentrés dans les zones urbaines, cela pourrait expliquer en partie pourquoi la végétation autour des établissements privés est généralement moins importante, car les zones urbaines ont tendance à être plus densément construites et donc moins végétalisées.

## Portion des établissements public et privés en fonction des types d'unités urbaines

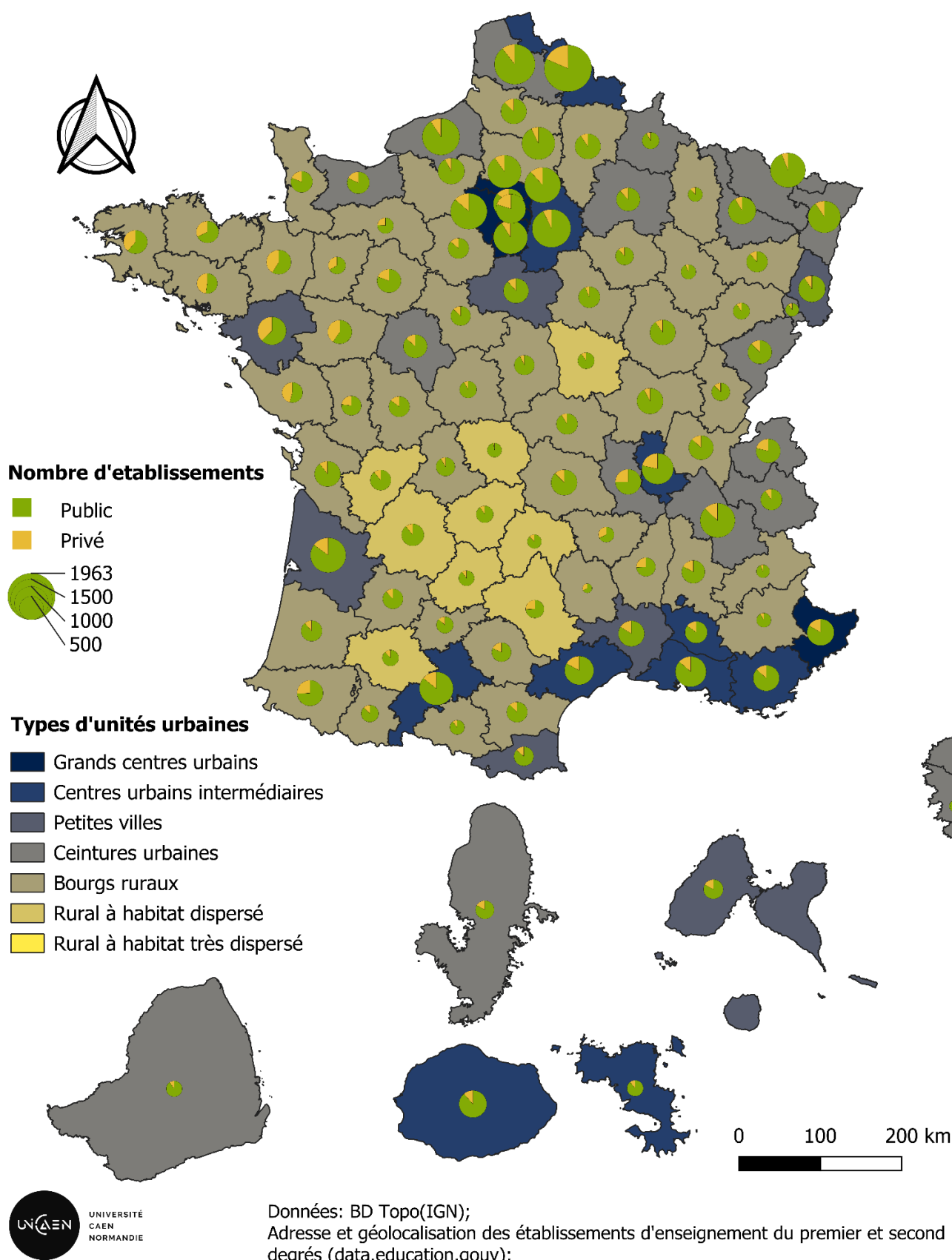


FIGURE 4 : PORTION D'ÉTABLISSEMENTS PUBLICS ET PRIVÉS EN FONCTION DES UNITÉS URBAINES

La carte présentée ci-dessus illustre la proportion d'établissements publics et privés dans chaque département de France. Chaque département est représenté par un graphique en camembert, où la taille de chaque portion reflète l'importance relative des établissements publics et privés dans cette région. De plus, chaque département est coloré en fonction du type de communes qui prédomine sur son territoire, déterminé par la médiane des valeurs d'unités urbaines des communes concernées.

Un constat évident à première vue est la prédominance des établissements publics dans tous les contextes urbains, y compris les zones rurales. Cette observation confirme l'idée largement acceptée selon laquelle les établissements publics sont plus nombreux que les établissements privés.

Un autre constat important réside dans le fait que dans les secteurs moins urbains, le nombre d'établissements publics est généralement moins élevé. Les valeurs les plus élevées sont généralement observées dans des zones plus urbaines. Cette observation suggère que la localisation géographique joue un rôle significatif dans la répartition des établissements publics et privés.

Ces observations diverses renforcent l'idée que la différence de végétalisation entre les établissements publics et privés est influencée par leur emplacement géographique. Les établissements publics, étant plus nombreux dans les espaces ruraux, pourraient ainsi bénéficier davantage des caractéristiques environnementales favorables à la végétalisation, telles que la disponibilité d'espaces verts et de zones moins densément construites. Cette relation entre la localisation géographique, la typologie des établissements et la végétalisation mérite une analyse plus approfondie pour mieux comprendre les facteurs qui influencent la présence de végétation autour des écoles.

En prenant en compte ces données sur les proportions d'établissements publics et privés selon leur contexte géographique, nos conclusions sur la relation entre le statut public ou privé des établissements et la végétalisation dans leur environnement proche sont renforcées. Cette approche permet d'appréhender de manière plus complète et solide les facteurs qui influencent la végétalisation autour des établissements scolaires en France.

c) Taux de végétalisation autour des écoles par départements.

### Taux de végétalisation des alentours d'établissements

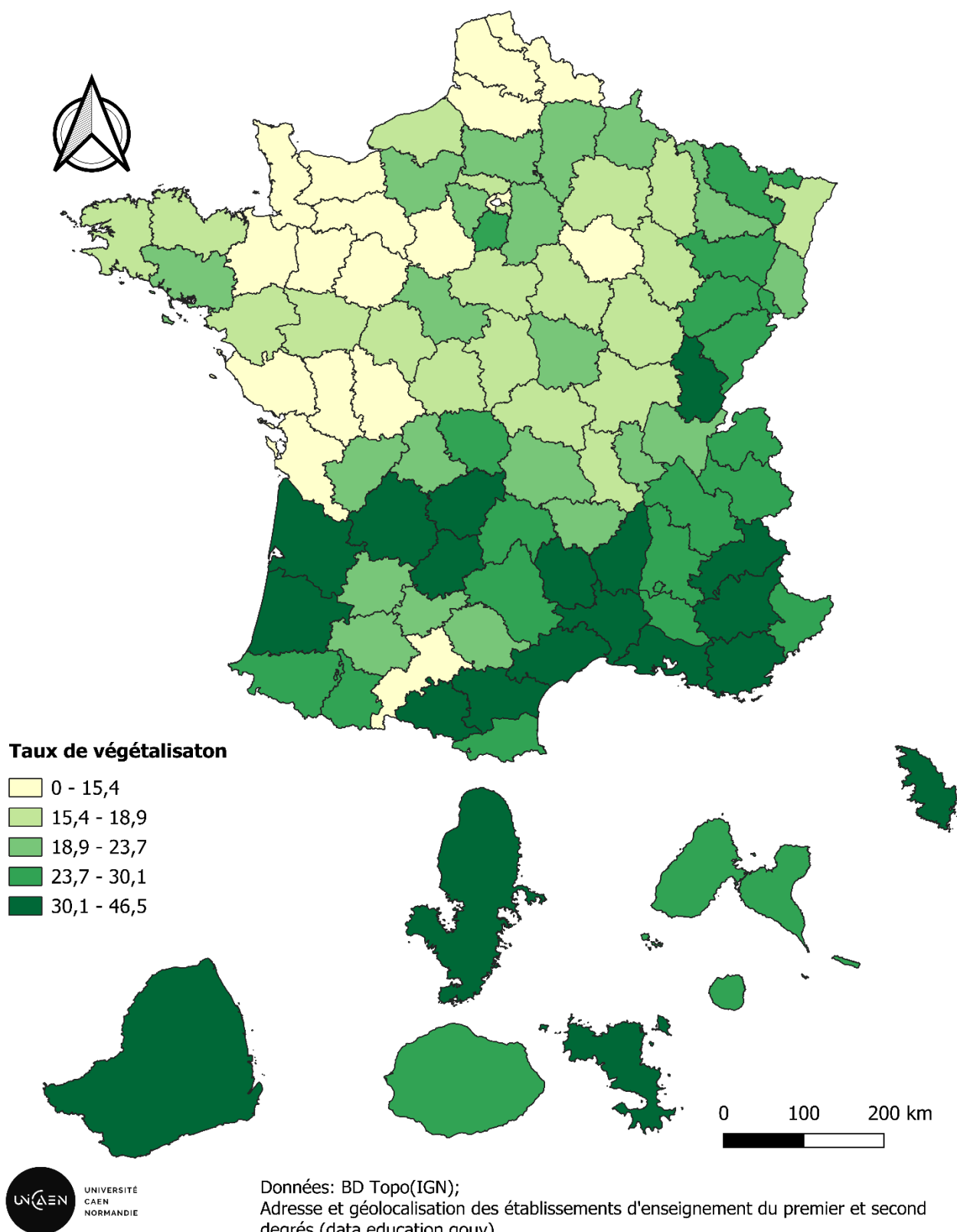


FIGURE 5: MOYENNE DEPARTEMENTALE DE VEGETATION AUTOUR DES ETABLISSEMENTS SCOLAIRES.

Cette carte offre une représentation visuelle de la moyenne de la végétation qui s'étend sur un rayon de 1000 mètres autour des écoles, organisée par département. En d'autres termes, pour chaque département, la valeur indiquée sur la carte représente la moyenne de la surface de végétation environnante pour les écoles situées dans ce département. Cette carte nous permet d'observer et de comparer les niveaux moyens de végétation dans les environs des écoles à travers les différents départements.

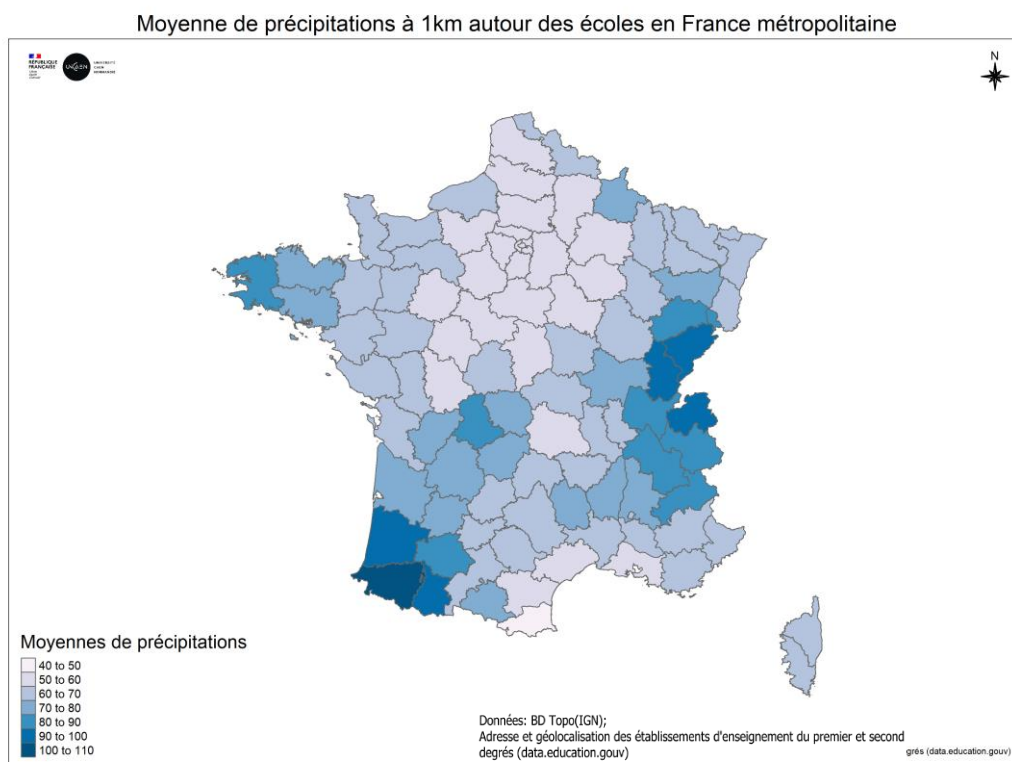
L'observation générale révèle que les établissements scolaires situés dans les départements du sud de la France sont en moyenne plus entourés de végétation que ceux situés dans les départements plus au nord du pays. Tous les départements situés au-dessus de la ceinture géographique de l'Hexagone affichent une moyenne de végétation à proximité de leurs écoles inférieure à 29%. C'est uniquement en dessous de cette démarcation que les établissements scolaires sont généralement plus végétalisés dans leur environnement proche.

Il convient de souligner que la région corse et ceux d'outre-mer se distinguent par la densité de végétation dans ses environs. Cette particularité découle de l'isolement de l'île, qui lui confère une biodiversité et une variété végétale uniques. Ces facteurs peuvent expliquer la prédominance de la végétation autour des écoles.

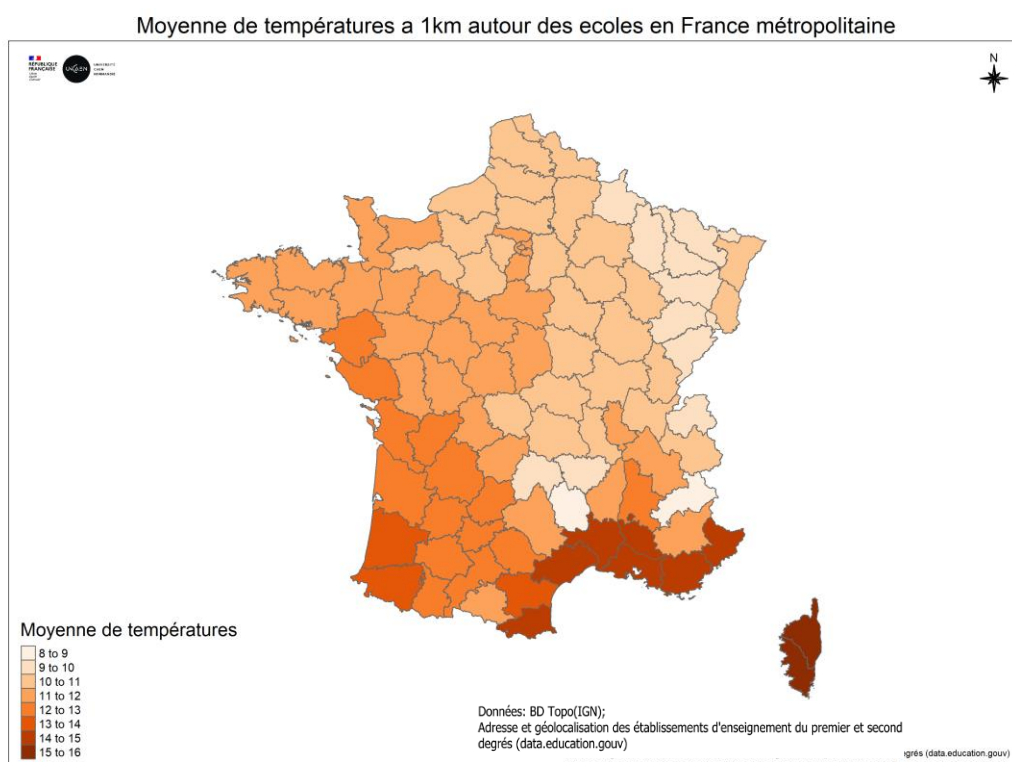
#### **d) Températures et précipitations moyennes autour des établissements scolaires**

Nous nous tournons à présent vers l'étude des températures et précipitations dans un périmètre d'un kilomètre autour des établissements scolaires, à travers l'ensemble du territoire français. Cette investigation s'inscrit dans le souci de comprendre les variations climatiques locales et leur impact potentiel sur les environnements éducatifs.

Les cartes présentant les températures et les précipitations moyennes se révèlent être des outils visuels inestimables. Les nuances de couleurs qui les composent fournissent une illustration visuelle des valeurs climatiques moyennes enregistrées dans les zones environnantes des écoles.



**FIGURE 6 : MOYENNE DEPARTEMENTALE DE PRECIPITATION AU TOUR DES ETABLISSEMENTS SCOLAIRES**



**FIGURE 7: MOYENNE DEPARTEMENTALE DE TEMPERATURES AUTOUR DES ETABLISSEMENTS**



La différenciation cartographique, basée sur les départements, apporte des informations cruciales pour déceler les tendances climatiques près des établissements éducatifs. Les données recueillies et regroupées par départements dévoilent des variations climatiques distinctes, susceptibles d'être influencées par des éléments tels que la topographie, l'urbanisation, la végétation et les effets microclimatiques. Bien que nous ne nous penchions pas sur ces facteurs dans le présent contexte, il est essentiel de reconnaître leur rôle potentiel dans les différences climatiques observées.

En se penchant sur les données relatives aux températures, il apparaît que les moyennes varient dans une fourchette allant de 8,72 à 15,68 degrés Celsius. Cette variation met en lumière une tendance significative : à mesure que l'on descend vers le sud du pays, les températures moyennes tendent à s'élever. Cette gradation thermique est particulièrement frappante sur la côte méridionale de la France. Un autre aspect notable est l'inclusion de la Corse parmi les départements où les températures sont notablement plus modérées, ajoutant une nuance intéressante à ce schéma climatique.

D'autre part, une analyse plus fine peut être opérée en distinguant les régions orientales des régions occidentales de la France métropolitaine. En général, les températures se montrent moins élevées dans l'est du pays que dans sa partie occidentale. Cette différenciation peut être attribuée à divers facteurs, tels que la géographie locale, la présence de masses d'eau ou encore les influences climatiques spécifiques à chaque région.

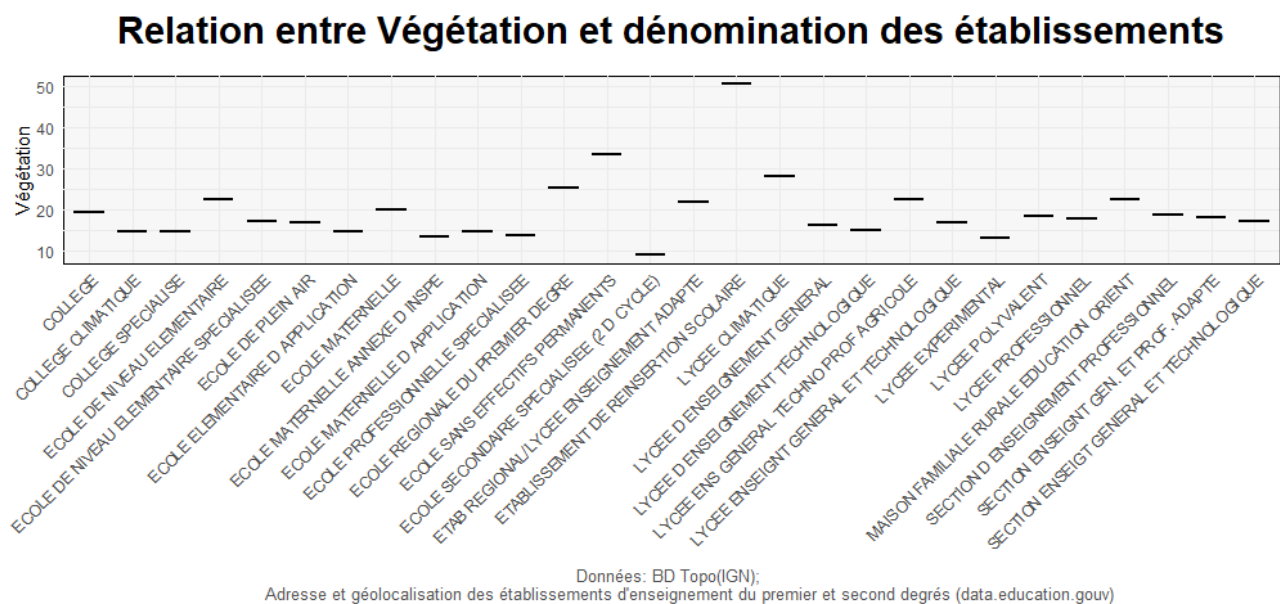
La corrélation entre des températures plus clémentes dans le sud de la France et une végétation plus abondante autour des écoles peut être explorée de manière fascinante. Le lien entre ensoleillement accru et processus de photosynthèse, qui alimente la croissance des végétaux, pourrait jouer un rôle essentiel. Les conditions climatiques plus favorables dans le sud offrent un environnement idéal pour que les espèces végétales prospèrent. Par conséquent, les alentours des établissements scolaires dans cette région pourraient bénéficier d'une végétalisation plus marquée, ce qui pourrait non seulement contribuer à l'embellissement visuel, mais également créer un cadre propice à l'apprentissage en plein air et à la sensibilisation à l'environnement.

Plongeant dans l'analyse des données de précipitations, il émerge un panorama de moyennes pluviométriques allant de 49,7 à 106,4 millimètres. Cette dispersion révèle des disparités régionales marquées, où les régions côtières ressortent comme des zones de concentration pluvieuse, en contraste avec la bande centrale de l'hexagone qui semble afficher des quantités moins significatives de précipitations. Cependant, ce schéma, contrairement aux corrélations plus évidentes entre températures et végétation, ne suit pas une trajectoire linéaire nord-sud aussi prononcée.

Il est cependant fascinant de souligner que la majorité des départements situés au sud de la ceinture centrale de la France affichent des moyennes de précipitations supérieures à 88,5 mm. En revanche, dans le nord, les quantités de précipitations ont tendance à se raréfier, en particulier aux abords de la région Île-de-France. Cette configuration, bien que prédominante, est nuancée par des exceptions qui méritent une attention particulière. La Bretagne, la Bourgogne et le Grand Est se détachent de cette tendance en enregistrant des niveaux de précipitations plus élevés, créant ainsi des nuances climatiques qui nécessitent une explication plus approfondie.

La Bretagne, par exemple, trouve son explication en partie dans sa proximité avec les masses d'eau de l'océan, qui peut induire des régimes de précipitations distincts. Cette proximité maritime peut jouer un rôle de modérateur climatique, apportant des précipitations plus abondantes. En outre, d'autres facteurs géographiques, topographiques et météorologiques, auxquels nous ne nous sommes pas encore attardés, pourraient également jouer un rôle dans ces variations observées, notamment dans les régions de l'est qui présentent des schémas inhabituels compte tenu de leur éloignement de l'océan.

#### e) Nature des établissements et végétation



**FIGURE 8: MOYENNE DE VEGETALISATION DES ALENTOURS D'ECOLES EN FONCTION DE LEUR NATURE**

Le schéma présenté ci-dessus illustre la moyenne de la végétation aux abords des écoles en relation avec leur catégorisation. Cette catégorisation est basée sur la dénomination principale des écoles.

La "dénomination principale des établissements scolaires" se réfère à la catégorisation ou à la classification générale qui identifie et caractérise le type spécifique d'établissement éducatif. Cette désignation est basée sur le niveau d'enseignement offert, la structure organisationnelle et parfois même l'orientation pédagogique de l'établissement. Elle englobe toute une gamme d'institutions éducatives allant des écoles primaires aux lycées en passant par les collèges, les écoles techniques et professionnelles, ainsi que d'autres formes d'écoles spécialisées. La "dénomination principale" d'un établissement peut également refléter sa vocation ou son domaine d'expertise. Par exemple, un lycée technique se concentrera davantage sur les programmes d'enseignement technique et professionnel, tandis qu'une école primaire mettra l'accent sur les premières étapes de l'éducation fondamentale. La catégorie à laquelle les établissements appartiennent détermine leur champ d'action dans l'éducation des

enfants. En englobant les collèges et les lycées, toutes les variétés d'institutions éducatives présentes dans notre ensemble de données sont prises en compte.

L'analyse approfondie de notre diagramme met en lumière une relation notable entre la diversité des types d'établissements et le degré de verdure qui les entoure. Les établissements éducatifs dédiés à la réinsertion des élèves ainsi que les collèges et lycées à vocation climatique se démarquent de manière frappante en affichant un niveau de végétation nettement plus élevé que les autres catégories d'établissements.

Cette distinction pourrait s'expliquer par l'importance accordée à la nature au sein de ces établissements particuliers. Les établissements de réinsertion scolaire, par exemple, peuvent reconnaître la valeur thérapeutique et apaisante de la nature pour les élèves qui peuvent avoir des besoins spécifiques sur le plan émotionnel et social. De plus, les collèges et lycées à orientation climatique manifestent peut-être une sensibilisation accrue à l'impact de la végétation sur les conditions climatiques locales. La végétation, en tant qu'élément clé des processus climatiques, peut contribuer à réguler la température, à favoriser la circulation de l'air et à atténuer les effets des îlots de chaleur urbains.

Cette variation marquée dans le niveau de verdure autour des établissements scolaires reflète la diversité de leurs valeurs pédagogiques et de leurs priorités. Elle souligne également l'effort de certains établissements pour tirer parti des avantages de la nature, que ce soit pour le bien-être des élèves, la sensibilisation environnementale ou l'impact climatique local.

#### **4. Résultats scolaires et végétation**

Nous atteignons maintenant un moment décisif de notre analyse, où nous plongeons profondément dans l'examen du lien essentiel entre la présence de végétation et la performance scolaire des enfants. Cette phase d'analyse est d'une importance cruciale, car elle englobe non seulement le bien-être des élèves, mais aussi l'évaluation de leurs résultats académiques et de leur épanouissement global dans le contexte éducatif. Notre objectif ultime est de démêler les liens entre la scolarité et une multitude de variables liées au cadre environnemental dans lequel l'apprentissage se déroule. Parmi ces variables, nous accordons une attention particulière aux conditions climatiques et au caractère urbain des municipalités où les élèves poursuivent leurs études.

La végétation, en particulier, suscite notre intérêt car elle pourrait jouer un rôle significatif dans le développement cognitif et émotionnel des élèves, ainsi que dans leurs performances académiques. Les bienfaits de la nature sur la santé mentale et le bien-être physique sont de plus en plus reconnus, et il est impératif de comprendre comment ces avantages peuvent se traduire dans le contexte éducatif. En explorant en profondeur cette variable, nous aspirons à saisir les nuances de son impact sur les élèves, de la manière dont elle peut influencer leur concentration en classe à sa capacité à créer un environnement propice à l'apprentissage.

De plus, en intégrant les conditions climatiques et le contexte urbain dans notre analyse, nous nous efforçons de prendre en compte l'ensemble du spectre environnemental dans lequel nos élèves évoluent. Les variations climatiques peuvent influencer non seulement la motivation des élèves, mais aussi leur capacité à se rendre à l'école et à participer pleinement à leurs activités scolaires. D'autre part, les aspects urbains des zones d'apprentissage peuvent avoir un impact sur la qualité de vie des élèves, de l'accessibilité aux parcs et espaces verts aux niveaux de bruit et de pollution atmosphérique. Tout cela peut affecter leur bien-être scolaire global.

En résumé, cette analyse exhaustive vise à établir une compréhension approfondie de la manière dont l'environnement physique peut influencer la trajectoire éducative des élèves. Les conclusions de cette étude pourraient ouvrir la voie à des recommandations politiques et à des initiatives éducatives visant à optimiser les conditions d'apprentissage pour les générations futures, en tenant compte des éléments tels que la végétation, le climat et le contexte urbain.

### **a) Rapport entre valeur ajoutée de réussite des établissements et végétation**

Dans notre analyse, nous allons examiner la valeur ajoutée des collèges et lycées en relation avec leurs taux de réussite. Les indicateurs de valeur ajoutée sont des outils d'évaluation utilisés à la fois pour les lycées et les collèges. Ils visent à mesurer l'impact spécifique de chaque établissement sur la réussite de ses élèves, que ce soit au baccalauréat pour les lycées ou au brevet des collèges pour les établissements du secondaire.

Ces indicateurs vont au-delà des simples taux de réussite bruts en examen, en cherchant à isoler l'influence des facteurs extérieurs à l'établissement pour évaluer son action propre. Pour juger de l'efficacité d'un lycée ou d'un collège, les performances de ses élèves sont comparées à celles d'élèves similaires dans des établissements comparables. Les indicateurs en "valeur ajoutée" sont donc utilisés pour analyser ces résultats.

Pour chaque lycée et collège, la valeur ajoutée mesure la différence entre les résultats réels obtenus par les élèves et les résultats attendus, en tenant compte de divers facteurs tels que l'âge, le sexe, le niveau scolaire à l'entrée, le profil socio-économique des élèves et d'autres caractéristiques propres à l'établissement. La valeur ajoutée est une mesure relative : si elle est positive, cela signifie que l'établissement a contribué de manière significative à la réussite de ses élèves au-delà des attentes, compte tenu de leur profil. En revanche, une valeur ajoutée négative indique que les résultats de l'établissement sont en deçà de la moyenne des établissements comparables.

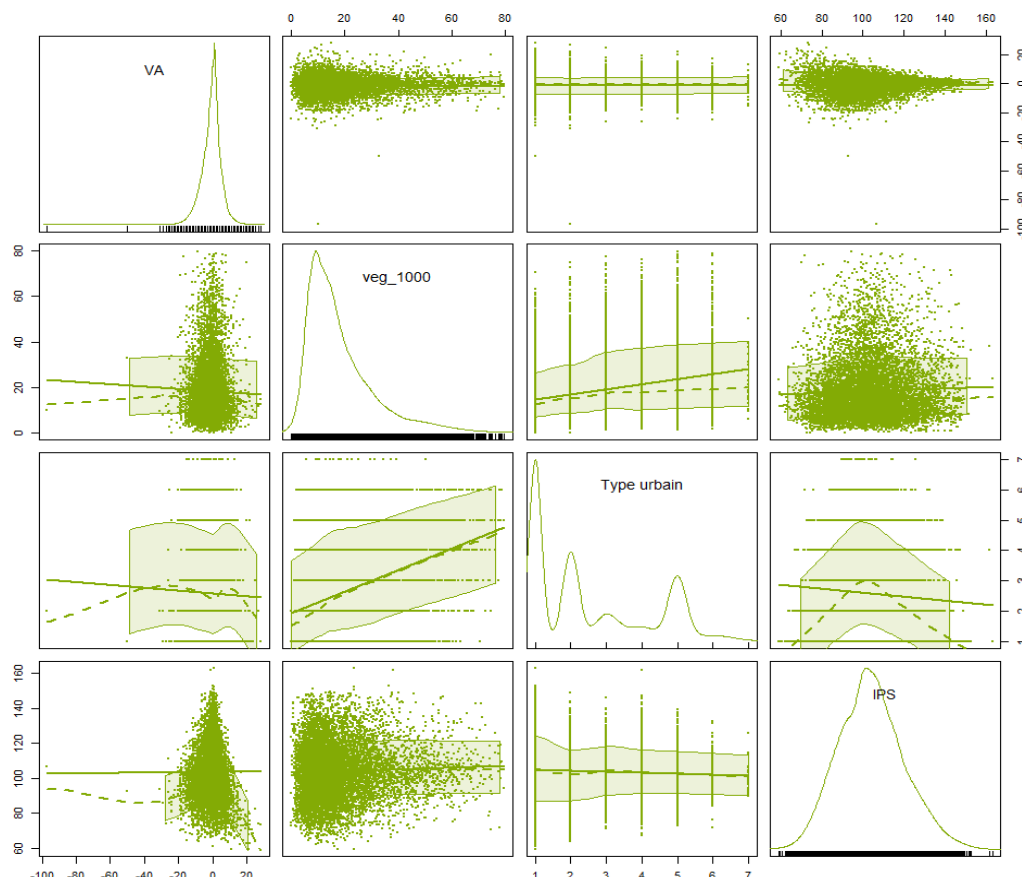
Il est important de noter que ces indicateurs sont publiés uniquement pour les lycées publics et privés sous contrat, tandis que pour les collèges, ils portent sur les résultats au brevet. De plus, les données de valeur ajoutée ne sont pas divulguées lorsque le nombre de candidats est en deçà d'un certain seuil, pour garantir la confidentialité des données.

Le cas échéant, notre objectif consistera à démontrer la relation entre la valeur ajoutée des collèges et lycées d'une part, et les variables exogènes associées au contexte environnemental et aux conditions socio-économiques, d'autre part. Nous examinerons plusieurs variables, notamment la proximité de la végétation dans un rayon de 1000 mètres autour des établissements, le type d'unité urbaine auquel ils sont rattachés, les données sur les températures et les précipitations moyennes, ainsi que le niveau socio-économique global des établissements.

Nous commencerons cette section par l'analyse d'une matrice de graphiques cartésiens. Cette matrice nous offrira une vision globale des relations entre les différentes paires de variables.

Nous nous concentrerons spécifiquement sur les variables de végétation, de type d'unité urbaine et de niveau socio-économique, que nous comparerons à la valeur ajoutée des établissements.

**Matrice de graphiques cartésiens de valeurs ajoutées liées aux taux de réussite des lycées et collèges**



**FIGURE 9: MATRICES CARTESIENNE DE VARIABLES LIEES A LA VALEUR AJOUTEE DE REUSSITE**

La figure présente ci-dessus représente une matrice de graphiques qui met en évidence les relations entre chaque paire de variables. Les graphiques situés en diagonale représentent la distribution de chaque variable par rapport à elle-même. Les graphiques en dehors de la diagonale illustrent les différentes relations entre les paires de variables. Une ligne de régression permet de visualiser la tendance générale de la relation pour chaque paire de variable.

La variable VA (valeur ajoutée), situé dans le coin supérieur gauche, représente le point central de notre analyse dans le graphique. Les autres graphiques de la même colonne nous offrent un aperçu des relations entre la valeur ajoutée et les différentes variables examinées.

En ce qui concerne la variable de végétation, nous observons une tendance notable : à mesure que le degré de végétation augmente, la valeur ajoutée des établissements a tendance à diminuer. Cette constatation suggère qu'une concentration plus élevée de végétation est associée à une baisse de la valeur ajoutée des établissements. Cette relation pourrait être le reflet de plusieurs facteurs d'ordre structurelle ou autre qui influent moins dans les espaces plus verts.

De même, en ce qui concerne la variable du type d'unité urbaine, nous constatons une tendance similaire. Plus nous nous dirigeons vers des zones qualifiées de "rurales" dans ce contexte, plus la valeur ajoutée des établissements a tendance à baisser. Cela suggère que les établissements situés dans des zones urbaines plus densément peuplées ou développées ont tendance à générer une valeur ajoutée plus élevée que ceux situés dans des zones rurales. Les raisons derrière cette relation pourraient inclure un accès plus facile aux infrastructures d'enseignement, une main-d'œuvre plus qualifiée, ou des conditions d'étude plus adéquates.

Cependant, en ce qui concerne la variable IPS (Indice de position sociale) dans les collèges et lycées, la relation est beaucoup moins prononcée. En fait, la relation semble presque linéaire, ce qui suggère qu'il n'y a pas de lien significatif entre le niveau socio-économique et la valeur ajoutée des établissements. Cela pourrait indiquer que, dans le contexte des collèges et lycées, d'autres facteurs sont prédominants pour expliquer les variations de la valeur ajoutée, tels que la qualité de l'enseignement, les ressources disponibles, ou d'autres variables spécifiques au secteur de l'éducation.

En jetant un coup d'œil rapide à la relation entre la végétation et le type d'unité urbaine (deuxième colonne, troisième ligne), une tendance frappante se dégage. En effet, on observe une augmentation nette de la végétation à mesure que l'on s'éloigne de l'urbain. Cette observation semble être en corrélation avec la densité de construction des espaces urbains. Plus précisément, les zones urbaines se caractérisent généralement par une densité de bâtiments plus élevée, ce qui pourrait expliquer la faible présence de végétation par rapport aux zones plus rurales, où la densité de construction est moindre. En d'autres termes, l'espace limité dans les zones urbaines laisse moins de place pour la végétation, tandis que les zones rurales offrent plus d'espace pour la croissance végétale en raison de leur caractère moins urbanisé.

Une autre observation pertinente concerne la variable IPS (Indice de position sociale) qui décroît à mesure que l'on s'éloigne vers des espaces plus ruraux. Cette tendance suggère que, dans cette étude ou dans ce contexte particulier, il existe une corrélation entre le niveau socio-économique et le degré d'urbanisation. En d'autres termes, les zones urbaines peuvent avoir tendance à avoir un niveau socio-économique plus élevé que les zones rurales. Il convient de noter que cette observation pourrait être le reflet de divers facteurs. Par exemple, les opportunités économiques, l'accès à l'éducation, ou les services sociaux peuvent varier entre les zones urbaines et rurales, influençant ainsi le niveau socio-économique des habitants de ces régions.

Nous allons maintenant nous concentrer sur l'estimation d'un modèle de régression linéaire multiple afin de mieux comprendre la véritable relation entre la valeur ajoutée et les diverses variables qui pourraient exercer une influence sur elle. Cette démarche vise à élucider de manière plus approfondie la manière dont ces variables interagissent avec la valeur ajoutée.

La régression linéaire multiple est une technique statistique utilisée en analyse de données et en apprentissage automatique pour modéliser la relation entre une variable dépendante (ou cible) et plusieurs variables indépendantes (ou explicatives) en supposant une relation linéaire entre elles. Cette méthode vise à trouver une équation linéaire qui décrit au mieux la variation de la variable dépendante en fonction des variables indépendantes.

Plus formellement, la régression linéaire multiple peut être représentée par l'équation suivante :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

où :

- Y représente la variable dépendante que l'on souhaite prédire. Dans le cas présent il s'agit de la valeur ajoutée
- $X_1, X_2, \dots, X_k$  représentent les variables explicatives.
- $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  sont les coefficients de régression qui mesurent l'impact de chaque variable explicative sur la variable dépendante.
- $\varepsilon$  représente l'erreur résiduelle, c'est-à-dire la différence entre les valeurs observées de la variable dépendante et les valeurs prédites par le modèle.

L'objectif de la régression linéaire multiple est de trouver les coefficients  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  qui minimisent l'erreur résiduelle, généralement mesurée en utilisant la méthode des moindres carrés. Une fois que le modèle est ajusté, il peut être utilisé pour faire des prédictions sur la variable dépendante en fonction des valeurs des variables indépendantes.

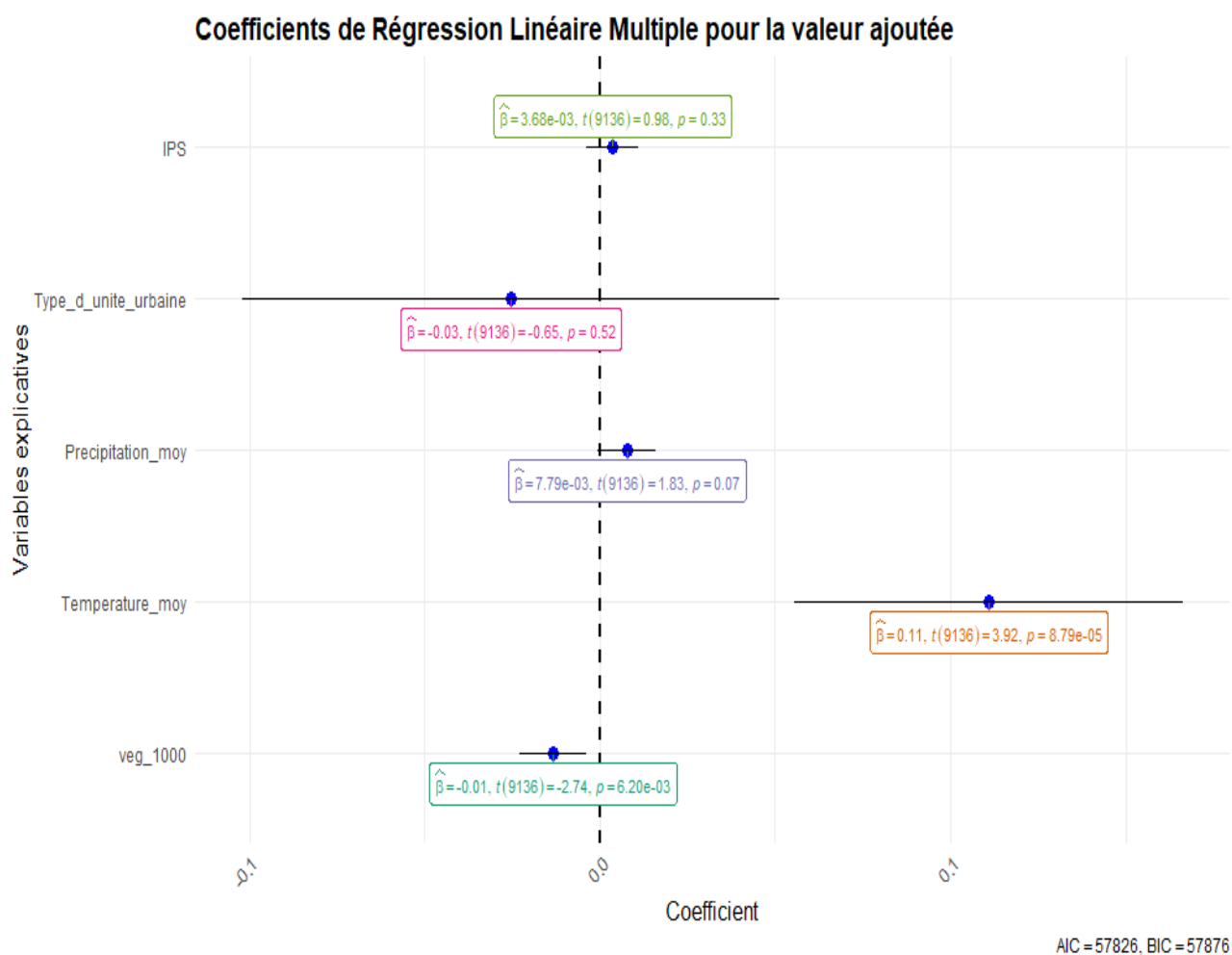
En résumé, la régression linéaire multiple est une méthode statistique qui permet de modéliser et de quantifier les relations linéaires entre une variable dépendante et plusieurs variables indépendantes, ce qui la rend utile pour la prédiction et l'analyse des données.

Nous avons appliqué cette méthode à la variable "valeur ajoutée". L'objectif principal est d'expliquer, d'un point de vue statistique, l'évolution de cette variable en fonction des autres variables, à savoir : la végétation, la température moyenne, les précipitations moyennes, le type d'unité urbaine et l'indice de position sociale. Cette démarche vise à établir des relations quantitatives entre la valeur ajoutée et ces diverses variables, permettant ainsi une analyse approfondie des facteurs qui influencent la valeur ajoutée dans le contexte étudié. Les résultats de cette régression se présentent comme suit :



<b>**Characteristic**</b>	<b>**Beta**</b>	<b>**95% CI**</b>	<b>**p-value**</b>	<b>**VIF**</b>
veg_1000	-0.01	-0.02, 0.00	0.006	1.1
Temperature_moy	0.11	0.06, 0.17	<0.001	1.6
Precipitation_moy	0.01	0.00, 0.02	0.067	1.6
Type_d_unite_urbaine	-0.03	-0.10, 0.05	0.5	1.2
IPS	0.00	0.00, 0.01	0.3	1.0

**TABLEAU 2 : REGRESSION DE LA VALEUR AJOUTEE DE REUSSITE DES ETABLISSEMENTS SCOLAIRES**



**FIGURE 10: REGRESSION MULTIPLE DE LA VARIABLE DE VALEUR AJOUTEE DE REUSSITE**

Plusieurs informations ressortent de ces différentes images.

Les coefficients bêta, dans une certaine mesure, reflètent l'impact relatif de chaque variable sur la valeur ajoutée lorsqu'elle augmente d'une unité. Lorsque le coefficient bêta est égal à -0,01 pour la végétation, cela signifie qu'une augmentation d'une unité de la valeur ajoutée s'accompagnera d'une diminution de 0,01 de la végétation. En conséquence, à mesure que la valeur ajoutée augmente d'une unité, la végétation diminuera de 0,01, la température augmentera de 0,11, les précipitations augmenteront également de 0,007, et l'indice de position sociale augmentera de 0,003. En d'autres termes, chaque fois que les variables mentionnées connaissent une augmentation correspondant aux coefficients précédemment mentionnés, la valeur ajoutée augmente d'une unité.

Le type d'unité urbaine n'est pas inclus dans cette explication en raison d'une p-value élevée. En effet, une p-value de 0,5 signifie qu'il y a une probabilité de 50 % de se tromper. Étant donné que cette variable est en réalité qualitative, une régression linéaire pourrait ne pas être la meilleure approche pour l'étudier. De plus, il convient d'émettre des réserves concernant l'Indice de Position Sociale (IPS), qui présente tout de même une marge d'erreur de 30 %.

L'équation qui explique le mieux la variable VA par rapport aux autres variables est la suivante :

$$VA = -2.703 + (-0.0133 \times \text{veg\_1000}) + (0.111 \times \text{Temperature\_moy}) + (0.0078 \times \text{Precipitation\_moy}) - (0.0253 \times \text{Type\_d\_unite\_urbaine}) + (0.0037 \times \text{IPS})$$

Nous allons maintenant nous pencher sur la signification de la valeur VIF. La Variance Inflation Factor (VIF), ou Facteur d'Inflation de la Variance en français, est une mesure essentielle pour évaluer la présence de multi colinéarité entre les variables indépendantes dans une régression linéaire multiple. La multi colinéarité survient lorsque deux ou plus de deux variables indépendantes dans un modèle sont fortement corrélées les unes avec les autres. En général, une VIF supérieure à 5 est considérée problématique.

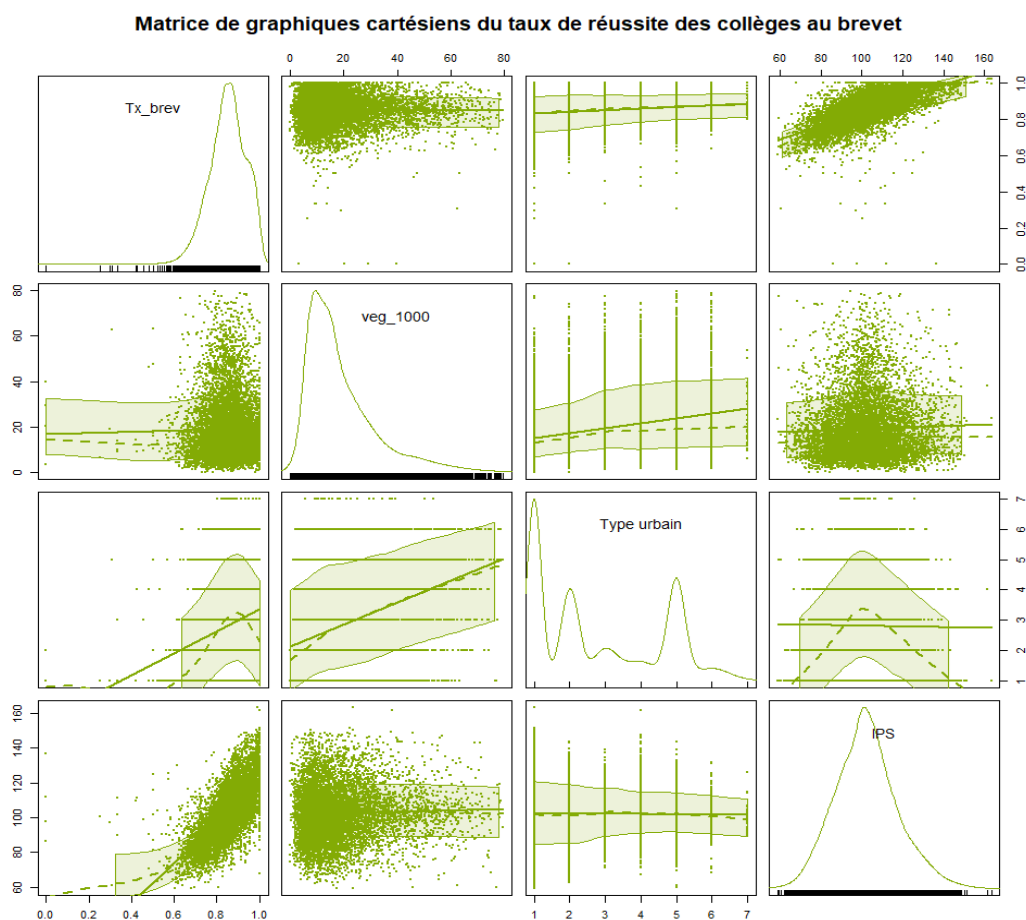
Dans notre cas, il n'y a pas de problème de VIF pour aucune de nos variables indépendantes, sauf pour la variable IPS. Toutes les autres variables présentent une certaine forme de multi colinéarité. Cela signifie que, en dehors de la variable IPS, les autres variables interagissent entre elles de manière significative, ce qui peut compliquer l'interprétation des coefficients de régression et nécessite une attention particulière lors de l'analyse des résultats.

### **b) Rapport entre résultats au brevet et végétation**

Pour aller plus en profondeur dans l'analyse des résultats des collèges et lycées, nous pouvons élargir notre perspective en examinant les performances des élèves aux examens nationaux, notamment le diplôme national du brevet. Depuis 2019, le ministère de l'Éducation met à disposition des données sur les résultats du brevet, permettant ainsi une analyse plus approfondie. Ces données comprennent les résultats par établissement, couvrant à la fois la France métropolitaine et les départements et régions d'Outre-Mer.

En utilisant une méthodologie similaire à celle précédemment employée pour évaluer la valeur ajoutée du taux de réussite des établissements, c'est-à-dire une régression linéaire multiple, nous avons obtenu des résultats presque identiques. Cette approche statistique nous permet de comprendre les facteurs qui influencent les performances des élèves au brevet, en prenant en compte les variables explicatives à savoir : la végétation, la température et les précipitations autour des écoles, le type d'unité urbaine auquel appartiennent les écoles et l'indice de position social dans les écoles. Ces variables pourraient inclure des facteurs tels que le niveau socio-économique des élèves, les ressources disponibles dans les établissements, la qualité de l'enseignement, et d'autres éléments qui peuvent jouer un rôle dans la réussite scolaire. Cela nous permettrait un examen plus détaillé des résultats au brevet, nous pouvons non seulement évaluer la performance globale des établissements, mais également obtenir une meilleure compréhension de l'impact de la végétation dans le monde de l'enseignement.

En appliquant la même méthodologie nous obtenons les résultats dont nous feront a présenté une analyse approfondie.



**FIGURE 11: MATRICE CARTESIENNE DE VARIABLES LIEES AUX RESULTATS DU BREVET.**

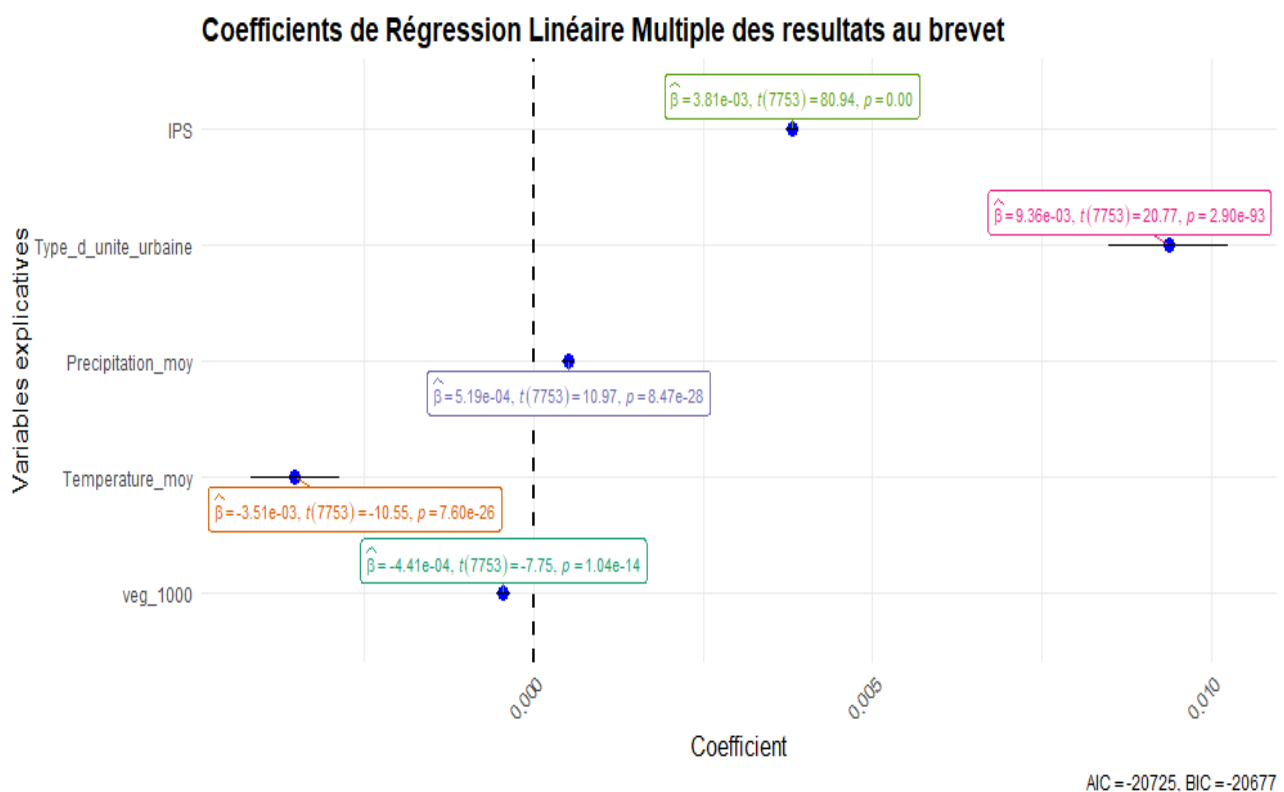
Ce graphique représente la matrice de régression bidimensionnelle entre trois variables : le taux de réussite au brevet, l'indice de position sociale et le type d'unité urbaine des établissements.

En ce qui concerne la variable "végétation", la ligne de régression présente une tendance quasi-linéaire. Cette observation suggère qu'il n'existe pas de lien significatif entre la quantité de végétation et le taux de réussite des élèves au cours de la période allant de 2006 à 2021. Cette constatation renforce l'idée selon laquelle la végétation n'a pas d'influence directe sur les performances scolaires. Il semble donc que d'autres facteurs, en particulier d'ordre structurel, jouent un rôle prédominant dans l'explication des variations du taux de réussite. Pour mieux comprendre ces facteurs et leur impact, une analyse plus approfondie serait nécessaire.

Quant à la relation entre le taux de réussite et le type d'unité urbaine auquel appartiennent les établissements, il n'existe pas d'un point de vue visuel une relation nette étant donné que la courbe de régression est quasiment horizontale (ne connaît aucune presque pas de variation). Il est important de noter que le taux de réussite est calculé en divisant le nombre d'admis par le nombre d'étudiants présentés. Dans ce contexte, il est intéressant de noter que les écoles situées dans des bourgs moins urbains n'ont pas nécessairement plus d'élèves que les établissements situés dans de grands pôles urbains. Par conséquent, le taux de réussite peut être influencé par cette disparité en termes de nombre d'élèves. Cependant, le constat demeure clair et indiscutable : les établissements dans des unités urbaines moins peuplées ont généralement des taux de réussite plus élevés.

Lorsqu'on examine l'indice de position sociale, le constat est effectivement similaire et même plus frappant. Le taux de réussite des établissements augmente à mesure que la position socio-économique globale augmente. Cette observation pourrait s'expliquer par plusieurs facteurs. Tout d'abord, dans les écoles situées dans des zones socio-économiquement plus aisées, il est probable que la structure d'enseignement soit plus développée et adaptée. Cela signifie que ces écoles peuvent offrir des ressources supplémentaires, un personnel enseignant plus expérimenté, des programmes éducatifs plus riches, et une infrastructure plus propice à l'apprentissage. Tout cela peut contribuer à de meilleures performances des élèves. De plus, les élèves issus de milieux socio-économiques plus élevés ont souvent un accès plus large à des opportunités éducatives en dehors de l'école, telles que des cours particuliers ou des activités parascolaires, ce qui peut également renforcer leur réussite académique. Cependant, il est essentiel de noter que cette observation ne signifie pas nécessairement que la réussite scolaire est uniquement déterminée par la position socio-économique. D'autres facteurs, tels que la qualité de l'enseignement, la motivation des élèves, le soutien parental, et l'environnement d'apprentissage, jouent également un rôle crucial.

La prochaine étape consisterait à effectuer une régression linéaire multiple afin de déterminer l'impact réel de ces différentes variables sur les résultats des élèves. Cette analyse permettrait de comprendre plus en profondeur comment les variables interagissent pour influencer les performances scolaires.



**FIGURE 12: REGRESSION MULTIPLE DE LA VARIABLE RESULTAT DU BREVET**

<b>**Characteristic**</b>	<b>**Beta**</b>	<b>**95% CI**</b>	<b>**p-value**</b>	<b>**VIF**</b>
<b>veg_1000</b>	<b>0.00</b>	<b>0.00, 0.00</b>	<b>&lt;0.001</b>	<b>1.1</b>
<b>Temperature_moy</b>	<b>0.00</b>	<b>0.00, 0.00</b>	<b>&lt;0.001</b>	<b>1.7</b>
<b>Precipitation_moy</b>	<b>0.00</b>	<b>0.00, 0.00</b>	<b>&lt;0.001</b>	<b>1.7</b>
<b>Type_d_unite_urbaine</b>	<b>0.01</b>	<b>0.01, 0.01</b>	<b>&lt;0.001</b>	<b>1.2</b>
<b>IPS</b>	<b>0.00</b>	<b>0.00, 0.00</b>	<b>&lt;0.001</b>	<b>1.0</b>

**TABEAU 3: TABLE DE REGRESSION DE LA VARIABLE RESULTAT AU BREVET**

Il est crucial de commencer par examiner les valeurs de p-valeur, qui sont remarquablement basses dans tous les cas ( $< 0,001$ ). Cette faible valeur de p-valeur indique une signification statistique considérable des résultats obtenus à partir de la régression linéaire. Une telle signification renforce de manière significative la crédibilité des conclusions à venir. De plus, la valeur relativement basse du VIF (inférieure à 2 dans tous les cas) apporte encore plus de poids à ces conclusions. Cela renforce la confiance dans l'interprétation des effets des variables indépendantes sur les résultats des élèves et contribue à une analyse plus robuste et convaincante.

Les résultats de la régression fournissent des informations importantes sur la relation entre plusieurs variables et le taux de réussite au brevet. Les influences variées des variables explicatives ont donc été mis à jour avec une significativité importante des conclusions qui en découlent.

En effet, la végétation influence négativement le taux de réussite dans les collèges. Lorsque le taux de réussite augmente d'une unité, la quantité de végétation diminue de  $-4.413e-04$ , et vice versa. Cela suggère que plus il y a de végétation autour des écoles, plus le taux de réussite est bas. Cependant, il est important de noter que cela ne signifie pas nécessairement que la végétation a un impact négatif sur les résultats scolaires. D'autres facteurs, tels que la structure du système éducatif local, doivent également être pris en compte.

Lorsque le taux de réussite augmente d'une unité, les températures moyennes diminuent de  $-3.514e-03$ , ce qui signifie que les étudiants ont tendance à mieux étudier dans des températures plus douces. Cette observation suggère que des températures plus clémentes peuvent favoriser un environnement d'apprentissage plus propice.

Lorsque le taux de réussite augmente d'une unité, les précipitations moyennes augmentent en moyenne de  $5.191e-04$ . Cela peut être interprété comme une indication que les zones sujettes à des précipitations plus élevées sont plus propices à l'apprentissage. Il est à noter que cela peut être lié aux températures plus douces dans ces zones.

Une augmentation d'une unité du taux de réussite est associée à une augmentation moyenne de  $9.362e-03$  du type d'unité urbaine. Cela confirme que, en général, plus une zone tend vers le rural, plus le taux de réussite au brevet est élevé.

L'indice de position sociale augmente de  $3.812e-03$  lorsque le taux de réussite au brevet augmente d'une unité. Cette observation renforce l'idée que la position socio-économique a une influence positive sur les résultats scolaires des élèves. En d'autres termes, les élèves issus de milieux socio-économiques plus élevés ont tendance à obtenir de meilleurs résultats.

En résumé, ces résultats de régression fournissent des informations utiles pour comprendre comment différentes variables influencent les performances scolaires des élèves. Cependant, il est important de noter que la corrélation ne signifie pas nécessairement une causalité directe, et d'autres facteurs peuvent également jouer un rôle dans les résultats scolaires. Une analyse plus approfondie pourrait être nécessaire pour explorer davantage ces relations.

Lorsque l'on examine les données recueillies et les analyses effectuées, il est tentant de conclure que la présence de végétation a un impact potentiellement néfaste sur les performances des lycées et collèges en France. Cependant, il est important de souligner que cette conclusion ne peut être assimilée à une relation de cause à effet. En réalité, d'autres variables jouent un rôle bien plus significatif dans la détermination des résultats scolaires. Parmi ces variables, le contexte urbain tient une place prépondérante. Les établissements situés dans des zones urbaines peuvent être confrontés à des défis différents de ceux situés en milieu rural. Les élèves des zones urbaines peuvent être exposés à des influences environnementales, sociales et économiques distinctes qui influencent leurs performances académiques. Ainsi, attribuer les variations des résultats scolaires uniquement à la présence de végétation serait une simplification excessive. De plus, il est impératif de prendre en compte la position socio-économique des élèves. Les inégalités économiques et sociales peuvent avoir un impact majeur sur les opportunités éducatives et les ressources disponibles pour les élèves. Par conséquent, il est plausible que les résultats scolaires soient davantage influencés par la situation socio-économique des élèves que par la quantité de végétation entourant leur école. Enfin, il est essentiel de considérer les disparités existantes entre les communes qui abritent les écoles. Les caractéristiques géographiques, économiques et sociales de chaque commune varient considérablement, ce qui signifie que les résultats scolaires peuvent être influencés par un large éventail de facteurs locaux.

En résumé, bien que les données suggèrent un lien entre la végétation et les résultats scolaires, il est nécessaire d'exercer la prudence et de reconnaître que d'autres variables, telles que le contexte urbain et la situation socio-économique, ont un impact plus important. Par conséquent, il serait prématuré de conclure de manière définitive que la végétation a un effet négatif sur la réussite scolaire, en particulier lorsque l'on tient compte des multiples variables structurelles et contextuelles qui interviennent dans ce contexte complexe.

### c) Régression par catégorie d'unités urbaines

L'idée fondamentale qui émerge est que la végétation ne semble pas nécessairement avoir un impact négatif sur l'éducation des enfants, et que le taux de réussite scolaire dépend fortement de facteurs autres que la présence de végétation. Pour étayer cette idée, nous avons mis en place une approche simple, à savoir une régression linéaire qui examine la relation entre la végétation et la valeur ajoutée en matière de réussite dans les lycées et collèges, en prenant en compte la catégorisation de la végétation en fonction des types d'unités urbaines. En outre, nous avons appliqué la même méthodologie à l'analyse des résultats du brevet, en cherchant à comprendre comment les écoles se comportent dans des contextes urbains qualifiables de similaires. Cette démarche vise à identifier les véritables déterminants des performances scolaires et à évaluer si la végétation, une fois d'autres facteurs contrôlés, conserve toujours son influence négative qu'elle tend à avoir. Ces analyses plus détaillées devraient nous aider à mieux comprendre le rôle relatif de la végétation par rapport à d'autres influences sur la réussite scolaire.

Unité urbaine	Coefficient	Unité urbaine	Coefficient
1	-0.0316015	1	0.0000766
2	0.0103937	2	-0.0002296
3	-0.0246567	3	-0.0006582
4	0.0187440	4	0.0001470
5	-0.0139946	5	-0.0005823
6	0.0264024	6	-0.0001460
7	0.1425727	7	0.0012830

**TABEAU 4: COEFFICIENTS DE REGRESSION EN FONCTION DES UNITE URBAINES**

Les résultats que nous avons obtenus confirment en grande partie cette idée. En effet, il est remarquable que dans différents contextes des communes (urbains comme rural), la présence de végétation semble avoir un impact positif sur les performances des étudiants et des établissements. Que ce soit en termes de valeur ajoutée ou de résultats au brevet, lorsque nous avons isolé les établissements en fonction de leur indice d'urbanisme, les résultats ont varié de manière significative. Ces observations suggèrent clairement que le paysage environnant a un effet non négligeable sur les résultats des élèves.



Cette constatation met en évidence l'importance de considérer d'autres variables comme un facteur influençant les performances scolaires. Bien que la végétation puisse jouer un rôle dans les résultats scolaires, il est crucial de reconnaître que d'autres variables exercent une influence nettement plus significative sur la réussite éducative. Parmi ces variables, on peut citer l'infrastructure éducative, les ressources techniques et humaines mises à la disposition des élèves. L'infrastructure éducative comprend des éléments tels que la qualité des bâtiments scolaires, l'accès aux technologies de l'information, les bibliothèques, les laboratoires, et bien d'autres. Une infrastructure éducative adéquate crée un environnement propice à l'apprentissage, ce qui peut avoir un impact majeur sur les performances des élèves. Les ressources humaines sont également cruciales. La qualité des enseignants, leur formation, leur motivation et leur capacité à enseigner sont des facteurs clés pour la réussite des élèves. Les ressources administratives et de soutien, telles que les conseillers d'orientation et les psychologues scolaires, jouent également un rôle essentiel. D'autres facteurs, tels que la motivation des élèves, le soutien familial, l'accès à des programmes éducatifs enrichissants en dehors de l'école, et la stabilité sociale, peuvent également être déterminants dans la réussite scolaire.

## Conclusion

La végétation, en tant qu'élément essentiel pour le bien-être physique et psychologique, joue un rôle capital dans le processus de croissance et de développement psycho-intellectuel des enfants. Ce développement est inextricablement lié à l'éducation, ce qui souligne l'importance d'établir une connexion entre le rapport à la végétation et les performances scolaires des enfants. Dans ce document, nous entreprenons d'explorer les liens potentiels entre la présence de végétation et les résultats académiques dans les lycées et collèges.

Une étape cruciale de ce travail a consisté en la création d'une base de données solide. En effet, pour mieux comprendre les relations entre la végétation et les performances scolaires des établissements d'enseignement, il est essentiel de prendre en compte l'environnement naturel environnant. À cet effet, nous avons développé un programme spécialement conçu pour collecter des données sur le taux de végétalisation des environs des établissements scolaires en France.

De plus, nous avons intégré des données relatives aux paramètres climatiques, notamment la température et les précipitations. Nous avons également inclus d'autres paramètres socio-économiques et structurels dans notre analyse, afin d'évaluer les possibles corrélations et influences que ces facteurs pourraient avoir sur la relation entre la végétation et les performances scolaires des élèves.

Après une analyse approfondie de la relation entre la végétation, les performances scolaires, et d'autres variables influençant l'éducation, plusieurs constatations significatives émergent. Il est crucial de noter que la végétation, bien que présentant certaines corrélations avec les résultats scolaires, ne peut pas être considérée comme un facteur déterminant majeur dans la réussite des élèves dans les établissements d'enseignement. Les résultats de nos analyses suggèrent que d'autres variables, telles que le contexte urbain, la position socio-économique des élèves, la qualité de l'enseignement, et les ressources disponibles dans les écoles, ont un impact bien plus significatif sur les performances scolaires.

Tout d'abord, en ce qui concerne la végétation, il est vrai que nous avons observé une tendance dans certaines analyses indiquant une corrélation négative entre la densité de la végétation et la valeur ajoutée ou les résultats au brevet. Cependant, cette corrélation est relativement faible et peut être expliquée par d'autres facteurs. Par exemple, les écoles situées dans des zones urbaines plus denses ont tendance à avoir moins d'espace vert, ce qui peut être lié à des défis structurels et environnementaux spécifiques à ces zones. Cependant, cette corrélation ne signifie pas que la végétation est la cause principale des variations des résultats scolaires.

En revanche, la position socio-économique des élèves et le contexte urbain se sont révélés être des facteurs bien plus importants pour expliquer les différences de réussite scolaire. Les écoles situées dans des zones urbaines plus denses ou des contextes socio-économiques plus élevés ont tendance à obtenir de meilleurs résultats, ce qui suggère que l'accès à des ressources éducatives, la qualité de l'enseignement, et d'autres éléments contextuels jouent un rôle crucial.

De plus, les analyses de régression linéaire multiple ont confirmé ces observations en montrant que la végétation, une fois d'autres facteurs contrôlés, n'a pas d'impact significatif sur les performances scolaires. En revanche, la température moyenne, les précipitations, le type d'unité urbaine, et l'indice de position sociale ont montré des relations significatives avec les résultats scolaires, bien que ces relations ne puissent pas être interprétées comme des relations de causalité directe. Si la présence de végétation peut jouer un rôle dans le cadre plus large de l'environnement scolaire, elle ne peut pas être considérée comme un facteur prédominant dans la réussite éducative. Les résultats scolaires sont

influencés par un ensemble complexe de variables, et il est essentiel de prendre en compte l'ensemble du contexte éducatif, environnemental et socio-économique pour comprendre pleinement les variations de la réussite scolaire. En se concentrant sur l'amélioration de l'enseignement, l'accès aux ressources éducatives, et la réduction des inégalités socio-économiques, les écoles peuvent mieux servir leurs élèves et favoriser leur réussite académique.

À la suite de ces diverses analyses, une question importante mérite d'être explorée. À présent que nous constatons qu'un lien direct entre les résultats scolaires et la végétation est peu explicatif, il devient intéressant d'explorer d'autres paramètres qui pourraient contribuer à la réussite scolaire des enfants. Le bien-être des élèves, le développement de leur mémoire, leur capacité d'attention et leur calme intérieur sont autant de facteurs qui méritent d'être étudiés. Des études menées par le National Academy of Sciences (PNAS) aux États-Unis, portant sur "Les espaces verts et le développement cognitif chez les élèves du primaire", ont mis en évidence une influence positive sur la mémoire de travail. Cela suggère qu'une attention particulière devrait être accordée à ce domaine afin de mettre en lumière les véritables bénéfices de la verdure pour les enfants. Ainsi, il devient de plus en plus crucial d'explorer comment les espaces verts peuvent contribuer non seulement aux performances académiques, mais aussi au bien-être général des élèves, à leur développement cognitif, et à leur capacité à gérer le stress et à maintenir leur concentration. Ces découvertes pourraient ouvrir la voie à des approches éducatives plus holistiques, mettant en avant le rôle de la nature dans l'épanouissement des enfants.

# Bibliographie

1. Manuel de Géographie quantitative (Cursus) : Thierry Feuillet, Etienne Cossart, Hadrien Commenges, Janvier 2019 Armand Colin.
2. Géographie de la végétation terrestre : Frédéric Alexandre, Alain Génin ; 2012 Collection U Armand Colin.
3. La végétation dans le tissu urbain : enjeux et conséquences, Gautier Forgerit, 14 Sep 2022
4. Pourquoi et comment végétaliser sa ville : Laura Paul | Nov 6, 2020 | Événement, Green, Nature, Végétalisation
5. Pourquoi les jeunes se mettent au vert ? : Laura Paul | Oct 1, 2020 | Événement, Green, Végétalisation, Workspaces
6. Enfants et Nature : Investigations Psychologiques, Socioculturelles et Évolutives. Peter H. Kahn, Jr. et Stephen R. Kellert. Cambridge, Massachusetts : The MIT Press. 2002.
7. Peter Verbeek & Frans B.M. de Waal. La relation des primates avec la nature : la biophilie comme un schéma général
8. Judith H. Heerwager & Gordon H. Orians. Le monde écologique des enfants
9. John D. Coley, Gregg E.A. Solomon, & Patrick Shafto. Le développement de la folkbiologie : une perspective des sciences cognitives sur la compréhension des enfants du monde biologique
10. Peter H. Kahn, Jr. Les affiliations des enfants avec la nature : structure, développement, et le problème de l'amnésie environnementale générationnelle
11. Stephen R. Kellert. L'expérience de la nature : le développement affectif, cognitif, et évaluatif chez les enfants
12. Olin Eugene Myers, Jr. & Carol D. Saunders. Les animaux comme liens pour développer des relations bienveillantes avec le monde naturel
13. Aaron Katcher. Les animaux dans l'éducation thérapeutique : des guides vers l'état liminal
14. Louise Chawla. Moments privilégiés dans le temps : les multiples façons d'être dans la nature pendant l'enfance
15. Rachel Kaplan & Stephen Kaplan. Les adolescents et l'environnement naturel : une pause ?
16. Cynthia Thomashow. Les adolescents et l'identité écologique : s'attacher à la nature sauvage
17. David W. Orr. L'économie politique et l'écologie de l'enfance.
18. Robert Michael Pyle. Le paradis dans un terrain vague : des endroits spéciaux, des espèces et des enfants dans le voisinage de la vie.
19. GGPlot2 Essentials for Great Data Visualization in R : Alboukadel Kassambara,
20. La régression linéaire simple avec le logiciel R : Delladata

## Liste des figures

Figure 1 : connexion d'une base de données PostgreSQL à Qgis .....	13
Figure 2 : comparaison entre végétation autour des établissements privés et publics .....	23
Figure 3 : comparaison de la végétation autour des établissements privés et publics en fonction des types d'unités urbaines.....	25
Figure 4 : portion d'établissements publics et privés en fonction des unités urbaines .....	27
Figure 5: Moyenne départementale de végétation autour des établissements scolaires.....	29
Figure 6 : Moyenne départementale de précipitation au tour des établissements scolaires .....	31
Figure 7: Moyenne départementale de températures autour des établissements.....	31
Figure 8: Moyenne de végétalisation des alentours d'écoles en fonction de leur nature .....	33
Figure 9: Matrices cartésienne de variables liées à la valeur ajoutée de réussite.....	37
Figure 10: Régression multiple de la variable de valeur ajoutée de réussite .....	40
Figure 11: Matrice cartésienne de variables liées aux résultats du brevet. ....	43
Figure 12: Régression multiple de la variable résultat du brevet.....	44

## Liste des tableaux

Tableau 1 : statistiques univariées de la végétation en fonction du secteur d'activité .....	24
Tableau 2 : régression de la valeur ajoutée de réussite des établissements scolaires .....	40
Tableau 3: Table de régression de la variable résultat au brevet.....	45
Tableau 4: coefficients de régression en fonction des unité urbaines .....	47

## Annexes

### Annexe 1 : Liste des packages utilisés dans l'environnement R

```
#Installation des packages
install.packages("sf")           # Installation du package sf
install.packages("tmap")         # Installation du package tmap
install.packages("httr")         # Installation du package httr
install.packages("leaflet")      # Installation du package leaflet
install.packages("dplyr")        # Installation du package dplyr
install.packages("terra")        # Installation du package terra
install.packages("ggplot2")      # Installation du package ggplot2
install.packages("readxl")       # Installation du package readxl
install.packages("tidyverse")    # Installation du package tidyverse
install.packages("ggpubr")       # Installation du package ggpubr
install.packages("rstatix")      # Installation du package rstatix
install.packages("RColorBrewer") # Installation du package RColorBrewer
install.packages("car")          # Installation du package car
install.packages("FactoMineR")   # Installation du package FactoMineR
install.packages("lwgeom")       # Installation du package lwgeom
install.packages("gtools")       # Installation du package gtools
install.packages("corrplot")     # Installation du package corrplot
install.packages("gtsummary")    # Installation du package gtsummary
install.packages("GGally")       # Installation du package GGally
install.packages("ggstatsplot")  # Installation du package ggstatsplot
install.packages("sjPlot")       # Installation du package sjPlot
install.packages("Hmisc")        # Installation du package Hmisc
install.packages("kableExtra")   # Installation du package kableExtra

#Chargement des packages dans L'environnement

library(sf)           # Manipulation et analyse de données géospatiales
library(tmap)         # Cartographie interactive
library(httr)         # Requêtes HTTP
library(leaflet)      # Création de cartes interactives
library(dplyr)        # Manipulation de données
library(terra)        # Analyse géospatiale avancée
library(ggplot2)      # Visualisation de données
library(readxl)       # Lecture de fichiers Excel
library(tidyverse)    # Suite de packages pour la science des données
library(ggpubr)       # Graphiques avancés avec ggplot2
library(rstatix)      # Outils statistiques
library(RColorBrewer) # Palettes de couleurs
library(car)          # Régression linéaire et diagnostics
library(FactoMineR)   # Analyse factorielle des correspondances
library(lwgeom)       # Géométrie spatiale
library(gtools)       # Outils généraux pour la manipulation de données
library(corrplot)     # Matrices de corrélation
library(gtsummary)    # Résumé de tableaux pour la statistique
library(GGally)       # Paires de graphiques avec ggplot2
library(ggstatsplot)  # Graphiques de statistiques avancées
library(sjPlot)       # Graphiques pour la régression linéaire
library(Hmisc)        # Outils de statistiques avancées
library(kableExtra)   # Personnalisation de tableaux
```

## Annexe 2 : Fonction de calcul de densité autour d'un point

```
dens_feature <- function(obs, feat, span, id) {  
  
  # Transformation des observations (obs) dans le système de coordonnées  
  # de feat  
  obs <- obs %>% st_transform(st_crs(feat))  
  
  # Extraire l'attribut id de la couche obs  
  obsId <- obs[[id]] %>% st_drop_geometry() %>% as.data.frame()  
  colnames(obsId) <- c(id)  
  result_list <- list() # Créer une liste pour stocker les résultats  
  
  for (d in span) {  
    # Créer un tampon (buffer) autour des observations à la distance d  
    buff <- st_buffer(obs, dist = d)  
  
    # Trouver l'intersection entre les tampons et la couche feat  
    tmp <- st_intersection(feat, buff)  
  
    # Calculer l'aire des zones intersectées  
    tmp$area <- st_area(tmp)  
  
    # Agréger les valeurs de l'attribut id par somme des aires pour chaque  
    # distance d  
    tmpAgr <- tmp %>%  
      group_by(!rlang::sym(id)) %>%  
      summarise(!paste0("veg_", d) := sum(area)) %>%  
      st_drop_geometry() %>%  
      as.data.frame()  
  
    # Joindre les résultats agrégés avec les observations par l'attribut  
    # id  
    obsId <- left_join(obsId, tmpAgr, by = id)  
  
    # Remplacer les valeurs manquantes par zéro  
    obsId <- obsId %>% mutate(across(everything(), ~ ifelse(is.na(.), 0,  
    .))) %>% as_tibble()  
  
    # Nommer l'objet de sortie avec le préfixe "dens_" et la distance d  
    output <- paste0("dens_", d)  
    result_list[[output]] <- obsId # Ajouter les résultats à la liste  
  }  
  
  # Retourner le dernier résultat de la liste (celui avec la plus grande  
  # distance)  
  return(result_list[[length(result_list)]])  
}
```



### Annexe 3 : Boucle de calcul de la végétation autour des écoles

```
# Vecteur de distances à utiliser pour le calcul de densité
d <- c(100,200,500,1000)
dens <- data.frame()

# Boucle pour traiter chaque ligne de la couche "etab"
for (i in 1:nrow(etab)) {
  tryCatch({
    print(i)

    # Sélection de la ligne i de la couche "etab"
    tmp_etab <- etab[i,] %>% st_transform(2154)

    # Crée une représentation en texte (WKT) d'une zone tampon autour de
    "tmp_etab"
    wkt <- st_as_text(st_geometry(st_buffer(tmp_etab, 1100) %>% st_transf
orm(4326)))

    # Charge les données végétales à l'intérieur de la zone tampon
    tmp_veg <- st_read("data/veg.gpkg", wkt_filter = wkt) %>%
      st_simplify() %>%
      st_union() %>%
      st_sf()

    # Calcul de la densité des caractéristiques végétales autour de "tmp_
    etab"
    tmp_dens <- dens_feature(tmp_etab, tmp_veg, d, "numero_uai")

    # Ajoutez le résultat à la liste
    dens_list[[i]] <- tmp_dens

    # Affichez un message pour indiquer la progression
    cat("Traitement de l'établissement", i, "terminé.\n")
  }, error = function(e) {
    # En cas d'erreur, affichez un message d'erreur et continuez avec l'i
    tération suivante
    cat("Erreur lors du traitement de l'établissement", i, ":", condition
Message(e), "\n")
  })
}
```

#### Annexe 4 : Traitement des rasters (cas des rasters ; méthode similaire pour les deux métriques)

```
# Importation successive des douze rasters de température pour tous les m
ois de l'année
for (i in 1:12) {
  file_name <- sprintf("tavg/%d.tif", i)
  obj_nm <- sprintf("tavg_%d", i)
  assign(obj_nm, (rast(file_name)))
}

# Empilement des rasters dans un même vecteur
tavg <- c(tavg_1, tavg_2, tavg_3, tavg_4, tavg_5, tavg_6, tavg_7, tavg_8,
  tavg_9, tavg_10, tavg_11, tavg_12)

# Moyenne des températures et intersection avec l'emprise de La France
tavg <- terra::mean(tavg) %>% crop(st_bbox(fr)) %>% project("epsg:2154")

# Définition du type de la matrice pour le calcul de la moyenne dans un r
ayon donné
typ <- "circle"

# Création d'un raster de dimensions 1000m avec le même système de coordo
nnées que le raster "tavg"
rst_prec <- rast(ext(prec), crs = crs(prec), resolution = 1000)

# Calcul de la matrice focale (focalMat) pour les températures moyennes
mat <- focalMat(tavg, d = d, type = typ)

# Calcul des moyennes dans un rayon de 1000m autour de chaque cellule
dens_tavg <- focal(x = tavg, w = mat, fun = mean)

# Extraction des valeurs de température moyenne pour les emplacements des
écoles (ate_point)
tavg_dens <- extract(tavg, etab, bind = TRUE) %>% terra::as.data.frame()
```

## Annexe 5 : Méthodologie de régression

```
# Fusionnez les tables de données "etab", "typo", et "ns" en utilisant des jointures left_join
res_etab <- etab %>%
  left_join(typo,
            by = c("code_commune" = "Code_Commune")
  ) %>%
  left_join(ns,
            by = c("numero_uai" = "UAI")
  )

##Va
# Effectuez une jointure entre la table "VA" et une sélection de colonnes de "res_etab"
VA <- VA %>% left_join(select(res_etab,
                             numero_uai,
                             code_commune,
                             veg_1000,
                             Temperature_moy,
                             Precipitation_moy,
                             Type_d_unite_urbaine,
                             IPS
                           ),
                      by = c("UAI" = "numero_uai")
)

# Supprimez les colonnes inutiles et les lignes avec des valeurs manquantes
va <- VA[, -(c(1,3))] %>% na.omit()

# Créez une matrice de graphiques cartésiens pour explorer les relations entre les variables
scatterplotMatrix(
  ~VA + veg_1000 + Type_d_unite_urbaine + IPS,
  data = va,
  smoother = loessLine,
  main = "Matrice de graphiques cartésiens de valeurs ajoutées liées aux taux de réussite des lycées et collèges",
  col = "#83AB05",
  pch = 16,
  cex = 0.2,
  grid = TRUE,
  diagonal = "histogram",
  legend.main = TRUE,
  legend.pos = "topright",
  smooth.method = "loess",
  var.labels = c("VA", "veg_1000", "Type urbain", "IPS"),
  hist.col = "lightblue",
  density = list(col = "pink"),
  span = 0.5
)

# Effectuez une régression linéaire multiple
va_lm <- lm(data = va,
            formula = VA ~ veg_1000 +
```

```

    Temperature_moy +
    Precipitation_moy +
    Type_d_unite_urbaine +
    IPS)

# Affichez un résumé des résultats de la régression
summary(va_lm)

# Créez une table de régression formatée
tbl_regression(va_lm) %>% add_vif()

# Supposons que vous ayez déjà généré votre table formatée
table_formatted <- tbl_regression(va_lm) %>%
  add_vif() %>%
  kbl()

# Personnalisez la table
table_formatted <- table_formatted %>%
  kable_styling(full_width = FALSE) %>%
  row_spec(0, bold = TRUE) %>%
  column_spec(1:3, bold = TRUE)

# Affichez la table
table_formatted

# Ajoutez un titre à la table
table_formatted <- table_formatted %>%
  title("Table de régression linéaire de la variable dépendante VA (valeur ajoutée)")

# Utilisez ggcoefstats pour visualiser les coefficients de la régression
graph <- ggcoefstats(va_lm, exclude.intercept = TRUE)

# Personnalisez le graphique
graph +
  theme_minimal() +
  labs(
    title = "Coefficients de Régression Linéaire Multiple pour la valeur ajoutée",
    x = "Coefficient",
    y = "Variables explicatives"
  ) +
  theme(
    plot.title = element_text(size = 14, face = "bold"),
    axis.text.x = element_text(angle = 45, hjust = 1),
    axis.title = element_text(size = 12)
  )

```

## **Annexe 6 : La nature et les enfants ( P.Kahn)**

***Children and Nature: Psychological, Sociocultural and Evolutionary Investigations.* Edited by Peter H. Kahn, Jr. and Stephen R. Kellert. xix + 348 pp. Cambridge, Massachusetts: The MIT Press. 2002. Includes name and subject indexes.**

**Reviewed by Kathryn C. McElveen, Department of Anthropology, University of Arizona.** arizona-kay@yahoo.com

Though the backgrounds of the contributors to this edited volume are diverse, all of the authors strive to support one thesis – that there is value in children being allowed to play in the dirt. The authors defend this thesis with approaches ranging from the poetic-magical, drawing on the importance of the magic and the mythic in developing memories of nature that endure into adulthood (Chawla, Ch.8) to the rigorously scientific, looking at children's contact with nature through direct, indirect, vicarious and symbolic means and inferring how this impacts children's emotional, intellectual and values-related development (Kellert, Ch. 5). Most authors in this collection, however, seek a balance between qualitative and quantitative data in their explorations of children's ecological experiences. Regardless of the field of the authors' expertise, research on the topic of children's interaction with nature seems somewhat limited. This volume attempts to collect what is known, and bring together a diverse group of thinkers to put together the pieces already available in the literature to support their inferences about the value of children developing a relationship with the natural world. A few of the authors do conduct research on the direct effects of interaction between children and the natural world, but these studies are more limited in scope and often focus upon the therapeutic functions of such interactions (Katcher, Ch. 7) or children's experience of the natural world in a more constrained setting such as a zoo (Myers, Jr. and Saunders, Ch. 6).

The first two chapters of the volume have a particularly strong evolutionary component, as Verbeek and de Waal (Ch. 1) first examine the primate relationship with nature, introducing the key term biophilia, "an innate tendency to affiliate with natural things" (p. 1). Their contribution is followed up Heerwager and Orians (Ch. 2), who explore how children's experience of the natural world could have shaped their survival in the past.

Several of the authors display a strong cognitive orientation, exploring the cognitive foundations of biological understanding (Coley, Solomon & Shafto, Ch. 3) and subsequently, the way interaction with the environment builds a structural framework of concepts and values in children (Kahn, Jr. Ch. 4). This last chapter may be of particular interest to anthropologists as it compares children's construction of concepts and values through environmental interaction across cultures. Also of anthropological interest may be Katcher's contribution (Ch. 7), as he draws on Victor Turner's ideas of liminality and communitas to explain why children in residential treatment facilities for behavior problems behave differently in the presence of animals.

Two essays specifically address the adolescent and nature, with the first (Kaplan & Kaplan, Ch.9) exploring the thesis that adolescents may take a "time-out" from appreciating and enjoying nature. They ultimately support this thesis, citing adolescents' penchant for social activity over solitary reflection, an activity more often associated with natural spaces. Thomashow (Ch. 10) supports the opposite contention, however, showing that adolescents do remain engaged with nature, particularly if their educational setting makes a hands-on project related to nature part of the curriculum.

For those interested in political ecology, the final chapters of this volume may be the most interesting. Orr's contribution (Ch. 11), entitled Political economy and the ecology of childhood, moves beyond lamenting that children are separated from nature and asks why this is the case. He addresses the main features of the world economy, focusing on how economic growth, material accumulation, development and commodification shape children's view of the world. Orr suggests that in today's world, children are isolated from nature, and deprived of a slower pace of life. Instead, children are exposed to virtual rather than actual reality, more violence, and a faster pace of life. His solution to this cycle of movement away from nature is to think about making the world a more child-centered place. He asks, "what would a child-centered political economy look like – how would we act if we were considering the welfare of unborn generations?" His contribution to the volume melds nicely with Pyle's reflections (Ch. 12) on "Eden in a vacant lot," calling for the maintenance of open spaces, and suggesting that butterflies may be an excellent index animal for judging when environments have become too "built-up."

This volume is a good starting point for anyone interested in children's interaction with the natural world. Because the authors are pulling together existing literature for the most part rather than reporting original research, the bibliographies following each essay typically provide a large number of useful references. The book as a whole is also valuable because it brings together authors of diverse academic backgrounds. These authors refrain from using jargon from their fields of specialization, instead focusing on the big picture and providing fresh perspectives to reader-researchers who seek to explore further the main point of the volume *Children and Nature*; that there are psychological, social, and possibly even survival benefits that accrue to children from being allowed to play in the dirt.

Journal of Political Ecology      Vol. 11 2004

## Annexe 7 : Les aires Terrestres Educatives

Dans le cadre de mon stage, j'ai également caractérisé le contexte paysager des ATEs (Aires Terrestres Éducatives). Mes encadrants sont en train d'effectuer des relevés dans les établissements scolaires associés à des ATEs mais ces données ne seront pas disponibles avant la fin de mon stage mais seront exploitées par les chercheurs dans les années à venir.

Le projet des Aires Terrestres Éducatives (ATEs) porté par l'OFB (Office Français de la Biodiversité) a pour objectif de sensibiliser la jeunesse aux bienfaits, à l'importance et à la préservation de la nature. Il se matérialise sous la forme de petits espaces verts dont la gestion est participative, impliquant à la fois les élèves et leurs enseignants, dans le but de préserver ces espaces. Au travers d'un "conseil des enfants", ce sont les élèves eux-mêmes qui se concertent et prennent toutes les décisions liées à la gestion de leur environnement. Cette expérience leur permet de découvrir leur territoire, d'appréhender les acteurs locaux, de développer des compétences transdisciplinaires, et de renforcer leur lien avec la nature. Dans le cadre du projet des ATEs, une meilleure compréhension du contexte paysager des écoles revêt une importance capitale. Cela englobe à la fois les éléments naturels et artificiels qui composent l'environnement quotidien des élèves, tels que les espaces verts, les cours de récréation, les bâtiments environnants, les rues adjacentes, etc. Parmi ces éléments, les composantes naturelles, notamment la végétation, jouent actuellement un rôle d'autant plus déterminant. En encourageant la gestion participative de ces espaces, le projet des ATEs favorise une véritable connexion entre les élèves et leur environnement. Il leur permet de comprendre la valeur des éléments naturels qui les entourent et de participer activement à leur préservation. De plus, cette démarche éducative transmet des compétences précieuses aux élèves, allant au-delà de la simple acquisition de connaissances académiques, et les prépare à devenir des citoyens responsables et conscients de l'importance de la nature dans leur vie quotidienne. Ce projet incarne une approche novatrice de l'éducation en mettant en avant le rôle essentiel de la nature dans le développement des jeunes tout en les impliquant activement dans la préservation du contexte paysager de leur établissement.

Le projet des Aires Terrestres Éducatives (ATE), porté par l'Office Français de la Biodiversité (OFB), est une initiative précieuse pour sensibiliser les jeunes générations à l'environnement et à la biodiversité tout en favorisant l'apprentissage pratique et l'engagement communautaire. Voici quelques points clés qui illustrent les objectifs et les avantages de ce projet :

**Sensibilisation à l'environnement** : Les ATE visent à sensibiliser les élèves dès leur plus jeune âge à l'importance de la nature et de la biodiversité. En développant un attachement à la nature dès l'enfance, le projet vise à créer une génération future d'éco-responsables qui prendront soin de leur environnement. Cette sensibilisation précoce est essentielle pour encourager des comportements respectueux de la nature.

**Apprentissage pratique** : Les ATE offrent une expérience d'apprentissage pratique. Les élèves interagissent directement avec la nature en observant la faune et la flore locales, en collectant des données scientifiques et en participant à des projets de conservation et de restauration écologique. Cela leur permet d'acquérir des compétences pratiques en environnement, mais aussi en travail d'équipe et en mise en place de stratégies participatives.

**Interdisciplinarité** : Les ATE favorisent une approche interdisciplinaire de l'éducation en intégrant des aspects scientifiques, environnementaux, artistiques, culturels et sociaux. Cette approche permet aux élèves d'acquérir des compétences variées et transdisciplinaires, ce qui est essentiel pour comprendre la complexité des enjeux environnementaux.

**Engagement communautaire** : Les ATE encouragent l'engagement communautaire en impliquant les familles, les enseignants, les experts locaux et les organismes environnementaux dans la création et la gestion des

aires éducatives. Les élèves apprennent ainsi à travailler en équipe, à comprendre l'importance de la collaboration et à reconnaître les différents acteurs impliqués dans la protection de l'environnement. Cela renforce les liens entre l'école et la communauté et favorise une prise de conscience collective de l'importance de l'environnement.

Les Aires Terrestres Éducatives répondent à plusieurs enjeux essentiels : l'éducation environnementale, l'acquisition de compétences pratiques, l'approche interdisciplinaire et l'engagement communautaire. Elles contribuent à former une génération consciente à la valeur de la biodiversité et à la nécessité de protéger notre environnement pour un avenir durable.



# Table des matières

<b>Remerciements</b>	1
<b>Introduction</b>	2
<b>I. STRUCTURE D'ACCEUIL</b>	4
1. Le laboratoire UMR CNRS IDEES	4
a) Les Axes de recherche IDEES Caen	5
b) Le Tuteur de Stage	6
2. Le Laboratoire SETE	6
<b>II. NOTRE TRAVAIL</b>	8
<b>III. CONSTRUCTION DES DONNEES</b>	9
1. Les données utilisées	9
• Adresse et géolocalisation des établissements d'enseignement du premier et second degré	9
• Zones de végétation BD topo	9
• Les précipitations et températures	10
• Diplôme national du brevet par établissement	10
2. Tutoriel d'utilisation de la BD topo SQL France entière dans R	11
3. Densité de végétation autour des écoles	14
• Fonction de calcul des superficies	15
• La fonction prend quatre arguments en entrée :	15
• Calcul des densités de végétation	16
4. Calcul des moyennes de températures et de précipitation	17
a) La fonction extract	17
b) Description du code	18
• Importation des rasters de température ou de précipitation pour chaque mois de l'année :	18
• Empilement des rasters dans un vecteur :	18
• Calcul de la moyenne des températures et intersection avec l'emprise de la France :	18
• Définition du type de la matrice pour le calcul de la moyenne dans un rayon donné :	19
• Création d'un nouveau raster de dimensions 1000m :	19
• Calcul de la matrice focale (focalMat) pour les températures moyennes :	19
• Calcul des moyennes dans un rayon de 1000m autour de chaque cellule :	19
• Extraction des valeurs de température moyenne pour les emplacements des écoles :	19
<b>IV. ANALYSE</b>	21
1. Caractéristiques des établissements et végétation	21
a) Méthodologie	22
b) Test statistique	22

c) Résultats .....	23
• Secteur public/privé et végétation .....	23
• Relation entre végétation et type urbain des communes .....	25
• Taux de végétalisation autour des écoles par départements.....	29
• Températures et précipitations moyennes autours des établissements scolaires .....	30
• Nature des établissements et végétation .....	33
2. Résultats scolaires et végétation .....	35
a) Valeur ajoutée et végétation .....	36
b) Résultats du brevet et végétation .....	42
c) Régression par catégorie d'unités urbaines .....	47
Conclusion .....	49
Bibliographie.....	51
Liste des figures .....	52
Liste des tableaux .....	53
Annexes.....	54
Annexe 1 : Liste des packages utilisés dans l'environnement R.....	54
Annexe 2 : Fonction de calcul de densité autour d'un point.....	55
Annexe 3 : Boucle de calcul de la végétation autour des écoles .....	56
Annexe 4 : Traitement des rasters (cas des rasters ; méthode similaire pour les deux métriques) ...	57
Annexe 5 : Méthodologie de régression .....	58
Annexe 6 : La nature et les enfants ( P.Kahn) .....	60
Annexe 7 : Les aires Terrestres Educatives .....	62

