

Master 1 - Économétrie Appliquée

Analyse de données et descriptive

Analyse de l'attractivité touristique
des départements en France
métropolitaine

Marie KERHOAS & Gloria Isabel PALACIO

Novembre 2023

Sommaire

I.	Introduction	3
II.	Analyse détaillée du sujet.....	4
III.	Analyse en Composantes Principales ACP.....	22
IV.	Corrélation entre la variable illustrative et les variables latentes	38
V.	Conclusion et discussion des résultats	47
VI.	Annexes	50
VII.	Bibliographie.....	69
VIII.	Table des matières.....	73

I. Introduction

Au regard du contexte global dans lequel nous évoluons : changement climatique, instabilité financière, appréhensions sanitaires... , le secteur du tourisme est aujourd'hui amené à se renouveler. Face à ces enjeux, il devient essentiel pour ce secteur d'évoluer vers des pratiques plus réfléchies et pérennes. Cette transition embrasse à la fois la soif de découverte et l'impératif de préserver notre patrimoine, tout en reconnaissant la fragilité des attraits qui le composent.

Dans un monde interconnecté, le tourisme ne se limite pas à un simple déplacement, il représente un pont entre les cultures et une composante vitale de l'économie. Selon les données de l'Organisation Mondiale du Tourisme, le secteur contribue à près de 10 % du PIB mondial et soutient un emploi sur dix dans le monde ¹. Ces chiffres témoignent de son impact colossal sur l'économie.

En 2022, année marquant une période de reprise post-Covid-19, la France s'est affirmée une fois de plus en tant que première destination touristique mondiale. Concrètement, cette place de leader s'est traduite par l'accueil de 79,4 millions de touristes internationaux, et 57,9 milliards d'€ de recettes².

L'attrait touristique de la France, apparaît donc comme indéniable. C'est pourquoi, dans le cadre de ce projet, nous avons choisi de décrypter cette attractivité, en nous focalisant sur les départements métropolitains. En effet, les départements d'outre-mer, en raison notamment de leur éloignement géographique, ont des caractéristiques distinctes qui méritent d'être étudiées spécifiquement.

Notre démarche reflète notre volonté de nous immerger dans la richesse des territoires, analysant chaque département pour en saisir les nuances et les particularités. Ainsi, notre objectif est de mettre en lumière les facteurs clé qui influencent l'attractivité touristique des différents départements.

¹<https://www.vie-publique.fr/eclairage/24088-le-surtourisme-quel-impact-sur-les-villes-et-sur-lenvironnement> (consulté le 22/10/2023)

²<https://www.gouvernement.fr/actualite/la-france-toujours-premiere-destination-touristique-au-monde> (consulté le 26/09/2023)

II. Analyse détaillée du sujet

A. Pertinence de la variable à expliquer

Afin de mesurer l'attractivité touristique des départements français, nous avons choisi comme variable à expliquer la fréquentation des hébergements collectifs touristiques en 2022³. Cette dernière s'exprime en nombre de nuitées et comprend à la fois la fréquentation des hôtels de tourisme ainsi que celle des campings.

A l'échelle des départements français, il s'agit de la seule variable mesurant la venue des touristes répertoriée par l'INSEE. D'autres statistiques étaient communiquées comme le nombre de chambres d'hôtel, d'hébergements collectifs ou encore d'emplacements de camping existants, mais ne prenaient pas en compte la présence ou non de touristes dans ces établissements.

Par ailleurs, il est important de savoir que cette variable ne comptabilise que de façon partielle (23 %) l'ensemble des nuitées touristiques de l'année 2022.

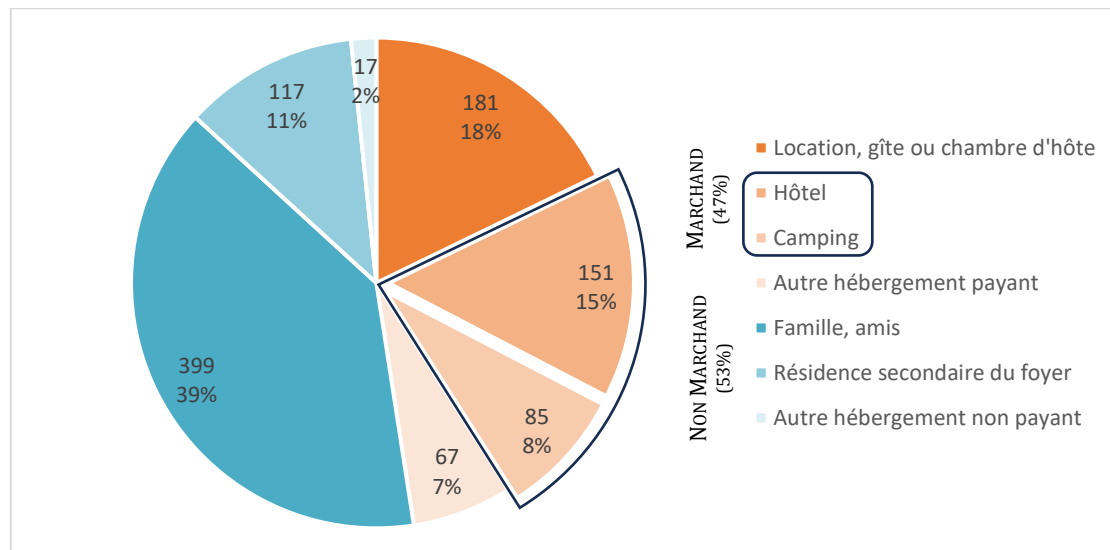


Figure 1 - Répartition du nombre de nuitées touristiques en 2022 selon les types d'hébergements (en millions)

Source : INSEE : <https://www.insee.fr/fr/statistiques/7653005#graphique-figure4> (consulté le 20/09/2023)

³ https://www.insee.fr/fr/statistiques/2012672#tableau-TCRD_020_tab1_regions2016/ (consulté le 20/09/2023)

Malgré cette restriction, la fréquentation des hébergements collectifs touristiques reste, selon nos recherches, la variable la plus pertinente à exploiter dans le cadre de notre analyse de l'attractivité touristique.

Sur un plan géographique, notre variable explicative se présente de la façon suivante :

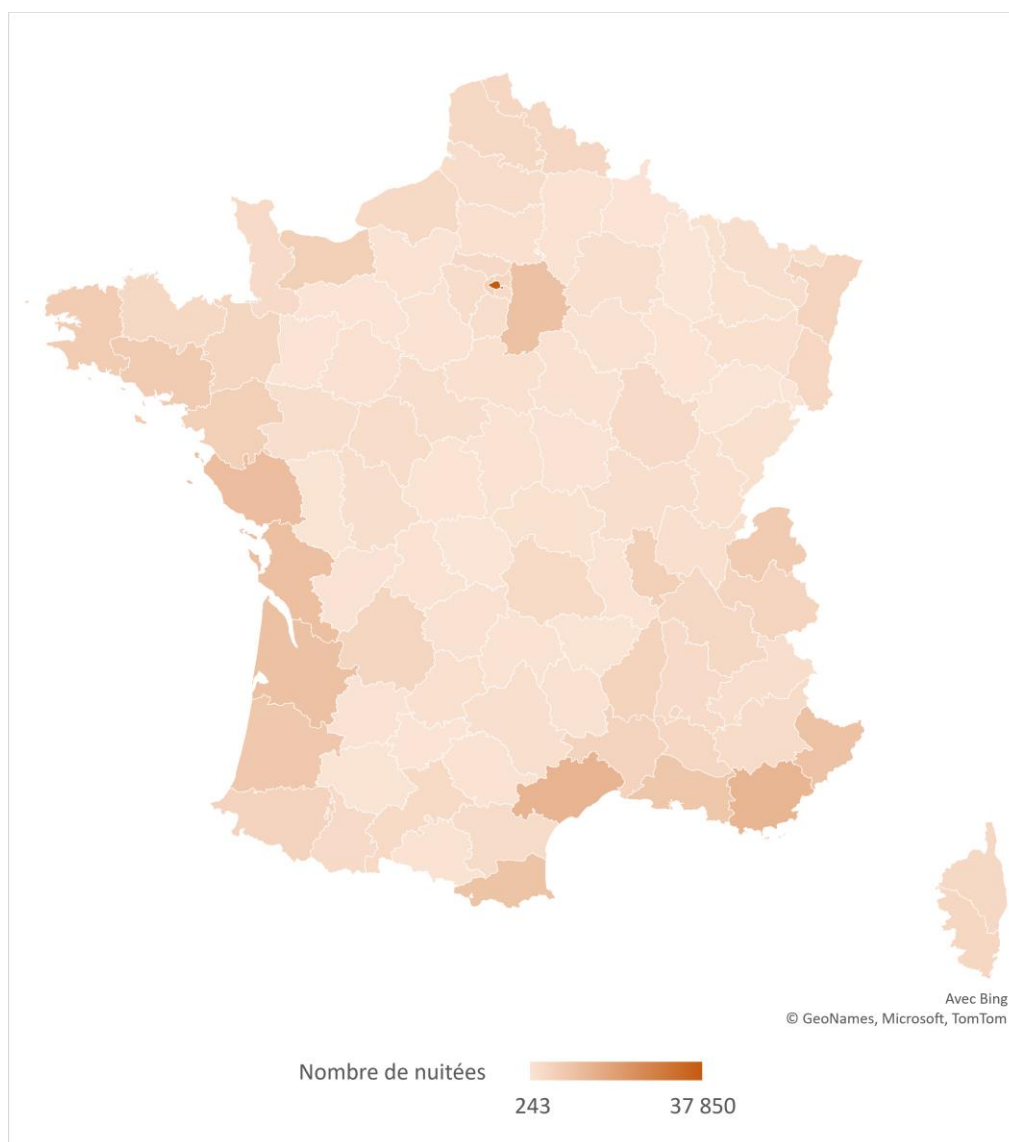


Figure 2 - La fréquentation des hébergements collectifs touristiques en nombre de nuitées sur 2022 selon les départements de France métropolitaine

Ce graphique, réalisé avec Excel, met en avant la singularité du département de Paris : à lui seul il représente 37 850 000 nuitées répertoriées. Cependant, même s'il se détache du lot, il n'a pas été question d'enlever cette observation de nos données. En effet, nous travaillons à l'échelle de la France métropolitaine, il est donc important d'en conserver chaque département.

B. Rôle des variables explicatives et corrélation entre la variable illustrative et les variables actives

1. Les commerces

Selon l'INSEE, le nombre de commerces représente le total des équipements pour chaque département (en 2021) tel que : supermarchés, épicerie, magasins de vêtements, etc. Ces données sont regroupées sous l'appellation "Base permanente des équipements (BPE)" et proviennent de la source SIRUS produite par l'INSEE⁴.

Nous avons opté pour la variable "nombre de commerces" car elle donne une indication claire de l'activité commerciale et de l'infrastructure économique. Une forte présence commerciale peut signifier un afflux régulier de résidents et de visiteurs, ce qui peut influencer directement le nombre de nuitées dans le département. De plus, un département avec une variété d'options commerciales peut être perçu comme plus attractive pour les touristes, offrant à la fois des commodités et des possibilités de shopping.

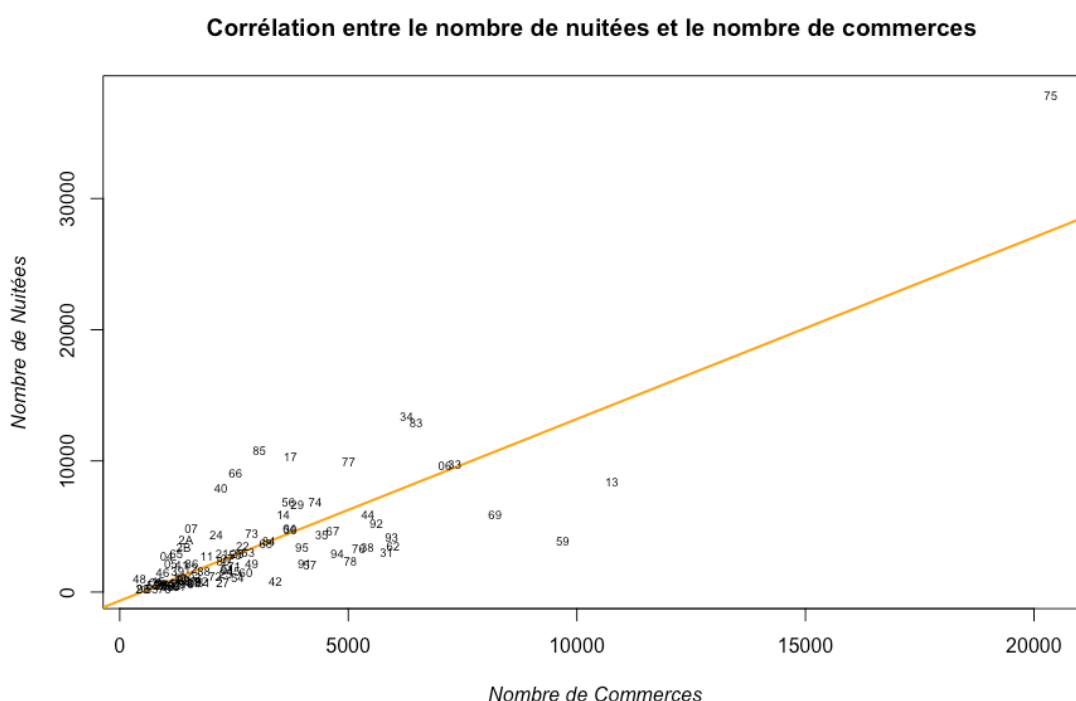


Figure 3 - Corrélation entre le nombre de nuitées et le nombre de commerces : $r = 0,8177$

⁴ <https://www.insee.fr/fr/statistiques/3568602?sommaire=3568656#dictionnaire> (consulté le 01/10/2023)

Une corrélation positive est observée entre ces deux variables, comme le souligne le coefficient de corrélation $r = 0,81$. Cette donnée suggère qu'une augmentation du nombre de commerces coïncide avec une hausse des nuitées dans les hôtels et campings.

Un nombre croissant de commerces, tels que restaurants, boutiques de souvenirs, centres de loisirs, etc., peut améliorer l'expérience des touristes. Si les visiteurs ont plus d'options pour manger, faire du shopping et s'engager dans des activités de loisirs, ils sont plus susceptibles de prolonger leur séjour, ce qui entraîne une augmentation du nombre de nuitées. Ce phénomène est clairement observable dans des départements tels que le "69 - Rhône", où Lyon offre une riche variété de commerces et d'activités culturelles; le "59 - Nord", avec Lille comme pôle commercial dynamique ; le "13 - Bouches-du-Rhône" où Marseille présente une combinaison unique de marchés, de boutiques et d'attractions naturelles ; et bien sûr, le "75 - Paris", qui, en tant que capitale mondiale de la mode et de la gastronomie, attire d'innombrables touristes chaque année.

Le shopping s'avère être un élément crucial de l'attrait touristique. Dès les années 2000, il représentait déjà 6 % des déplacements touristiques, caractérisés par des visites fréquentes - à peu près une fois par trimestre - souvent motivées par les tendances saisonnières de la mode. Ces séjours consacrés au shopping sont typiquement brefs, se limitant en général à une journée, bien que certains s'étendent sur le week-end. Par ailleurs, pour 15 % des touristes français en visite urbaine, le shopping est une activité intégrale de leur expérience, venant s'ajouter aux 19 % qui privilégient les expériences gastronomiques ⁵.

Cette préférence pour le shopping et la gastronomie souligne la relation entre le commerce et le tourisme, deux secteurs qui s'influencent mutuellement de manière significative. D'une part, le commerce répond aux besoins des touristes en proposant des services et des produits indispensables, ce qui améliore considérablement leur expérience globale. D'autre part, le tourisme agit comme un moteur pour l'activité commerciale locale, créant une demande accrue pour une gamme variée de biens et de services.

⁵ https://www.persee.fr/doc/AsPDF/htn_0018-439x_2001_num_2_1_2766.pdf (consulté le 15/10/2023)

2. Les médecins généralistes

Par la suite, nous nous sommes intéressées à la relation existante entre le tourisme et la santé.

Selon Sébastien FLEURET, géographe de la santé, il existerait une corrélation positive entre ces deux domaines. En effet dans l'un de ses articles⁶, la santé apparaît comme un élément d'attractivité touristique. Cela concerne principalement une population de retraités sérénisée par un cadre de sécurité sanitaire.

Afin d'étudier ce lien dans notre analyse, nous avons choisi comme variable explicative le nombre de médecins pour 100 000 habitants.⁷

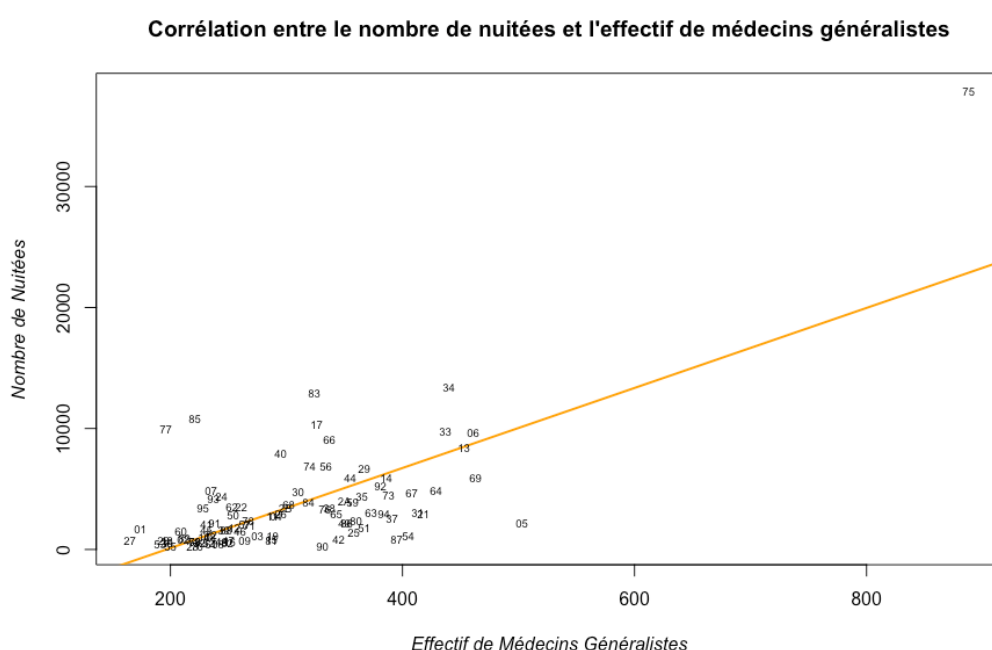


Figure 4 - Corrélation entre le nombre de nuitées et l'effectif de médecins généralistes : $r = 0,7004$

Ce graphique présente une relation linéaire positive entre le nombre de nuitées et l'effectif de médecins généralistes. Ce résultat est confirmé par le coefficient de corrélation des deux variables, il est de $r = 0,7004$. Aussi, le nombre de nuitées augmente avec l'effectif de médecins généralistes exerçants et cela va dans le sens de la littérature.

⁶ <https://socgeo.com/2017/05/19/du-tourisme-medical-a-une-geographie-des-interactions-entre-tourisme-et-sante/> (consulté le 04/10/2023)

⁷ <https://www.insee.fr/fr/statistiques/2012677/> (consulté le 01/10/2023)

3. Les équipements culturels

La variable que nous avons choisie pour mesurer “la culture” de nos départements est le nombre total d'équipements culturels, tel que défini par la base permanente des équipements sportifs, de loisirs et culturels. Cette base englobe des éléments essentiels de la vie culturelle, comme les cinémas, salles de spectacle et des espaces remarquables et du patrimoine⁸.

Nous avons choisi de nous intéresser à cette variable en reconnaissant le rôle primordial que joue le tourisme culturel dans le monde actuel. Car tout comme il existe un tourisme balnéaire, sportif ou d'affaires, il existe un tourisme dit « culturel », dont le déplacement est motivé par l'existence d'équipements culturels.

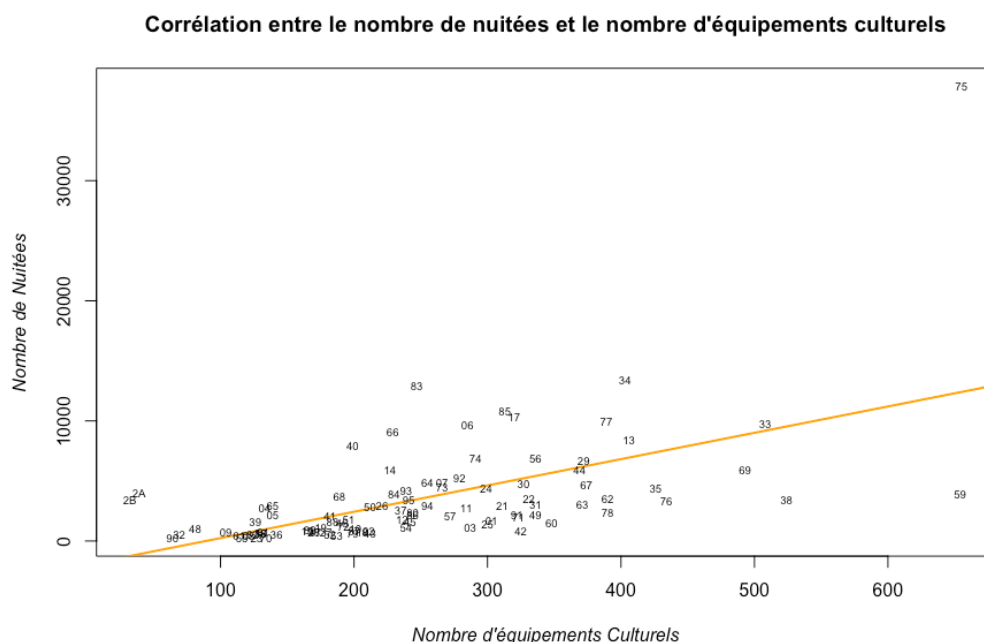


Figure 5 - Corrélation entre le nombre de nuitées et le nombre d'équipements culturels : $r = 0,5654$

Le coefficient de corrélation $r = 0.57$ signifie qu'il existe une tendance à ce que, lorsque le nombre d'équipements culturels augmente dans un département, le nombre de nuitées touristiques dans ce département ait également tendance à augmenter. Cette corrélation suggère que l'offre culturelle joue un rôle positif dans l'attrait touristique d'un département, incitant ainsi les visiteurs à passer plus de nuitées.

⁸ <https://www.insee.fr/fr/statistiques/3568599?sommaire=3568656> (consulté le 27 /09 /2023)

L'une des premières motivations des touristes étrangers qui visitent notre pays reste la culture : la richesse de son patrimoine, matériel et immatériel, la vitalité de sa création et de sa scène artistique, ses industries culturelles et créatives, sa langue et son art de vivre⁹.

Notre intérêt pour cette variable réside dans l'analyse de la contradiction entre la dimension culturelle et l'attractivité touristique à divers égards. En théorie, l'ensemble des biens culturels est ouvert au tourisme. Cependant, la réalité révèle des nuances. Toute l'offre culturelle ne se prête pas nécessairement au tourisme. Certaines activités culturelles ne suscitent que peu d'intérêt touristique (par exemple : les bibliothèques), sont incompatibles avec les pratiques touristiques, restent inaccessibles (fermeture saisonnière) ou excluent cette clientèle¹⁰.

De plus, la gestion culturelle n'est pas forcément compatible avec les intérêts économiques inhérents aux acteurs du tourisme. En effet, le secteur du tourisme est guidé principalement par des acteurs motivés par la création d'une activité commerciale pour atteindre des objectifs de rentabilité économique, et impliquant une forte proportion d'acteurs issus du secteur privé et commercial. Or, la gestion culturelle met davantage l'accent sur la préservation et la conservation des éléments représentatifs de notre culture, ce qui peut parfois justifier des restrictions en matière d'accès et d'affluence touristique.

Les points de rencontre entre l'attractivité touristique d'un département et la culture sont donc nombreux mais présentent des limites.

4. Le niveau de vie médian

Dans le cadre de nos recherches sur l'attractivité touristique, nous avons choisi le "niveau de vie médian" comme une variable explicative car il est un indicateur des ressources économiques. Calculée en divisant le revenu disponible du ménage par le nombre d'unités de consommation (UC), elle prend en compte la taille et la composition des ménages à

⁹ <https://www.culture.gouv.fr/Thematiques/Developpement-culturel/Actualites/La-culture-un-veritable-atout-pour-le-succes-touristique-de-nos-territoires/> (consulté le 12 /10 /2023)

¹⁰ https://memoires.sciencespo-lyon.fr/internet/Masters/SECI/juanchich_1/pdf/juanchich_1.pdf (consulté le 17 /10 /2023)

Et, de manière plus générale, le "niveau de vie" est une représentation du bien-être économique moyen des résidents d'un département. Il peut ainsi servir de baromètre pour évaluer la qualité des infrastructures et des services offerts. Un niveau de vie plus élevé est susceptible d'indiquer des infrastructures de meilleure qualité, une variété d'attractions touristiques plus riches et une meilleure capacité d'accueil des touristes.

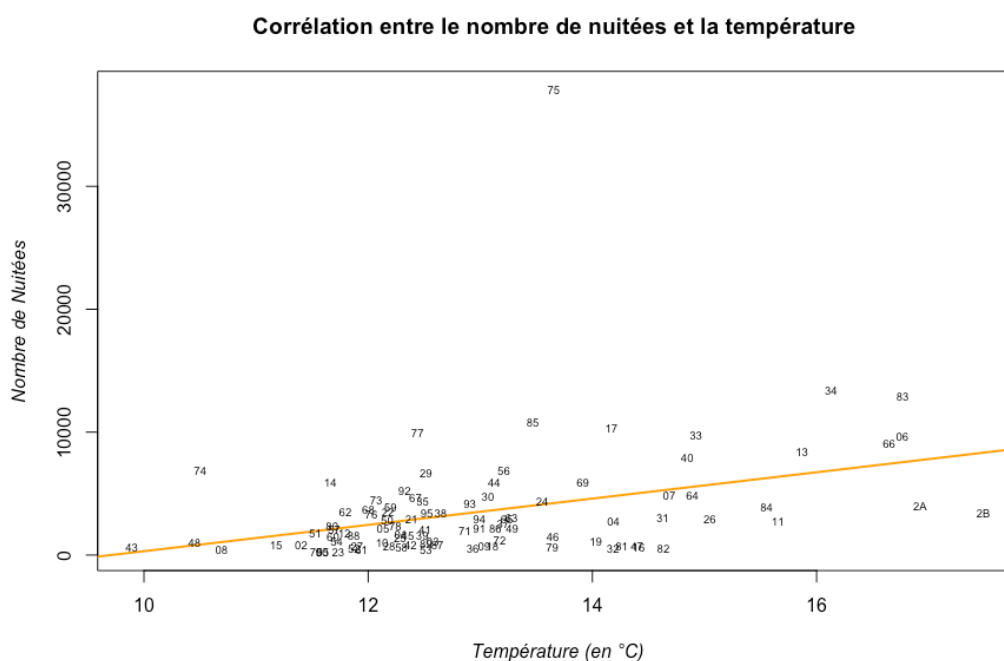


¹¹ <https://www.insee.fr/fr/statistiques/6692414:sommaire=6692394#documentation> (consulté le 27/09/2023)

5. La température

Notre variable explicative suivante s'intéresse cette fois-ci au climat et plus précisément à la température moyenne observée dans chacun des départements de France métropolitaine¹².

Comme le présentent Jean-Paul CERON et Ghislain DUBOIS dans une communication¹³ parue à la suite de leur participation à une conférence sur le tourisme et le changement climatique, la relation entre le tourisme et le climat est loin de se résumer à une unique corrélation positive. En effet, à première vue, une hausse des températures semblerait être favorable à l'essor du tourisme avec des destinations plus ensoleillées et donc plus attractives. Cependant cette modification du climat entraîne des conséquences impactant négativement le secteur du tourisme telles la diminution de la couverture neigeuse et des réserves en eau ou bien encore des chaleurs excessives. Il est important de noter que l'impact des effets négatifs s'amplifiera significativement au fil des ans.



¹² <https://www.data.gouv.fr/fr/reuses/temperature-moyenne-mensuelle-par-departement/> (consulté le 01/10/2023)

¹³ <https://www.tec-conseil.com/lab/publications/colloques-conferences-et-ateliers/tourisme-et-changement-climatique-une-relation-a-double-sens.htm?lng=fr> (consulté le 22/10/2023)

Ce graphique présente la relation observée entre notre variable à expliquer et les températures moyennes relevées. Avec un coefficient de $r = 0,35$, cette corrélation est positive. Concrètement, cela signifie que, pour le moment, les départements ayant les températures moyennes les plus importantes ont tendance à attirer les touristes à y séjourner. On constate ici que le premier aspect mis en avant par la littérature, que les touristes apprécient les lieux ensoleillés, est davantage pris en compte que le second actuellement.

6. Les équipements sportifs

Les équipements sportifs font référence aux infrastructures et installations conçues spécifiquement pour la pratique d'activités sportives. Ces installations englobent une variété d'espaces tels que les courts de tennis, les piscines, les gymnases et les pistes d'athlétisme¹⁴.

Dans le contexte de notre projet, nous donnons de l'importance de ces équipements car ils ne sont pas seulement des lieux de loisirs pour les résidents locaux, mais aussi des attractions majeures pour les touristes, contribuant ainsi au dynamisme économique et social d'un département.

Nous tenons à préciser qu'il y a diverses formes de tourisme sportif. Nous en identifions principalement quatre : le tourisme sportif de spectacle, celui de visite, d'action, et le tourisme de séjour/découverte axé sur un sport¹⁵. Dans le cadre de notre analyse, lorsque nous évoquons "les équipements sportifs", nous faisons spécifiquement référence au tourisme sportif de type séjour/découverte centré sur une discipline sportive.

¹⁴ <https://www.insee.fr/fr/statistiques/3568599?sommaire=3568656#dictionnaire> (consulté le 27 /09 /2023)

¹⁵ <https://www.sport-et-tourisme.fr/corporate/essor-du-tourisme-sportif/> (consulté le 21 /10 /2023)

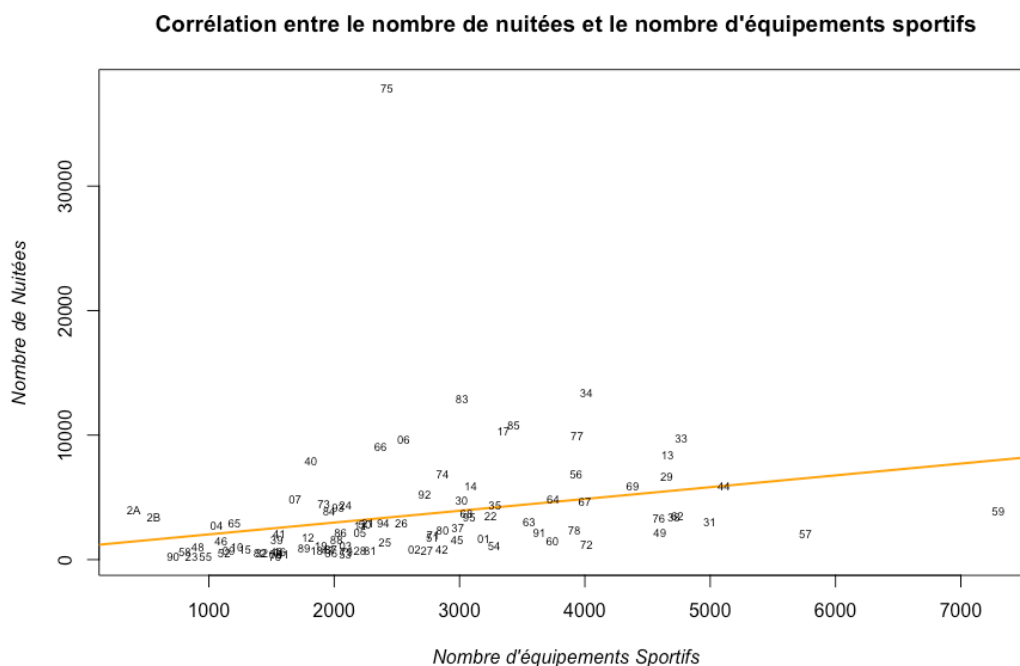


Figure 8 - Corrélation entre le nombre de nuitées et le nombre d'équipements sportifs : $r = 0,2628$

La corrélation positive observée de $r=0,26$ suggère qu'il existe une association faible à modérée entre le nombre d'équipements sportifs et le nombre de nuitées dans les départements.

Pour comprendre cette corrélation, la littérature nous dit que le sport de nature ludique s'intègre dans le tourisme en s'écartant de sa vocation compétitive. Bien que la pratique sportive de loisir ait augmenté ces vingt dernières années et que le sport ait une dimension économique, il est perçu par le secteur touristique comme un simple complément. Le tourisme sportif manque de structures, d'organisation et de stratégie de communication, malgré la croissance du secteur. De nombreux experts considèrent qu'il ne constitue pas une offre de service à part entière¹⁶.

7. La population active

Une autre variable explicative à laquelle nous nous sommes intéressées est la population active. Elle est issue du recensement effectué en 2020 par l'INSEE¹⁷ et nous indique le taux d'activité par départements pour l'ensemble des hommes et femmes âgés de 15 à 64 ans.

¹⁶ <https://theses.hal.science/tel-00482139/document> (consulté le 27/09/2023)

¹⁷ https://www.insee.fr/fr/statistiques/2012710#tableau-TCRD_015_tab1_departements (consulté le 27/09/2023)

Selon l'Organisation de Coopération et de Développement Économiques (OCDE) ¹⁸, la population active est définie comme « la somme des personnes en âge de travailler qui sont employées et celles qui sont au chômage. » Cette définition place l'accent sur la disponibilité et la volonté de travailler, qu'une personne ait déjà un emploi ou soit en recherche active d'emploi.

Nous nous intéressons à comment cette variable peut influencer le nombre de nuitées, en tenant compte du fait qu'une main-d'œuvre abondante est indispensable pour le développement et l'entretien des infrastructures telles que les hôtels, les restaurants et les attractions. De plus, une population active compétente dans des domaines clé tels que le marketing ou le service, peut rehausser la qualité de l'expérience touristique, ce qui attire davantage de visiteurs. En outre, l'apport en compétences et en innovation de cette population active est déterminant pour créer de nouveaux produits et services dans l'industrie du tourisme.

Cependant, nos données ne définissent pas les catégories socioprofessionnelles, ce qui limite la possibilité d'une analyse approfondie sur les différents types de population active. Par conséquent, nous examinerons cette corrélation de manière globale.

Dans la littérature, et précisément dans un contexte post-COVID, nous constatons que les capacités de reprise du secteur du tourisme sont menacées par la pénurie de main-d'œuvre.¹⁹ Une augmentation de cette population pourrait aider à couvrir cette pénurie, y compris pour les chômeurs qui représentent un réservoir potentiel de main-d'œuvre. Une fois formés et embauchés, ils peuvent aider à combler les lacunes en compétences et répondre à la demande croissante de services de voyages, permettant ainsi de garantir les capacités d'accueil²⁰ et de générer une stabilité voire une augmentation des nuitées.

¹⁸ <https://data.oecd.org/fr/emp/population-active.htm> (consulté le 26/10/2023)

¹⁹ <https://www.senat.fr/questions/base/2022/qSEQ220628097.html> (consulté le 26/10/2023)

²⁰ https://www.lepoint.fr/economie/les-vacances-menacees-par-la-penurie-de-main-d-oeuvre-24-05-2022-2476807_28.php#11 (consulté le 26/10/2023)

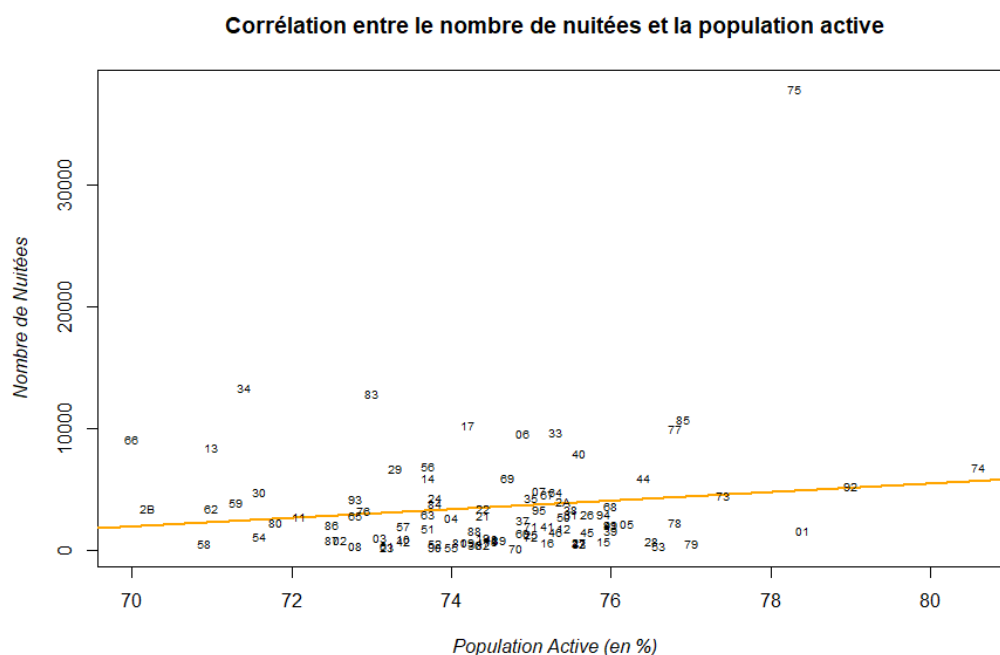


Figure 9 – Corrélation entre le nombre de nuitée et la population active : $r = 0,1409$

Ce graphique présente une faible corrélation positive entre l'attractivité touristique mesurée par le nombre de nuitées réservées et le taux d'activité des 15 à 64 ans. A mesure que la population active augmente, le nombre de nuitées augmente donc, mais faiblement. Le signe de cette corrélation est donc conforme à la littérature.

8. Le réseau routier

Notre variable, "le réseau routier", correspond à la somme totale, en kilomètres, des autoroutes et des routes nationales départementales au 31 décembre 2021.²¹

Le tourisme dépend de deux facteurs principaux. Le premier est le désir ou l'intention de voyager, influencé par l'attrait des destinations et la possibilité de voyager (en termes de temps et de ressources financières). Le second est la capacité physique d'accéder à ces destinations, ce qui implique la disponibilité et l'utilisation d'infrastructures de transport²² et de communication adéquates.

Nous avons déjà abordé à plusieurs reprises le premier paramètre. En introduisant la variable "le réseau routier", nous plongeons dans l'exploration du second, axée

²¹ <https://www.insee.fr/fr/statistiques/2012705> (consulté le 28/09/2023)

²² [https://www.researchgate.net/publication/325256026 Réseaux de transport et mobilités touristiques](https://www.researchgate.net/publication/325256026_Réseaux_de_transport_et_mobilites_touristiques) (consulté le 14/10/2023)

principalement sur l'infrastructure. En mettant l'accent sur l'infrastructure routière, notre approche se penche directement sur les questions d'accessibilité et de mobilité départementale. Cela nous permet de cerner avec précision comment l'infrastructure routière peut impacter le tourisme, sans la distraction potentielle des transports aériens, sujet que nous aborderons par ailleurs dans notre discussion.

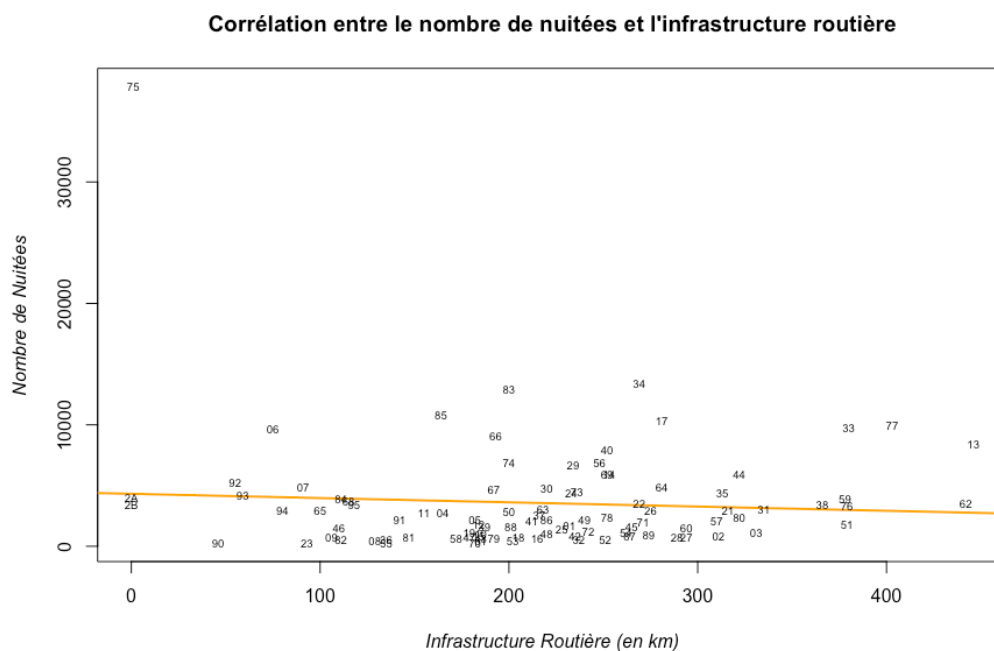


Figure 10 - Corrélation entre le nombre de Nuitées et l'infrastructure routière : $r = -0,071$

La corrélation négative entre l'infrastructure routière et le nombre de nuitées, telle que représentée par le coefficient de corrélation $r = -0.071$, suggère qu'à mesure que l'infrastructure routière d'un département augmente, le nombre de nuitées tend légèrement à diminuer. Cependant, il est important de noter que cette corrélation est faible, comme l'indique la valeur proche de zéro de r .

Lors de l'examen des résultats, nous avons été surprises de découvrir une corrélation négative entre notre variable illustrative et notre variable active. Cette observation va à l'encontre de notre intuition initiale et des affirmations courantes dans la littérature, qui suggèrent que c'est le réseau routier qui facilite les grands flux touristiques ²³.

Cependant, lorsque l'on réfléchit en termes de mobilité touristique, c'est-à-dire à la manière dont les touristes se déplacent au sein de l'infrastructure, plusieurs observations

²³https://www.caissedesdepots.fr/fileadmin/PDF/finance_carbone/etudes_climat/etude_climat_n18-e.pdf (consulté le 14/10/2023)

intéressantes émergent. Premièrement, il apparaît que les touristes ont tendance à privilégier les voyages de courte distance par rapport à ceux de moyenne distance, et les voyages de moyenne distance par rapport à ceux de longue distance ²⁴.

Deuxièmement, la tendance de « la sobriété », qui vise à réduire une activité spécifique, dans ce cas, la longueur d'un trajet ²⁵.

Et enfin des études montrent que les résultats sont assez inégaux et contrastés d'un département à l'autre lorsqu'on cherche à affirmer qu'une augmentation de l'infrastructure routière entraîne systématiquement une hausse de l'activité touristique²⁶.

9. Les aménagements cyclables

Notre variable Aménagements Cyclables comptabilise la distance en km spécifiquement adaptés à la circulation des vélos. Cela comprend à la fois les pistes cyclables, mais également les voies vertes et les bandes cyclables²⁷.

Si nous avons choisi d'étudier cette variable, c'est qu'elle est synonyme de mobilité douce²⁸. En effet en installant ces équipements, dont la première vocation est de favoriser le déplacement des locaux par un moyen non motorisé, les départements ont également créé des objets touristiques. Ces aménagements sont perçus comme des infrastructures de loisir permettant de se rendre dans un lieu à visiter par exemple, voire même de pratiquer du « vélotourisme ». Ils ont également l'avantage de pouvoir être utilisés de façon totalement gratuite et permettent soit de découvrir des territoires naturels d'une façon plus authentique, soit de circuler plus facilement dans des centres-villes encombrés. Dans tous les cas, leur usage est respectueux de l'environnement. Aussi, la présence de ces aménagements contribue à améliorer l'image des départements qui les possèdent.

²⁴

https://www.researchgate.net/publication/325256026_Réseaux_de_transport_et_mobilites_touristiques (consulté le 14/10/2023)

²⁵https://presse.ademe.fr/2021/06/le-tourisme-durable-en-france-un-levier-de-relance-ecologique.html#_ftnref1 (consulté le 15/10/2023)

²⁶https://temis.documentation.developpement-durable.gouv.fr/docs/Temis/0060/Temis-0060963/7664_Rapport.pdf (consulté le 15/10/2023)

²⁷ <https://www.amenagements-cyclables.fr/fr/stats> (consulté le 01 /10 /2023)

²⁸ https://doi.org/10.3406/edyte.2006.982www.persee.fr/doc/edyte_1762-4304_2006_num_4_1_982 (consulté le 12 /10 /2023)

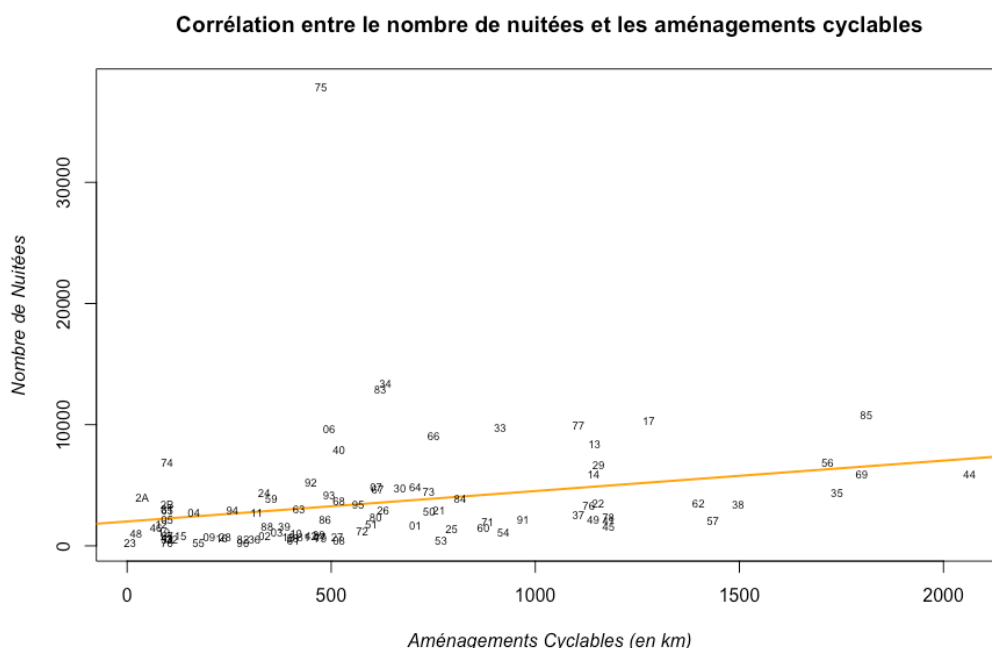


Figure 11 - Corrélation entre le nombre de Nuitées et la distance d'Aménagements Cyclables mise en place : $r = 0,2541$

Ce graphique présente une relation linéaire positive entre le nombre de nuitées et la distance en km d'aménagement cyclistes mise en place, avec un coefficient de corrélation de 0,25. Cette corrélation pourrait même être bien supérieure sans l'impact de Paris. En accord avec la littérature, nous observons donc concrètement que la mobilité douce est un facteur d'attractivité touristique pour les départements français.

10. La densité de population

La variable explicative qui vient ensuite est la densité de population. Pour l'obtenir dans les départements métropolitains, nous avons divisé la population²⁹ de chaque département par sa superficie³⁰.

Selon la littérature, la politique publique tend à établir un lien entre la densité et le dynamisme économique³¹. En conséquence, une grande partie des investissements est dirigée vers les zones urbaines les plus peuplées, qui promettent des rendements financiers plus élevés. Cette orientation a donné naissance à une économie métropolitaine autosuffisante, qui attire encore plus de personnes dont des touristes, et favorise une

²⁹ <https://www.insee.fr/fr/statistiques/1893198> (consulté le 27 /09 /2023)

³⁰ <https://www.observatoire-des-territoires.gouv.fr/superficie> (consulté le 27 /09 /2023)

³¹ https://www.senat.fr/rap/r16-565/r16-565_mono.html (consulté le 30 /10 /2023)

croissance soutenue. Cependant, il est paradoxal de voir que, malgré cette concentration sur l'urbanisation, nous assistons à un déclin des compétences et à un manque de planification efficace dans les zones rurales. Cette négligence est surprenante, étant donné que le cadre de vie offert par ces campagnes correspond aussi à des attentes de nos habitants et des nombreux touristes qui visitent notre pays chaque année.

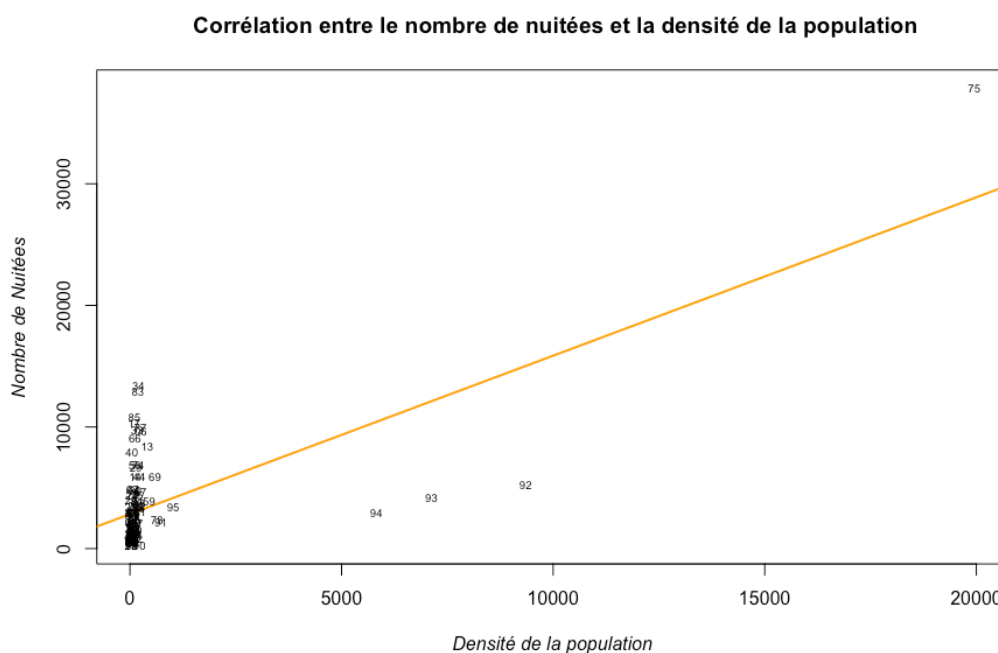


Figure 12 - Corrélation entre le nombre de nuitées et la Densité de Population : $r = 0,6741$

Ce graphique présente une forte corrélation : 0,67, entre la fréquentation des hébergements collectifs de tourisme et la densité de population. A mesure que la densité d'un département augmente, le nombre de nuitées suit : cela concorde avec nos recherches littéraires. Toutefois il est intéressant de constater que dans cette représentation, le coefficient est très largement influencé par les départements de la région Île de France.

C. Statistiques et valeurs aberrantes des variables

Afin de mieux comprendre les observations de nos variables, nous avons calculé leurs principales statistiques.

	Min	1st Qu..25%	Médiane	Moyenne	3rd Qu..75%	Max	Ecart type
Nuitées	243	877	2282	3573	4373	37850	4623
Commerces	435	1337	2332	3057	3910	20366	2729
Médecins Généralistes	165	235	288	304	357	888	98
Equipements Culturels	32	174	238	252	322	655	119
Niveau de vie médian	18470	20990	21665	22052	22520	28810	1649
Température moyenne	10	12	13	13	14	17	2
Equipements Sportifs	400	1741	2380	2628	3299	7298	1281
Population Active	70	73	75	75	76	81	2
Routes	0	165	217	216	270	446	94
Vélo	7	284	505	622	876	2064	467
Densité Population	15	50	83	566	169	19949	2391

Figure 13 - Principales statistiques des variables

Notre variable Nuitées, baromètre de l'attractivité touristique, s'étend d'un minimum de 243 à un maximum impressionnant de 37 850 nuitées : cela met en lumière la disparité de l'attractivité touristique entre les départements.

Ces statistiques nous permettent également d'en savoir plus sur les variables explicatives. On observe entre autres qu'il existe au moins un département ne possédant aucun km d'autoroutes ni de route nationale, mais aussi que la température moyenne des départements et le taux de population active parmi les 15 – 64 ans sont relativement stables respectivement autour de 13 °C et de 75 %.

Par ailleurs, comme pour la variable illustrative, nous observons que pour la plupart des variables explicatives, il existe un écart très important entre le troisième quartile et le maximum : cela laisse présager de l'existence de valeurs aberrantes.

Afin de les visualiser, nous avons ensuite réalisé les boxplots des différentes variables (Annexe 1). Dans chacune d'elles, des individus potentiellement atypiques étaient visibles. En effectuant des tests de Rosner (Annexe 2) nous avons pu confirmer la présence de certaines valeurs aberrantes pour des départements comme Paris (principalement) ou le Nord. Si nous avons choisi de les conserver, c'est pour traduire fidèlement la réalité de la France métropolitaine dans notre analyse.

III. Analyse en Composantes Principales ACP

A. Valeurs propres et nombre d'axes

La première étape de l'ACP a été de déterminer les axes à prendre en compte. Dans cet objectif, nous avons examiné les valeurs propres de l'ensemble des dimensions analysées.

	Valeurs propres	% de variance	% de variance cumulée
Dim 1	3,95	39,49	39,49
Dim 2	2,26	22,56	62,05
Dim 3	1,52	15,18	77,23
Dim 4	0,8	7,97	85,2
Dim 5	0,49	4,91	90,11
Dim 6	0,36	3,6	93,71
Dim 7	0,24	2,36	96,06
Dim 8	0,19	1,87	97,93
Dim 9	0,13	1,34	99,27
Dim 10	0,07	0,73	100

Tableau 1 – Les valeurs propres

Nous constatons que les trois premières composantes expliquent respectivement 39,49 %, 22,56 % et 15,18 % de la variance totale, cumulant ainsi 77,23 %. Par conséquent, nous choisirons de retenir uniquement ces trois premières dimensions pour notre analyse, car elles capturent l'essentiel de l'information et sont les plus significatives pour notre étude.

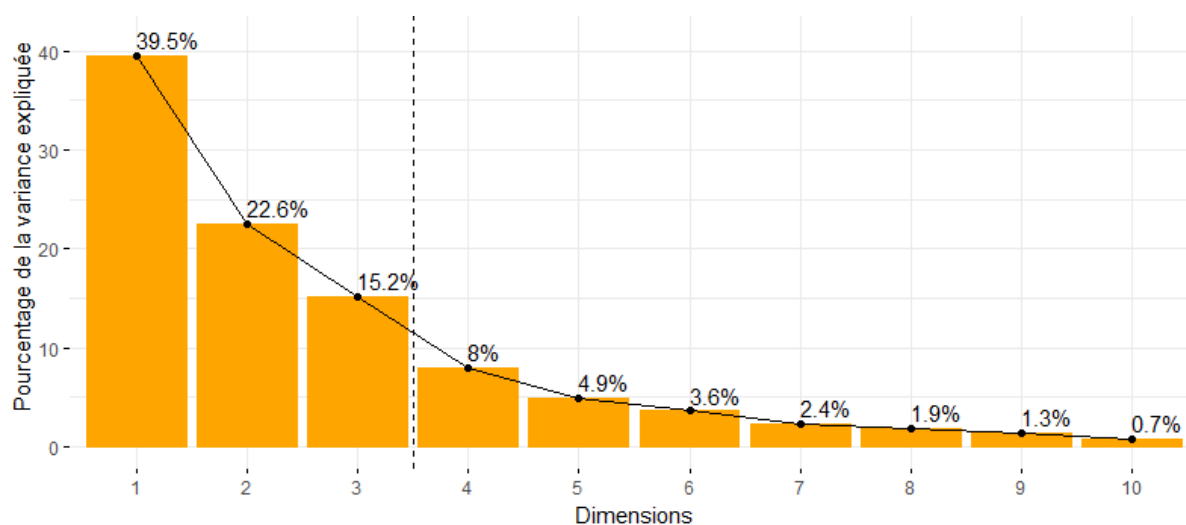


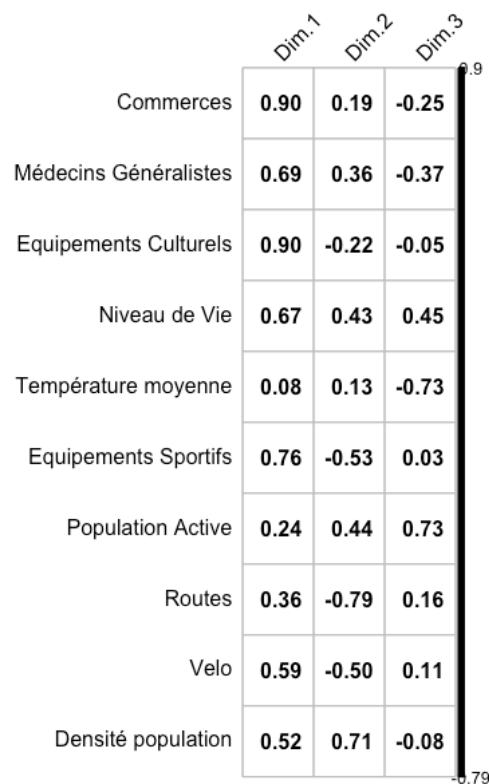
Figure 14 – Répartition de la variance

B. Variables : coordonnées, contributions, cosinus carrés

Dans cette nouvelle étape de l'ACP, nous allons étudier les caractéristiques des variables selon les 3 dimensions retenues.

1. Les coordonnées

En premier lieu, nous nous sommes intéressées aux coordonnées de nos variables.



The table displays the coordinates of ten variables across three dimensions (Dim.1, Dim.2, Dim.3). A vertical color scale bar on the right indicates the magnitude of the coordinates, ranging from -0.79 (dark blue) to 0.9 (dark red), with 0 being white.

	Dim.1	Dim.2	Dim.3
Commerces	0.90	0.19	-0.25
Médecins Généralistes	0.69	0.36	-0.37
Equipements Culturels	0.90	-0.22	-0.05
Niveau de Vie	0.67	0.43	0.45
Température moyenne	0.08	0.13	-0.73
Equipements Sportifs	0.76	-0.53	0.03
Population Active	0.24	0.44	0.73
Routes	0.36	-0.79	0.16
Velo	0.59	-0.50	0.11
Densité population	0.52	0.71	-0.08

Figure 15 - Les coordonnées des variables

Avec les signes de ces coordonnées, nous avons pu placer les différentes variables dans les cadrans adaptés sur les plans des dimension 1 et 2 ainsi que des dimension 1 et 3.

II	I
	Population Active Niveau de vie Médecins Généralistes Commerces Température Densité Population
III	IV
	Routes Equipements Sportifs Equipements Culturels Vélo

Tableau 2 - Les coordonnées des variables par cadran sur le plan dim1/dim2

Dans le plan comprenant les dimensions 1 et 2, nous observons que l'intégralité de nos variables se place dans la partie positive de la dimension 1. Concernant la dimension 2, certaines de nos variables l'impactent positivement, mais également négativement.

II	I
	Population Active Niveau de vie Routes Equipements Sportifs Vélo
III	IV
	Equipements Culturels Médecins Généralistes Commerces Température Densité Population

Tableau 3 - Les coordonnées des variables par cadran sur le plan dim1/dim3

Dans le plan comprenant les dimensions 1 et 3, le constat est similaire au précédent : les variables impactent uniquement positivement la dimension 1 et dans les deux sens la dimension 3.

2. Les contributions

Dans un second temps, nous nous sommes concentrées sur les contributions des variables. La contribution d'une variable à un axe indique dans quelle mesure chaque variable contribue à la construction de chaque composante principale. Les variables avec des contributions plus élevées sont considérées comme exerçant une influence plus forte sur la définition de l'axe.

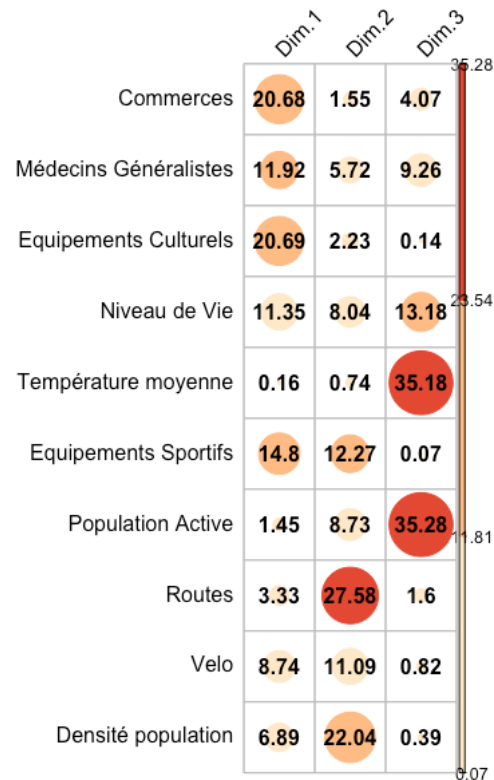


Figure 16 - Les contributions des variables

Pour l'axe 1, nous remarquons que les variables équipements culturels et commerces ont les contributions et donc les poids les plus importants, à 20,69 % chacune, suivies des équipements sportifs à 14,80 %. Comme nous l'avons vu dans la partie des coordonnées ces trois variables impactent positivement la dimension 1, elles font donc valoir leur poids de manière de façon positive sur cet axe.

En ce qui concerne l'axe 2, la variable route est la variable la plus contributive avec un poids de 27,58 %. Ce poids s'applique négativement sur l'axe 2 au vu de ses coordonnées. Vient ensuite la variable densité de population avec une contribution de 22,04 % qui joue de façon positive sur l'axe.

Pour l'axe 3, les contributions les plus importantes sont celles de la population active et de la température, avec respectivement 35,28 % et 35,18 %, elles sont suivies du niveau de vie à 13,18 %. Si la température exerce ce poids de manière négative sur l'axe 3, les deux autres variables avec leur contribution impactent positivement cette dimension.

3. Les cosinus carrés

La dernière caractéristique propre aux variables que nous avons exploitée est le cosinus carré associé à chaque dimension. Un cosinus carré mesure la qualité de représentation de la variable sur l'axe concernée. Plus il est proche de 1 et meilleure est cette qualité.

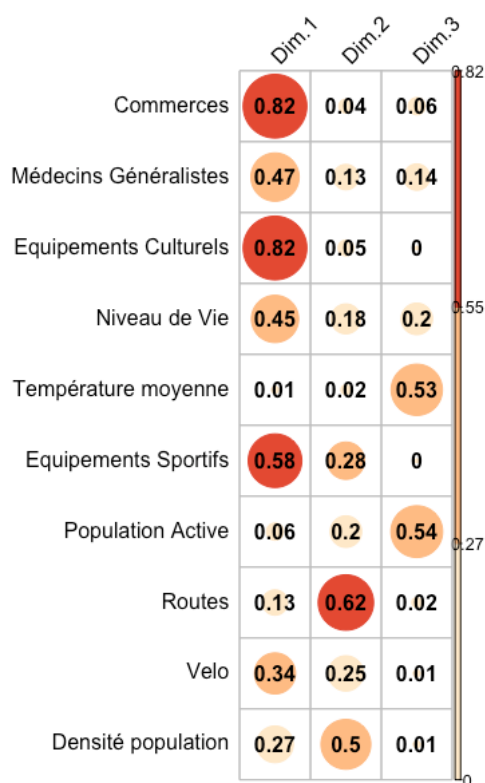


Figure 17 – Les cosinus carrés des variables

Pour la dimension 1, nous nous intéressons aux variables les plus contributives que nous avons repérées précédemment, à savoir les équipements culturels, les commerces et les équipements sportifs. Les deux premières variables ont un cosinus carré élevé de 0,82 qui témoigne de leur qualité dans la représentation de l'axe 1. La variable des équipements sportifs, quant à elle, présente un cosinus carré de 0,58, certes moins important, mais qui reste à prendre en considération. Il est à noter que cette variable a un cosinus carré de 0,28 pour la dimension 2. Dans le plan des dimensions 1 et 2, le fait qu'elle ne définissent pas un unique axe sera donc visible.

Dans la dimension 2, on se concentre sur les variables route et densité de population. Elles possèdent respectivement des cosinus carrés de 0,62 et de 0,5 qui sont les plus importants de la dimension. Cependant ces deux variables et principalement densité de population ont également un cosinus carré dans la dimension 1 qui entre en considération dans la représentation sur le plan des dimensions 1 et 2. Ces deux variables ne seront donc pas représentées directement sur l'axe 2 mais légèrement en diagonale avec l'axe 1.

Pour la dimension 3, les cosinus carrés à observer sont ceux des variables population active et température. Ils sont respectivement de 0,54 et de 0,53, et il s'agit des deux cosinus carrés les plus élevés de la dimension. Aussi, bien que ces deux valeurs s'éloignent de 1, elles sont celles qui traduisent le plus qualitativement la dimension 3. Par ailleurs, les cosinus carrés de ces deux variables sont proches de 0 dans la dimension 1, donc sur le plan des dimensions 1 et 3 des cosinus de 0,53 et 0,54 se démarqueront fortement.

C. Cercle de corrélation

Après avoir défini les caractéristiques de nos variables explicatives, il est maintenant tant de les placer sur les cercles de corrélations des plan dimensions 1 et 2, ainsi que dimensions 1 et 3.

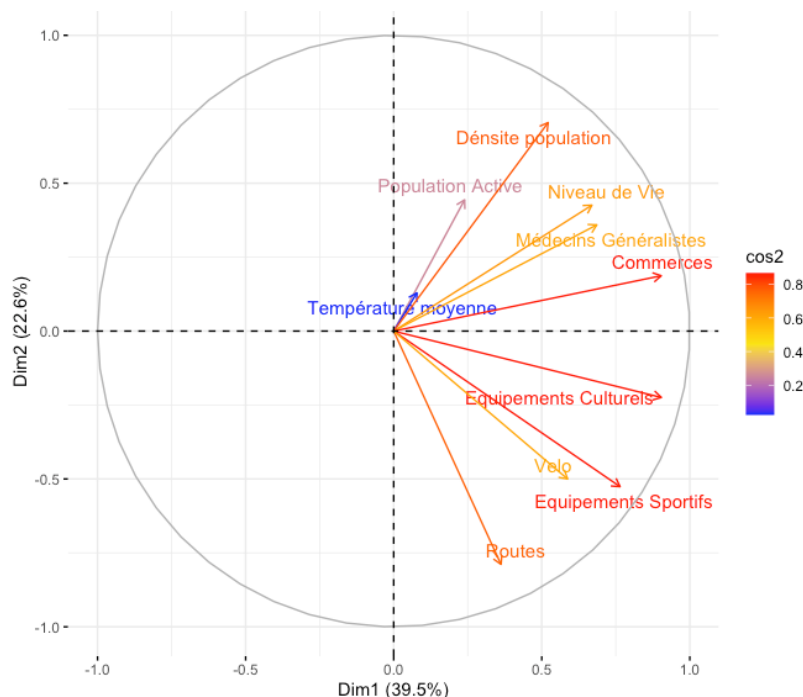


Figure 18 - Cercle de corrélation des dimensions 1 et 2

Ce premier cercle de corrélation dans le plan des dimensions 1 et 2 est conforme à nos attentes. 'Commerces' et 'Culture' y apparaissent comme les principales variables représentatives de la dimension 1, toutes deux l'influençant positivement. La dimension 2 a quant à elles comme variables significatives la 'Densité Population' et 'Routes' l'influençant respectivement, positivement et négativement.

Par ailleurs, on observe dans ce cercle que des variables comme 'Niveau de Vie', 'Médecins Généralistes' ou encore 'Equipements Sportifs' sont projetées en diagonales : on ne peut donc pas les associer à une unique dimension car elles contribuent à la fois à la définition des axes 1 et 2.

On observe également qu'à elles seules les dimensions 1 et 2 expliquent 62,05 % de la variance totale.

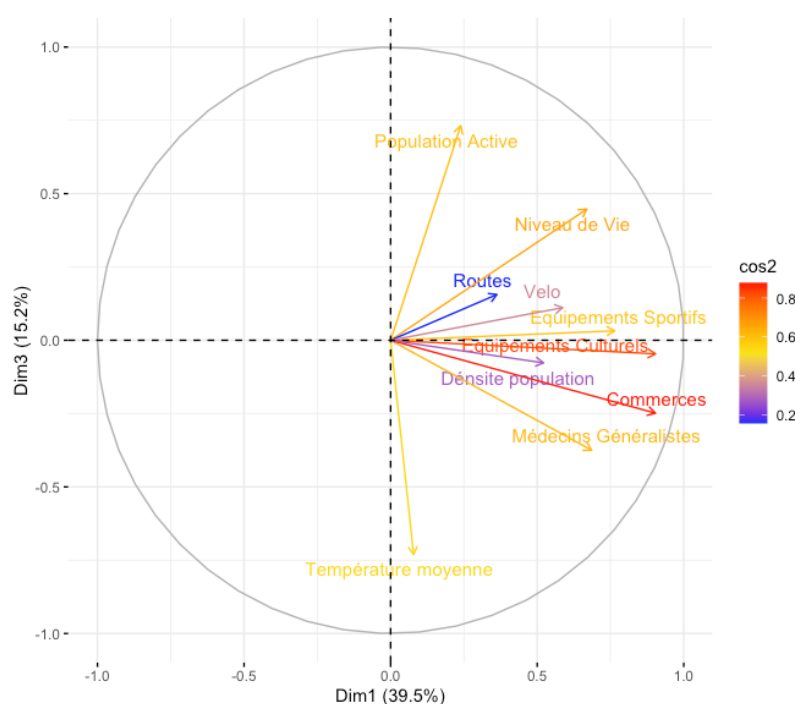


Figure 19 - Cercle de corrélation des dimensions 1 et 3

Le second cercle de corrélation, se trouve dans le plan des dimensions 1 et 3 (plan qui explique 54,7 % de la variance totale). Son constat reste conforme à nos analyses précédentes : les variables 'Population Active' et 'Température' sont celles qui influencent principalement l'axe 3 respectivement positivement et négativement. En effet, elles se

positionnent presque dans l'axe de cette dimension. De manière moindre, le 'Niveau de Vie' participe également à la définition de l'axe 3.

Concernant la dimension 1, on s'aperçoit que la variable 'Equipements Sportifs' s'aligne davantage dans l'axe : dans le plan précédent elle apportait également une participation à la construction de l'axe 2, mais dans ce plan elle participe uniquement à celle de la dimension 1 et de façon positive. On retrouve bien entendu la 'Culture' et les 'Commerce' dans la définition de cette dimension.

D. Définition des variables latentes

Afin de définir nos variables latentes, nous allons synthétiser les informations les plus significatives obtenues pour chaque axe.

	Négatif	Contribution	Positif	Contribution
Axe 1			Commerces	20,68
			Equipements Culturels	20,69
			Equipements Sportifs	14,80
Axe 2	Route	27,58	Densité Population	22,04
Axe 3	Température	35,18	Population Active	35,28

Tableau 4 - Les variables principales définissant chaque axe

Notre démarche pour identifier les variables latentes débute avec l'axe 1. Nous constatons que les variables les plus influentes, positivement, sont les commerces, mais aussi les équipements culturels et sportifs. L'absence de variables significatives sur le côté négatif de cet axe renforce notre compréhension. Ces éléments nous ont permis de définir notre premier axe comme étant **l'accès aux services de consommation et de loisirs**. A mesure que l'on se déplace vers la droite de l'axe, l'offre de services commerciaux, culturels et de divertissement augmente.

Nous poursuivons nos discussions en nous concentrant sur l'axe 2. Nous constatons que la variable densité de population augmente avec cet axe et qu'à contrario la température elle diminue avec l'axe 2. La nature de la variable latente ne nous apparaissait pas clairement au premier abord. Néanmoins, notre réflexion nous a amenés à considérer que la densité de population dans nos départements, lorsqu'elle est mise en relation avec le kilométrage des infrastructures routières, pourrait influencer les modes de déplacement au sein du territoire. Dans les zones à forte densité démographique, il semble que les transports en commun soient privilégiés, probablement en raison de l'efficacité et de la praticité dans les espaces urbains denses. À l'inverse, dans les départements bien pourvus en infrastructures routières, les déplacements individuels, comme l'usage de la voiture, pourraient être favorisés, suggérant un mode de vie où la mobilité personnelle est plus aisée et peut-être plus nécessaire en raison d'une moindre concentration de services et d'habitants.³² Nous avons donc défini cette variable latente comme mesurant les **pratiques de mobilité collective**. A mesure que l'on se déplace vers le haut de cet axe, l'importance des infrastructures de transport et les habitudes de déplacement en commun augmentent.

Enfin, nous avons considéré l'axe 3. Les principales variables qui le définissent sont négativement, la température et positivement, la population active. Nous avons alors défini cet axe comme baromètre d'un **environnement propice au travail**. A mesure que l'on se déplace vers le haut de l'axe 3, nous nous trouvons dans des conditions plus favorables à l'emploi.

E. Projection des individus sur le plan factoriel

Nous avons ensuite cherché à savoir si le positionnement des départements dans les plans des dimensions 1 et 2 ainsi que 1 et 3 correspondait aux définitions des variables latentes.

Nous avons alors déterminé les départements les plus significatifs de chaque axe.

Dans cette partie, les numéros des observations ne correspondent pas aux codes des départements, mais à leur positionnement parmi les départements. Aussi, la Haute-Corse devient le 21, et ainsi de suite.

³² https://www.senat.fr/rap/r20-313/r20-313_mono.html

	Négatif	Contribution	Positif	Contribution
Axe 1			Paris (76)	24,95
			Rhône (70)	4,38
			Nord (60)	3,99
	Haute-Corse (21)	2,39	Bouches-du-Rhône (13)	3,55
	Creuse (24)	2,29	Gironde (34)	3,44
			Isère (39)	2,98
			Loire-Atlantique (45)	2,82
			Hauts-de-Seine (93)	2,44
Axe 2	Pas-de-Calais (63)	6,03	Paris (76)	35,1
	Nord (60)	3,84	Hauts-de-Seine (93)	10,86
	Seine-Maritime (77)	2,51	Val-de-Marne (95)	3,81
	Moselle (58)	2,5	Corse-du-Sud (20)	3,78
	Bouches-du-Rhône (13)	2,05	Haute-Savoie (75)	1,95
Axe 3	Haute-Corse (21)	9,02		
	Pyrénées-Orientales (67)	7,7		
		6,21	Haute-Savoie (75)	9,87
	Hérault (35)	5,46	Ain (01)	4,24
	Bouches-du-Rhône (13)	3,51	Yvelines (79)	3,81
	Aude (11)	3,37	Hauts-de-Seine (93)	3,63
	Alpes-Maritimes (6)	3,33		
	Var (84)			

Tableau 5 – Les départements les plus significatifs de chaque axe

Afin de mieux visualiser ces départements en croisant les dimensions, nous allons les projeter sur les plans des dimensions 1 et 2 ainsi que 1 et 3 avant de les interpréter.

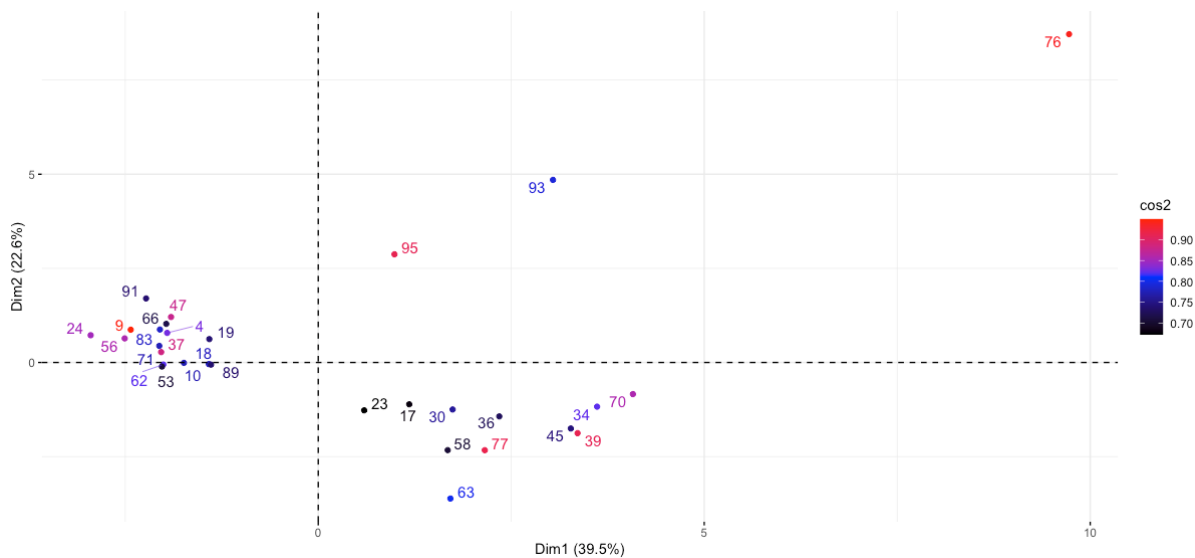


Figure 20 -Projection des 30 départements les plus significatifs sur le plan des dimensions 1 et 2

Projetés à droite de notre axe 1, nous avons des individus comme : Paris (76), le Rhône (70) (avec Lyon comme préfecture), les Hauts-de-Seine (93) (avec des villes comme Boulogne-Billancourt et Nanterre), et Gironde (34) (avec Bordeaux) .

Paris, Le Rhône, et Hauts-de-Seine sont des centres économiques et culturels majeurs, tandis que la Gironde et l'Isère se distinguent par leur riche patrimoine historique et culturel. Ensemble, ces départements reflètent une combinaison d'importance économique et culturelle au sein de la France.

À gauche de l'axe 1, nous retrouvons les départements de l'Indre (37), l'Ariège (09), l'Orne (62), la Meuse (56), le Cher (18), l'Aube (10), la Haute-Saône (71), la Creuse (24), et le Lot (46). Ces départements partagent plusieurs caractéristiques communes, notamment leur nature majoritairement rurale. Leur économie est principalement axée sur l'agriculture, reflétant un mode de vie plus traditionnel et moins urbanisé.

En haut, sur l'axe 2, comme nous l'attendions, nous retrouvons des individus tels que Val-de-Marne (95), des Hauts-de-Seine (93), et de Paris (76). Ces départements sont parmi les plus densément peuplés et sont étroitement intégrés dans le tissu métropolitain de l'Île-de-France. La mobilité collective y est donc un enjeu majeur.

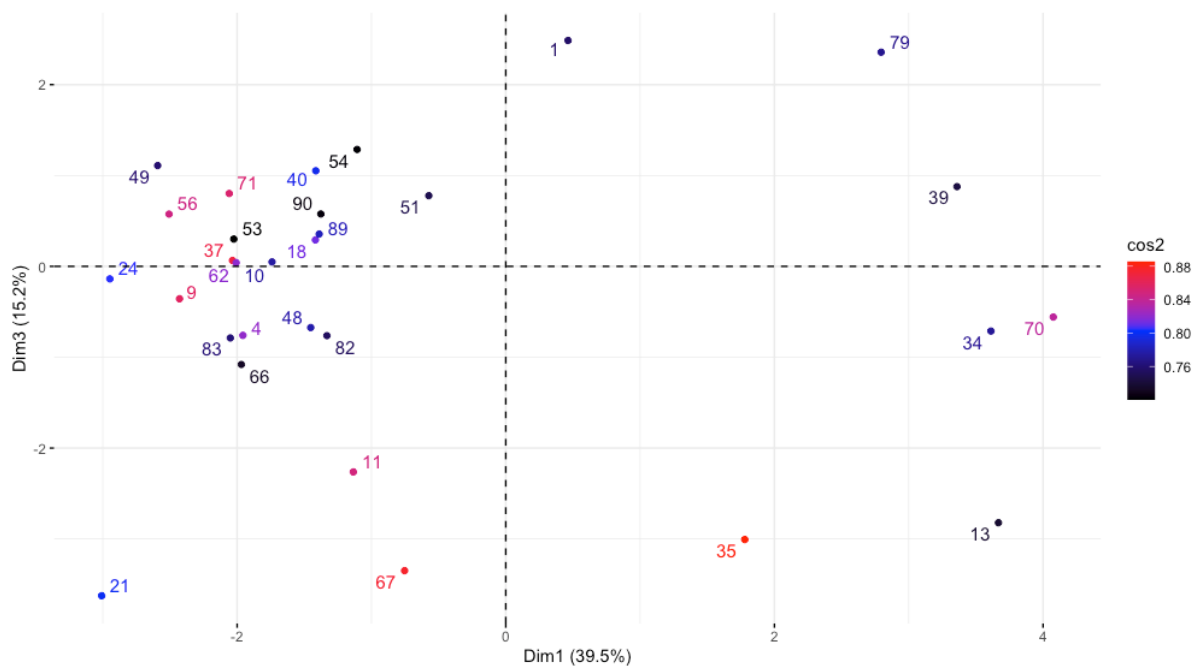


Figure 21 - Projection des 30 départements les plus significatifs sur le plan des dimensions 1 et 3

La projection des départements, en accord avec notre troisième variable latente : environnement propice au travail ; montre des résultats intéressants. Nous observons que les départements avec un niveau de vie supérieur, et où le pourcentage de la population active est plus concentré, tels que la Haute-Savoie (74), l'Ain (1) et la Manche (50), sont situés dans des zones où la température moyenne est inférieure à 13 °C. Cette observation suggère que des températures plus fraîches pourraient favoriser un environnement de travail plus productif. En revanche, dans la partie inférieure de notre projection, nous retrouvons des départements tels que les Pyrénées-Orientales (66), l'Aude (11) et l'Hérault (34), caractérisés par des températures moyennes d'environ 16 °C. Toutefois, il est important de souligner que ces résultats ne confirment pas la notion généralisée selon laquelle des températures plus élevées réduiraient systématiquement la croissance économique et les revenus. Affirmer que les départements les plus chauds sont les plus pauvres serait une simplification excessive. En effet, au niveau interdépartemental, une telle affirmation n'a pas été démontrée. Par conséquent, bien que le climat puisse influencer certains aspects de l'économie, il ne détermine pas à lui seul le destin économique d'un département.

F. Projection de la variable illustrative sur le cercle de corrélation

Maintenant que nos variables latentes sont définies et ont été appuyées par les observations, nous allons à présent regarder, où se projette notre variable illustrative sur les deux plans étudiés.

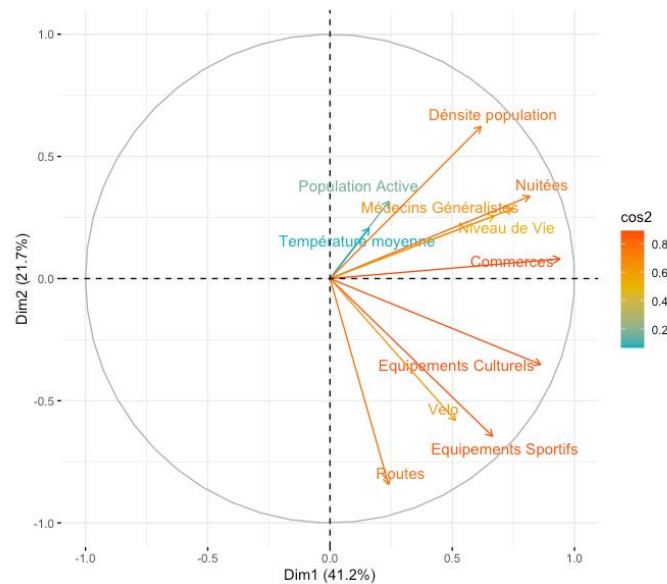


Figure 22 - Cercle de corrélation des dimensions 1 et 2 comprenant la variable illustrative

Dans le plan des dimensions 1 et 2, la fréquentation des hébergements collectifs touristiques se positionne dans le cadran I. Cela signifie qu'elle augmente à la fois avec l'accès aux services de consommation et de loisir mais aussi avec la pratique de mobilité collective.

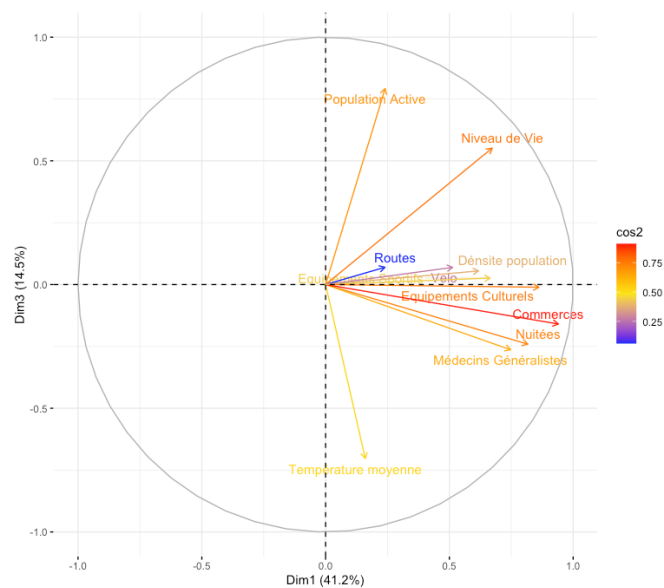


Figure 23 - Cercle de corrélation des dimensions 1 et 3 comprenant la variable illustrative

Dans le plan des dimensions 1 et 3, c'est cette fois-ci dans le cadran IV que se projette la variable illustrative. Concrètement, cela veut dire qu'elle croît lorsque l'accès aux services de consommation et de loisir augmente, mais également lorsque l'environnement propice au travail diminue.

G. ACP : le cas sans Paris

Nous avons décidé d'établir une nouvelle version de cette ACP en enlevant Paris. Cette démarche vise à réduire les distorsions dues au statut exceptionnel du département de la capitale et nous permettra de comparer de manière plus équilibrée les résultats et caractéristiques des autres départements.

Au vu d'un aperçu rapide des nouvelles statistiques de nos variables (Annexe 3), nous constatons que plusieurs d'entre elles présentent des changements mineurs, voire restent inchangées, telles que la température, les équipements sportifs ou la population active. Toutefois, pour d'autres variables, comme anticipé, il y a des changements considérables, notamment en ce qui concerne la densité de population ou le nombre de commerces.

Dans le cadre l'ACP, nous examinons les valeurs propres et le nombre d'axes à étudier (Annexe 4). Il est intéressant de noter que l'exclusion de Paris de notre analyse n'affecte pas significativement notre Axe 1, qui continue d'expliquer 39 % de l'inertie. De même, notre Axe 3 maintient sa contribution à 15 % de la variance. Cependant, le constat diffère pour l'Axe 2, qui passe de 23 % à 18 %, entraînant ainsi une réduction de 4 % dans la variance totale expliquée par nos trois axes principaux. Avec ces trois axes, nous capturons désormais 73 % de la variance totale, ce qui demeure un résultat satisfaisant.

L'étape suivante est celle de la réalisation des matrices des coordonnées, des contributions ainsi que des cosinus carrés.

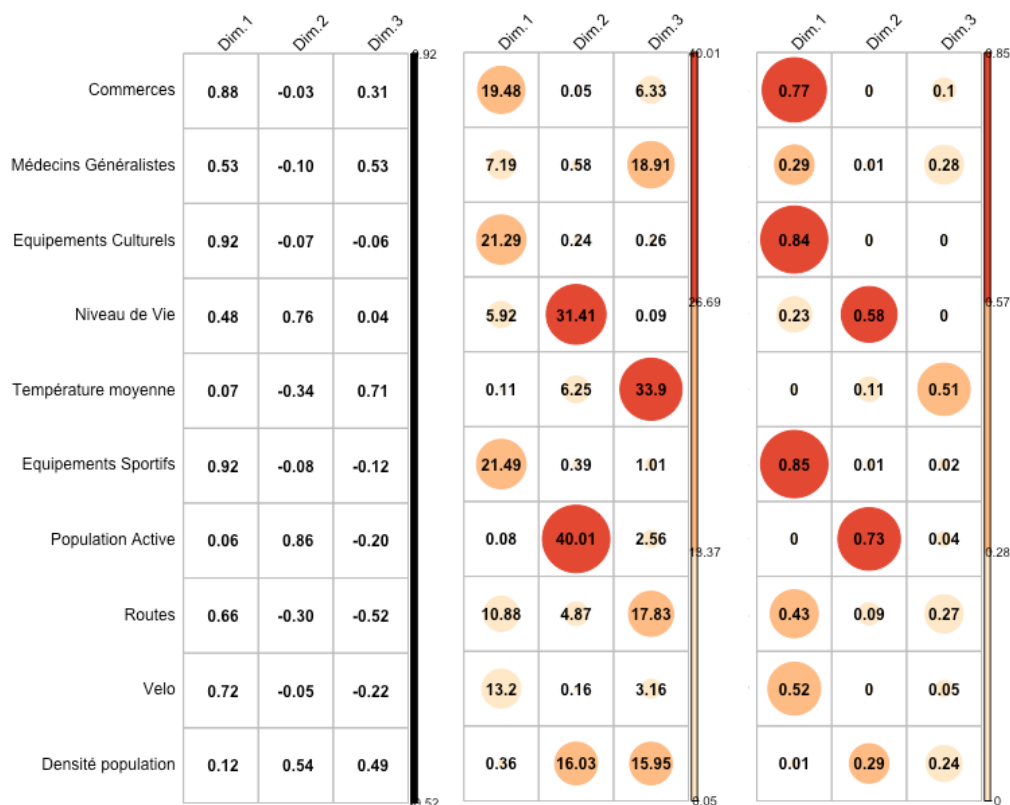


Figure 24- Coordonnées, contributions et cosinus carrés sans Paris

En observant ces matrices, nous constatons avec surprise, qu'avoir exclu Paris de notre analyse a complètement bouleversé la définition des variables latentes 2 et 3.

En effet si les variables Commerces, Equipements Culturels ainsi qu'Equipements Sportifs définissent toujours positivement l'axe 1 comme en témoigne leur positionnement ainsi leurs contributions et cosinus carré élevés, il est loin d'en être de même pour les autres dimensions.

L'axe 2 est désormais défini uniquement positivement par les variables Population Active ainsi que Niveau de Vie : elles possèdent des coordonnées positives dans cette dimension, des contributions très importantes de 40,01 % et 31,41 %, sans oublier des cosinus carrés de 0,73 et 0,58 qui témoignent de leur qualité dans la représentation. Si on les représente sur le plan des dimensions 1 et 2, la Population Active, compte tenu de ses très faibles contribution et cosinus carré pour la dimension 1, sera à elle seule un parfait baromètre de l'axe 2.

Par ailleurs, l'axe 3 qui était auparavant défini positivement par la Population Active, n'est plus aucunement impacté par cette dernière : la contribution et le cosinus carré de cette variable sont respectivement 2,56 % et 0,04 pour cet axe. La dimension conserve toutefois comme impactante la Température, mais cette fois-ci de façon positive. Dans la constitution de cet axe, on pourrait également retrouver les Médecins Généralistes et les Routes qui possèdent une contribution à prendre en compte cependant leurs cosinus carrés plus élevés dans la dimension 1 que 3 ne jouent pas en leur faveur. Dans ce contexte, on observe la variable qui suit en termes de contribution : il s'agit de la Densité de Population à 15,95 %. Avec un cosinus carré de 0,24 pour l'axe 3, de 0,01 pour l'axe 1 et un positionnement dans le cadran I du plan des dimensions 1 et 3, elle participera également positivement à la définition de l'axe 3.

Dans un tel contexte, nos variables latentes définies précédemment pour les axes 2 et 3 n'ont plus lieu d'être. Et cela est troublant que l'éviction d'un seul individu ait autant impacté nos résultats. En réalisant cette seconde ACP, nous souhaitons comparer la répartition des départements dans des plans similaires avec et sans Paris. Cependant, compte tenu des forts bouleversements observés qui rendent impossible des comparaisons pertinentes, nous ne pousserons donc pas plus loin la réflexion de cette nouvelle ACP.

IV. Corrélation entre la variable illustrative et les variables latentes

A. Régression linéaire multiple

Dans cette partie, nous avons établi 2 modèles à partir de nos variables latentes. Ils ont la forme suivante :

$$Y = \alpha + \beta_1.Axe1 + \beta_2.Axe2 + \varepsilon$$

Le premier combine l'accès aux services de consommation et de loisir avec les pratiques de mobilité collectives (modèle 1) tandis que le second reprend l'accès aux services de consommation, mais cette fois ci avec l'environnement propice au travail (modèle 2). Leur objectif a été de tester la significativité de nos axes et plus globalement de nos modèles dans l'explication de la variable illustrative.

Ci-dessous, se trouvent respectivement les résultats obtenus sur R en effectuant les régressions linéaires multiples des modèles 1 puis 2.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      3573.0      287.1  12.445 < 2e-16 ***
Accès_services_consommation_loisir  1629.0      144.5  11.275 < 2e-16 ***
Pratiques_mobilité_collective      1154.4      191.1   6.039 3.15e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2813 on 93 degrees of freedom
Multiple R-squared:  0.6376, Adjusted R-squared:  0.6298
F-statistic: 81.8 on 2 and 93 DF, p-value: < 2.2e-16
```

Figure 25 - RLM du modèle 1

Notre premier modèle est pertinent selon le test de Fisher : la p-value observée est inférieure à 1 %. Cette régression nous permet également d'interpréter, grâce aux p-values des test de Student, inférieures à 0,01 que nos variables latentes sont toutes deux significatives au seuil de 1 %, et même en deçà. Par ailleurs, le R^2 ajusté nous indique que ce modèle explique 63 % de la variance de notre variable illustrative : il est donc d'une qualité satisfaisante.

Aussi, selon ce modèle, quand l'accès à des services de consommation et de loisir d'un département augmente d'une unité, la fréquentation des hébergements collectifs de tourisme du département augmente de 1 629 nuitées. Si ce sont les pratiques de mobilité collectives qui augmentent d'une unité, ce sont 1 154 nuitées qui viennent s'ajouter.

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      3573.0      305.8   11.685 < 2e-16 ***
Accès_services_consommation_loisir  1629.0      153.9   10.587 < 2e-16 ***
Environnement_travail    -1140.7      248.2   -4.597 1.35e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2996 on 93 degrees of freedom
Multiple R-squared:  0.5889,    Adjusted R-squared:  0.58
F-statistic: 66.61 on 2 and 93 DF,  p-value: < 2.2e-16

```

Figure 26 - RLM du modèle 2

Nos observations du premier modèle s'avèrent être en partie vraies pour le second : il est pertinent au regard au test Fisher et ses axes sont significatifs au seuil de 1 %. Mais cette fois-ci, ce sont 58 % de la variance de la fréquentation des hébergements collectifs touristiques qui sont expliqués par ce modèle : nous le considérons d'une qualité acceptable.

Selon le modèle 2, quand l'accès à des services de consommation et de loisir d'un département augmente d'une unité, la fréquentation des hébergement collectifs de tourisme de ce département augmente également de 1 629 nuitées. Mais à l'inverse, s'il s'agit de l'environnement de travail qui augmente d'une unité, ce sont 1 141 nuitées qui sont déduites.

Par la suite, nous nous sommes intéressées aux résidus de nos modèles. Sous, la forme de boxplots, ils se présentent de la façon suivante :

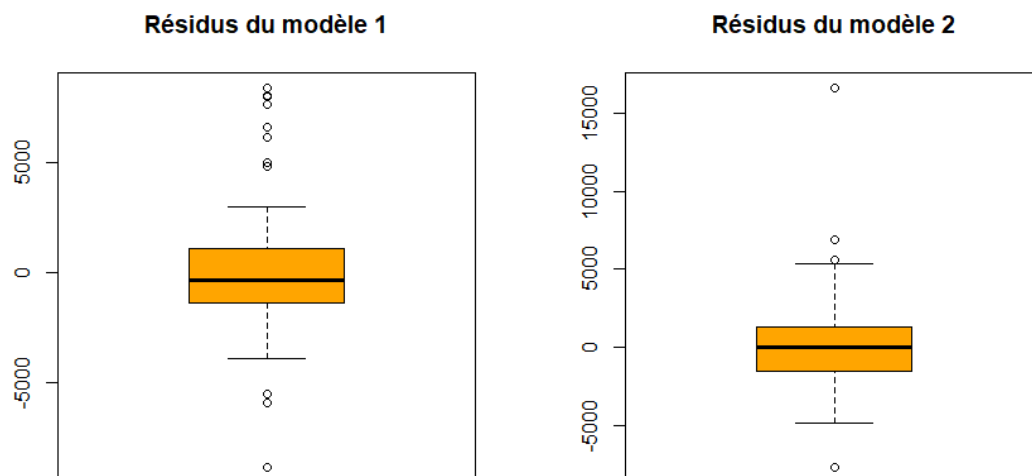


Figure 27 - Distribution des résidus des modèles 1 et 2

Ces deux graphiques nous donnent une idée générale de la distribution empirique des résidus des modèles 1 et 2. Dans les 2 modèles, il semble exister une relative symétrie des

queues de distributions (du min au max). Concernant les médianes, nous observons que si dans les résidus du modèle 2 elle semble proche de 0 et qu'il y a approximativement autant de valeurs positives que négatives, elle est en revanche inférieure à 0 dans les résidus du modèle 1 : il y a donc ici plus de valeurs négatives que positives.

Ces boxplots nous permettent également de mettre en évidence la présence de valeurs atypiques dans ces distributions à la fois positives et négatives. Dans les résidus du modèle 2, nous observons des valeurs atypiques positives particulièrement élevées.

Un moyen complémentaire afin de visualiser ces distributions est de les présenter sous la forme d'histogrammes.

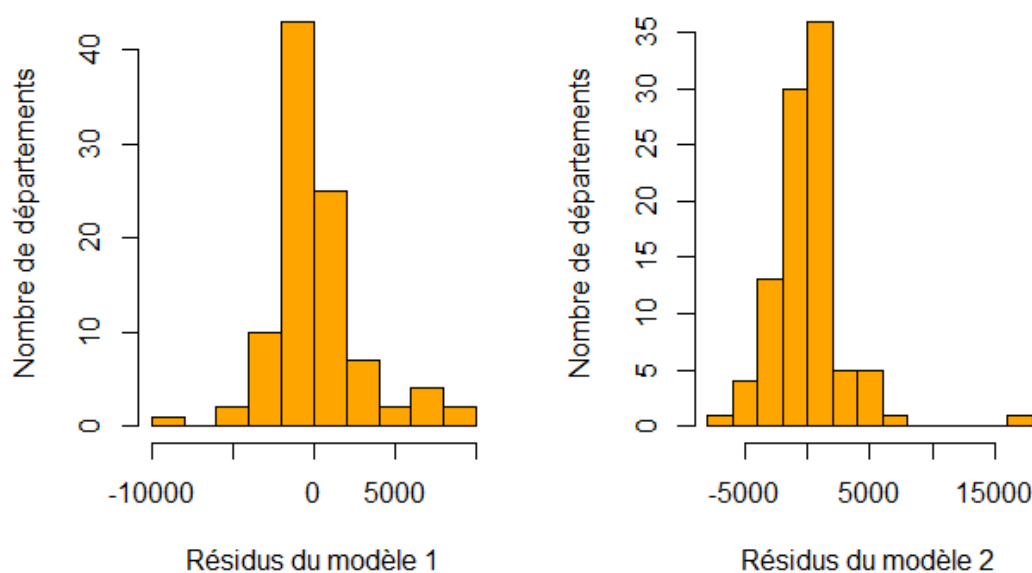


Figure 28 – Distribution des résidus des modèles 1 et 2 par des histogrammes

Si les interprétations de ces graphiques sont similaires à celles des boxplots, ils ont toutefois l'avantage de mettre en avant les classes comptabilisant le plus d'observations, ou celles au contraire qui n'en possèdent que très peu.

Nous avons ensuite analysé la symétrie et l'aplatissement des résidus respectivement grâce aux mesures Skewness et Kurtosis.

	Skewness	Kurtosis
Résidus Modèle 1	0,74	5,43
Résidus Modèle 2	1,82	12,24

Figure 29 – Skewness et Kurtosis des résidus des modèles 1 et 2

Les Skewness des résidus sont dans les deux modèles supérieurs à 0, cela signifie que dans les deux cas, la majorité des observations se situent en réalité à gauche et que les outliers sont eux à droite.

Par ailleurs, les résultats de la mesure Kurtosis sont positifs pour les résidus des deux modèles. Il y a donc de nombreuses valeurs éloignées de la moyenne dans ces distributions : elles sont dites leptokurtiques.

Afin d'évaluer la pertinence de l'ajustement des distributions empiriques des résidus, nous avons par la suite réalisé des diagrammes Quantile-Quantile.

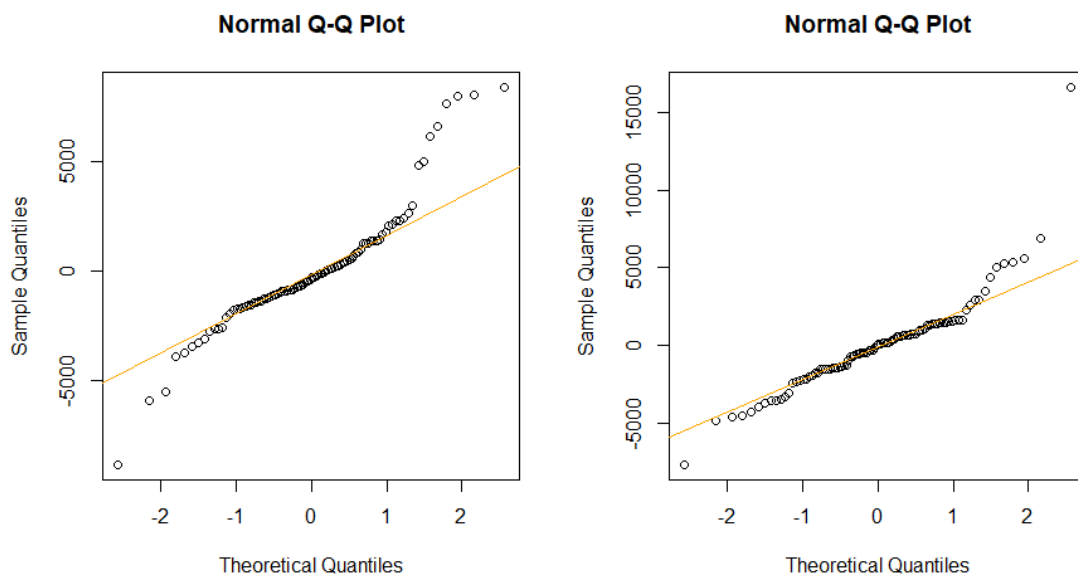


Figure 30 – Quantile-Quantile des résidus des modèles 1 et 2

Dans cet outil graphique, un modèle est considéré comme valable s'il suit la première diagonale. Dans notre cas, bien qu'une majorité des observations soient sur la diagonale, nous observons pour les valeurs les plus importantes des écarts entre les observations et la théorie, particulièrement pour les résidus du modèle 2. Il est donc difficile de donner une conclusion sur la validité des modèles avec ces graphiques.

Aussi, après avoir observé la distribution des résidus de nos modèles de manière graphique, il est donc nécessaire de réaliser des tests de Kolmogorov-Smirnov afin de tester leur normalité. En effet, ce test est à appliquer aux échantillons supérieurs à 50 observations et est donc adapté à notre analyse départementale. Les résultats obtenus dans R pour les modèles 1 et 2 sont les suivants :

```
Exact one-sample Kolmogorov-Smirnov test
data:  base$residu1
D = 0.12687, p-value = 0.08324
alternative hypothesis: two-sided
```

Figure 31 - Test de Kolmogorov-Smirnov des résidus du modèle 1

Dans le premier modèle la p-value obtenue est supérieure à 0,05. Aussi, les résidus suivent la loi normale au seuil de 5 %.

```
Exact one-sample Kolmogorov-Smirnov test
data:  base$residu2
D = 0.16424, p-value = 0.009899
alternative hypothesis: two-sided
```

Figure 32 - Test de Kolmogorov-Smirnov des résidus du modèle 2

Dans le second modèle, le résultat est inverse : la p-value est inférieure à 0,05. Cela signifie que les résidus ne suivent pas la loi normale et qu'il est nécessaire de leur apporter une correction, cependant, nous ne la réaliseront pas dans le cadre de cette analyse.

B. Relation entre la variable illustrative observée et prédite par RLM

Après avoir réalisé et examiné nos régressions linéaires multiples, nous avons souhaité connaître les relations entre la fréquentation des hébergements collectifs touristiques observée et les nuitées prédites par ces deux modèles.

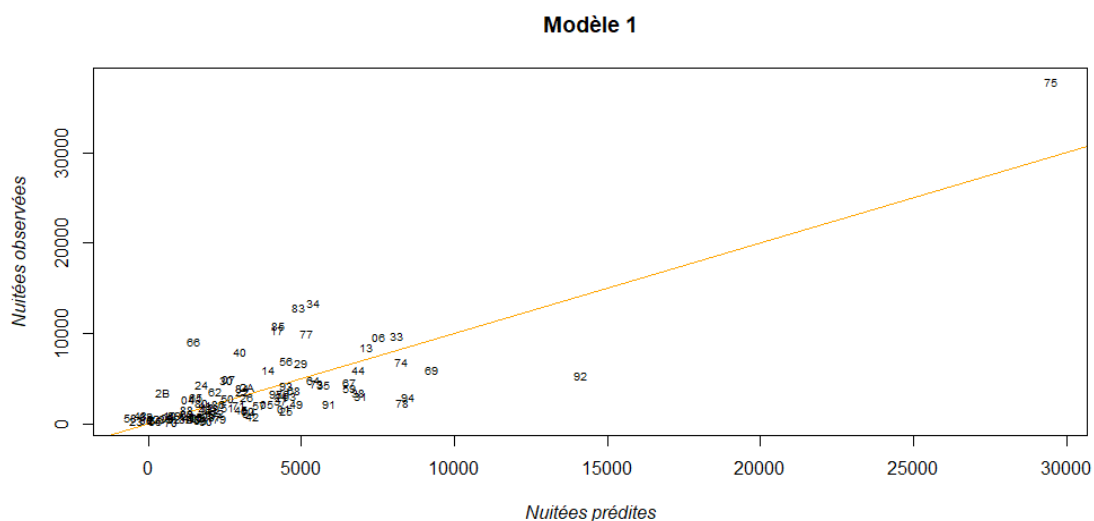


Figure 33 –Les Nuitées observées et prédites par le modèle 1

Dans le graphique ci-dessus, la droite orange représente les prédictions du modèle 1 selon la RLM. Les observations départementales du modèle 1 s'organisent autour de cette droite, cependant si certaines coïncident parfaitement : le 86 (Vienne), le 50 (Manche), le 26 (Drôme) ; d'autre s'en éloignent davantage : le 83 (Var), le 34 (Hérault) ou encore le 92 (Hauts-de-Seine). Ce graphique permet donc de visualiser les départements dans lesquels notre variable illustrative est la mieux expliquée par le modèle 1, mais également ceux pour lesquels le modèle 1 ne permet pas une prédiction juste.

Par ailleurs, pour les observations qui ne correspondent pas aux prédictions, nous constatons que les prédictions peuvent être supérieures ou inférieures à la valeur observée. Nous ne pouvons donc pas établir une relation strictement positive ou négatives entre observations et estimations dans le modèle 1.

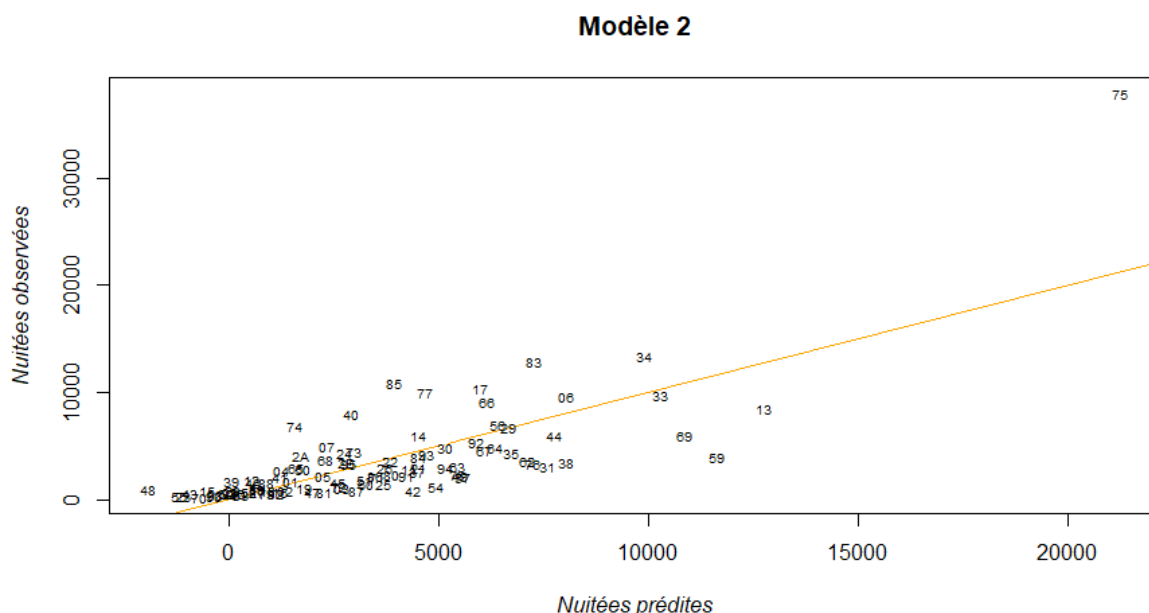


Figure 34 - Les Nuitées observées et prédites par le modèle 2

Dans le modèle 2, les départements où les prédictions correspondent aux observations sont entre autres le 01 (Ain), le 05 (Hautes-Alpes) et le 29 (Finistère). Et à l'inverse parmi ceux qui divergent, on retrouve le 85 (Vendée), le 59 (Nord) mais également et plus largement le 75 (Paris).

Cette fois aussi, il n'est pas possible de définir une relation claire entre les observations et les prédictions pour les départements s'éloignant de la droite de correspondance entre observations et prédictions.

C. Régression logistique binaire

Dans cette partie, nous avons établi cette fois-ci des régressions logistiques binaires de notre variable illustrative selon les variables latentes 1 et 2 puis 1 et 3. Au préalable, il a été nécessaire de coder la variable illustrative sous forme d'une variable binaire (0,1) et de choisir un seuil séparant en deux groupes nos départements pour lequel nos variables latentes sont significatives. Afin de choisir le seuil à retenir, nous en avons essayé plusieurs.

1. Le premier quartile

Le premier seuil que nous avons testé est le quartile 1 du nombre de nuitées dans des hébergements collectifs touristiques : 877,2.

```
Coefficients:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                   2.6564      0.6314   4.207 2.59e-05 ***
Accès_services_consommation_loisir 1.7900      0.4598   3.893 9.90e-05 ***
Pratiques_mobilité_collective    0.6485      0.4707   1.378  0.168
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 107.968  on 95  degrees of freedom
Residual deviance:  66.528  on 93  degrees of freedom
AIC: 72.528

Coefficients:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                   2.8258      0.6640   4.256 2.09e-05 ***
Accès_services_consommation_loisir 1.7732      0.4173   4.249 2.15e-05 ***
Environnement_travail          -0.8332      0.3397  -2.453  0.0142 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 107.968  on 95  degrees of freedom
Residual deviance:  61.516  on 93  degrees of freedom
AIC: 67.516
```

Figure 35 - Régressions logistiques binaires des deux modèles pour un seuil de 877,2

Pour ce seuil, nous observons que la variable latente Accès aux services de consommation et de loisir et les constantes sont significatives dans les deux modèles au seuil de 1 %, c'est également le cas de l'Environnement propice au travail dans le second modèle. Cependant, la variable latente Pratiques de mobilité collective n'est pas significative dans le premier modèle. Aussi, nous ne retenons pas le premier quartile comme seuil et n'interprétons pas ces résultats.

2. La médiane

Le second seuil testé est la médiane des nuitées dans des hébergements collectifs touristiques : 2 281,5.

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.4020    0.2997   1.342  0.1797
Accès_services_consommation_loisir  1.3221    0.2838   4.659 3.18e-06 ***
Pratiques_mobilité_collective      0.8800    0.3507   2.509  0.0121 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 133.084  on 95  degrees of freedom
Residual deviance:  88.635  on 93  degrees of freedom
AIC: 94.635

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.3633    0.3045   1.193 0.232841
Accès_services_consommation_loisir  1.3249    0.2704   4.900 9.58e-07 ***
Environnement_travail      -1.1730    0.3156  -3.717 0.000202 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 133.084  on 95  degrees of freedom
Residual deviance:  75.584  on 93  degrees of freedom
AIC: 81.584
```

Figure 36 - Régressions logistiques binaires des deux modèles pour un seuil de 2 281,5

Nous observons, pour ce nouveau seuil, que l'intégralité des variables latentes est significative au seuil de 1 % dans les deux modèles. Seulement les AIC qui mesurent la qualité des modèles sont importants, nous allons donc chercher à les diminuer tout en conservant des variables latentes significatives.

3. Le seuil optimal

Par tâtonnement, nous avons cherché la première valeur de seuil entre le Q1 et la médiane dans laquelle toutes les variables latentes étaient significatives dans les deux modèles. Nous avons ainsi choisi la valeur 879 que nous considérerons comme notre seuil optimal.

```

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      2.8051    0.6704   4.184 2.86e-05 ***
Accès_services_consommation_loisir  2.0233    0.5045   4.011 6.06e-05 ***
Pratiques_mobilité_collective      1.0117    0.5105   1.982  0.0475 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 110.111  on 95  degrees of freedom
Residual deviance:  64.882  on 93  degrees of freedom
AIC: 70.882

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      2.7143    0.6392   4.247 2.17e-05 ***
Accès_services_consommation_loisir  1.7648    0.4084   4.322 1.55e-05 ***
Environnement_travail      -0.8300    0.3349  -2.478  0.0132 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 110.111  on 95  degrees of freedom
Residual deviance:  62.334  on 93  degrees of freedom
AIC: 68.334

```

Figure 37 - Régressions logistiques binaires des deux modèles pour un seuil de 879

Avec notre seuil optimal, où la significativité des variables latentes est respectée au seuil de 5 %, nous pouvons enfin interpréter ces régressions logistiques binaires.

Le premier modèle nous indique que la probabilité pour un département d'avoir un nombre de nuitées dans des hébergements collectif de tourisme supérieur à 879 sera 7,56 ($= e^{2,0233}$) fois supérieure pour les départements ayant un accès important aux services de consommations et de loisir, et 2,75 ($= e^{1,0117}$) fois supérieure pour ceux ayant des pratiques de mobilités collectives développées.

Dans le second modèle, cette probabilité sera 5,84 ($= e^{1,7648}$) fois supérieure pour les départements ayant un accès important aux services de consommations et de loisir, et 0,44 ($= e^{-0,83}$) fois inférieure pour ceux ayant un environnement propice au travail.

V. Conclusion et discussion des résultats

A. Conclusion

Réaliser cette ACP, nous aura permis de comprendre la dynamique qui existe entre nos différentes variables explicatives. En effet, en analysant leurs caractéristiques, puis en les projetant dans les plans des dimensions 1 et 2 ainsi que 1 et 3, nous sommes parvenues à définir des variables latentes cohérentes avec les répartitions observées de ces variables sur les cercles de corrélation. Si nous avons choisi de travailler avec ces trois axes, c'est qu'ils expliquent à eux seuls 77,23 % de la variance totale. Pour les axes 1, 2 et 3, nous avons établi nos facteurs synthétiques comme étant respectivement, l'accès à des services de consommation et de loisir, les pratiques de mobilité collectives et la mesure d'un environnement propice au travail. Si ces variables latentes sont pertinentes au niveau des variables, il a ensuite été nécessaire de vérifier si elles l'étaient également au niveau des individus. Nous avons obtenu cette confirmation avec des départements comme Paris à la fois lieu de prédilection dans l'accès aux services et dans l'utilisation des transports en commun, mais surtout positionné dans le cadran I du plan des dimensions 1 et 2, ou encore la Haute-Corse, département insulaire appartenant au cadran III du plan des dimensions 1 et 3 qui possède une offre moindre d'accès à des services commerciaux ou de loisir et dont le cadre est plus propice à du tourisme qu'à du travail,

L'intérêt de la définition des axes est d'analyser comment notre variable illustrative, la fréquentation des hébergements collectifs de tourisme, y évolue. En la projetant sur les cercles de corrélations, nous avons constaté qu'elle augmentait avec l'accès à des services de consommation et de loisir ainsi qu'avec les pratiques de mobilité collectives, mais qu'elle diminuait lorsque que l'environnement devient plus propice au travail.

Par la suite, nous avons testé la qualité de ces corrélations en réalisant des régressions, linéaires multiples dans un premier temps puis logistiques binaires. Avec toutes ses composantes significatives au seuil de 1 %, un R^2 ajusté de 63 % et des résidus suivant la loi normale au seuil de 5 %, c'est le modèle comprenant les variables latentes 1 et 2 qui s'est distingué dans la RLM. Mais dans la régression logistique binaire, avec un seuil optimal déterminé de 879 nuitées, il s'agit du modèle impliquant les variables latentes 1 et 3 qui s'est avéré de meilleure qualité, présentant un AIC légèrement inférieur et des composantes significatives au seuil de 5 %. Ces

résultats présentent nos variables latentes et plus globalement les modèles établis comme significatifs dans l'explication de notre variable latente. Nous sommes donc satisfaites des résultats obtenus.

B. Discussion

Dans cette dernière partie, nous nous pencherons sur les diverses problématiques et limites rencontrées au cours de notre projet, tout en partageant nos impressions et nos idées.

Commençons par la base de notre analyse : la base de données. Suite au choix de notre sujet, nous avons pris l'initiative de constituer nous-même notre jeu de données. Toutefois, notre quête de données de qualité a rapidement mis en lumière certaines complexités. Nous avons constaté que l'accès à des données actuelles et fiables était difficile, voire parfois impossible, en raison de leur obsolescence ou de leur absence totale. En dépit de ces obstacles, nous avons réussi à former une base de données constituée d'informations allant de 2019 à 2023.

Dans le processus de délimitation de notre étude, nous avons pris la décision de nous concentrer uniquement sur les départements métropolitains, excluant ainsi les départements d'outre-mer (DOM). Cette décision a été principalement motivée par des considérations de distance et d'accessibilité. De plus, en ce qui concerne les données de transport, nous avons choisi de nous limiter à l'infrastructure routière. Les autres moyens de transport, tels que les aéroports ou les ports, ne pouvaient pas être représentés efficacement au niveau départemental, ce qui a influencé notre choix.

Considérant l'importance économique liée à l'analyse des données, nous avons souhaité intégrer des données économiques supplémentaires, que nous jugeons pertinentes. Cependant, nous avons été confrontés à un obstacle notable : tous les départements ne divulguent pas ces informations.

Après avoir constitué notre première base de données, nous avons réalisé des tests initiaux qui ont donné des résultats convenables. Toutefois, dans un esprit d'amélioration, nous avons cherché des combinaisons de données susceptibles d'optimiser les résultats en termes d'inertie. Ce processus nous a amenés à explorer des variables qui, finalement,

n'ont pas été incluses dans la base de données finale. Parmi celles-ci figuraient le taux de boisement, la présence de communes touristiques et les unités urbaines.

Au fil de cette réflexion et des tests, nous avons également décidé d'écarter certaines variables, telles que la superficie et la population, au profit de la densité. Cette étape d'ajustement et de réévaluation des variables a constitué une période longue et exigeante, nécessitant une réflexion considérable, mais s'est avérée être un processus essentiel pour affiner notre projet.

Concernant nos individus, la diversité des caractéristiques a marqué notre étude. Des disparités notables existent même au sein des départements. Ces variations internes ont complexifié notre analyse, mettant en évidence la nécessité de reconnaître les spécificités de chaque contexte départemental.

Pour approfondir notre analyse, l'examen des données relatives aux logements particuliers proposés sur des plateformes comme Airbnb, aujourd'hui un acteur incontournable du marché, représenterait une piste intéressante. Cette exploration pourrait offrir des perspectives nouvelles sur l'évolution du secteur de l'hébergement touristique.

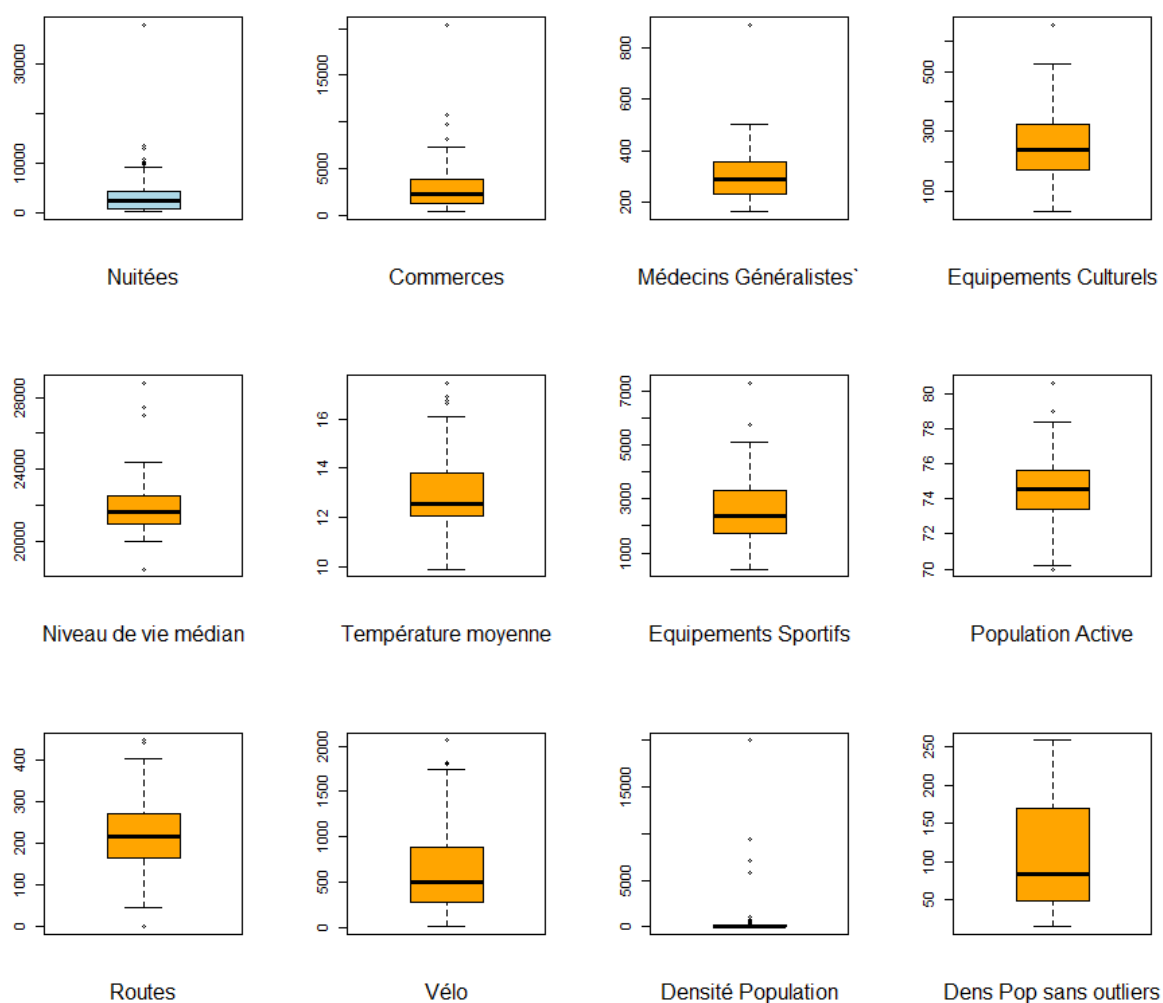
Dans notre réflexion, nous avons jugé paradoxal que malgré la crise du COVID-19, la reprise s'oriente fortement vers nos anciennes habitudes. Bien qu'il y ait une conscience collective croissante pour un tourisme plus responsable, il semble que, en tant que société, nous continuons majoritairement à percevoir le tourisme davantage comme un bien de consommation que comme un vecteur de changement.

Toutefois, nous avons également appris qu'il existe des distinctions importantes dans le domaine du tourisme. « Tous les voyages ne sont pas touristiques, toutes les pratiques touristiques ne résultent pas d'un voyage, et tout ce qui est agréable ou lié au loisir n'est pas nécessairement du tourisme ».

³³ Le tourisme : acteurs, lieux et enjeux » M. Stock. Belin, Paris 2003, p10 (consulté le 08/11 /2023)

VI. Annexes

A. Tableaux et graphiques supplémentaires



Annexe 1 - Distribution des variables

Test Method:		Rosner's Test for Outliers						
Data:		base\$Nuitées						
Number of Outliers Detected:		3						
i	Mean.i	SD.i	Value	Obs.Num	R.i+1	lambda.i+1	Outlier	
1 0	3572.969	4623.182	37850	76	7.414165	3.370097	TRUE	
2 1	3212.158	2995.084	13424	35	3.409535	3.366490	TRUE	
3 2	3103.521	2816.691	12898	84	3.477300	3.362836	TRUE	
Data:		base\$Commerces						
Number of Outliers Detected:		3						
i	Mean.i	SD.i	Value	Obs.Num	R.i+1	lambda.i+1	Outlier	
1 0	3056.771	2729.040	20366	76	6.342608	3.370097	TRUE	
2 1	2874.568	2075.091	10766	13	3.802933	3.366490	TRUE	
3 2	2790.617	1917.175	9690	60	3.598724	3.362836	TRUE	
Data:		base\$`Médecins Généralistes`						
Number of Outliers Detected:		1						
i	Mean.i	SD.i	Value	Obs.Num	R.i+1	lambda.i+1	Outlier	
1 0	304.4167	97.89325	888	76	5.961426	3.370097	TRUE	
Data:		base\$`Equipements Culturels`						
Number of Outliers Detected:		2						
i	Mean.i	SD.i	Value	Obs.Num	R.i+1	lambda.i+1	Outlier	
1 0	251.7500	119.1543	655	76	3.384267	3.370097	TRUE	
2 1	247.5053	112.2527	654	60	3.621245	3.366490	TRUE	
Data:		base\$`Niveau de vie médian`						
Number of Outliers Detected:		4						
i	Mean.i	SD.i	Value	Obs.Num	R.i+1	lambda.i+1	Outlier	
1 0	22052.19	1649.147	28810	93	4.097762	3.370097	TRUE	
2 1	21981.05	1502.556	28790	76	4.531575	3.366490	TRUE	
3 2	21908.62	1333.467	27470	79	4.170619	3.362836	TRUE	
4 3	21848.82	1207.337	27030	75	4.291414	3.359136	TRUE	
Data:		base\$`Equipements Sportifs`						
Number of Outliers Detected:		1						
i	Mean.i	SD.i	Value	Obs.Num	R.i+1	lambda.i+1	Outlier	
1 0	2627.562	1281.139	7298	60	3.645536	3.370097	TRUE	
Data:		base\$`Densité Population`						
Number of Outliers Detected:		10						
i	Mean.i	SD.i	Value	Obs.Num	R.i+1	lambda.i+1	Outlier	
1 0	565.50417	2391.21408	19949.2	76	8.106215	3.370097	TRUE	
2 1	361.46526	1318.92227	9350.3	93	6.815288	3.366490	TRUE	
3 2	265.83936	938.19516	7124.5	94	7.310484	3.362836	TRUE	
4 3	192.09032	610.70611	5822.7	95	9.219835	3.359136	TRUE	
5 4	130.88804	157.73563	1022.8	96	5.654474	3.355387	TRUE	
6 5	121.08681	127.36122	729.4	92	4.776283	3.351588	TRUE	
7 6	114.32778	110.44968	639.8	79	4.757571	3.347738	TRUE	
8 7	108.42360	95.73277	591.2	70	5.042959	3.343836	TRUE	
9 8	102.93750	80.99791	453.9	60	4.332982	3.339880	TRUE	
10 9	98.90345	72.02931	406.7	13	4.273212	3.335869	TRUE	

Annexe 2 – Identification des outliers par des tests de Rosner

(les numéros correspondent ici au positionnement dans la liste : Haute-Corse = 21)

	Min	1st Qu..25%	Médiane	Moyenne	3rd Qu..75%	Max	Ecart type
Nuitées	243	876	2222	3212	4302	13424	2995
Commerces	435	1326	2315	2875	3809	10766	2075
Médecins Généralistes	165	235	287	298	356	503	78
Equipements Culturels	32	173	236	248	321	654	112
Niveau de vie	18470	20990	21630	21981	22495	28810	1504
Température moyenne	10	12	13	13	14	17	2
Equipements Sportifs	400	1724	2370	2630	3316	7298	1288
Population Active	70	73	74	74	76	81	2
Routes	0	168	218	218	270	446	92
Vélo	7	284	515	623	878	2064	469
Densité Population	15	49	83	361	167	9350	1319

Annexe 3 – Principales statistiques des variables sans Paris

	Valeurs propres	% de variance	% de variance cumulée
Dim 1	3,97	39,68	39,68
Dim 2	1,83	18,34	58,01
Dim 3	1,5	15,01	73,03
Dim 4	0,85	8,58	81,56
Dim 5	0,66	6,56	88,12
Dim 6	0,45	4,52	92,64
Dim 7	0,27	2,67	95,3
Dim 8	0,21	2,11	97,41
Dim 9	0,14	1,43	98,85
Dim 10	0,12	1,15	100

Annexe 4 - Valeurs propres sans Paris

B. Base de données

Code	Département	Nuitées	Commerces	Médecins Généralistes	Equipements Culturels	Niveau de vie	Température moyenne	Equipements Sportifs	Population Active	Routes	Vélo	Densité Population
01	Ain	1685	2355	174	303	24030	12,29	3193	78,4	232	706	116,6
02	Aisne	879	1789	211	211	20300	11,40	2640	72,6	311	337	71,0
03	Allier	1131	1642	275	287	20990	12,58	2091	73,1	331	367	45,3
04	Alpes-de-Haute- Provence	2720	1029	291	133	21130	14,19	1062	74	165	164	24,1
05	Hautes-Alpes	2222	1116	503	139	21420	12,13	2206	76,2	182	98	25,2
06	Alpes-Maritimes	9667	7116	461	285	22630	16,77	2554	74,9	75	495	258,3
07	Ardèche	4919	1566	235	266	21450	14,69	1688	75,1	91	610	60,1
08	Ardenne	454	1045	241	121	20310	10,69	1532	72,8	129	519	50,8
09	Ariège	736	778	264	104	20430	13,04	1158	74,2	106	201	31,6
10	Aube	996	1348	229	201	20960	12,12	1224	73,4	185	414	52,1
11	Aude	2785	1903	288	284	19980	15,66	2209	72,1	155	317	61,9
12	Aveyron	1850	1555	255	236	21260	11,79	1790	75,4	184	84	32,0
13	Bouches-du-Rhône	8465	10766	453	406	22210	15,87	4660	71	446	1145	406,7
14	Calvados	5947	3579	386	227	22180	11,66	3089	73,7	253	1143	126,6
15	Cantal	827	840	287	165	21140	11,18	1288	75,9	184	131	25,1
16	Charente	702	1477	251	130	21410	14,41	1566	75,2	215	232	58,9
17	Charente-Maritime	10334	3733	326	320	22080	14,17	3350	74,2	281	1278	97,0
18	Cher	798	1223	197	206	21560	13,11	1860	74,5	205	395	41,1
19	Corrèze	1095	1304	288	175	21590	14,03	1894	74,4	179	90	40,5
2A	Corse-du-Sud	4027	1447	350	39	22350	16,92	400	75,4	0	37	40,8
2B	Haute-Corse	3487	1399	299	32	20740	17,49	561	70,2	0	98	40,1
21	Côte-d'Or	2937	2243	418	311	22940	12,39	2267	74,4	316	764	61,2

22	Côtes-d'Armor	3539	2688	261	331	21850	12,18	3248	74,4	269	1154	88,4
23	Creuse	297	495	219	127	20130	11,73	862	73,2	93	7	20,3
24	Dordogne	4398	2113	244	299	20830	13,55	2090	73,8	233	336	45,4
25	Doubs	1399	2315	358	300	23260	12,28	2405	75	228	795	105,1
26	Drôme	3001	2569	295	221	21790	15,05	2536	75,7	275	627	80,1
27	Eure	744	2253	165	179	22240	11,90	2740	75,6	294	515	99,3
28	Eure-et-Loir	785	1632	194	170	22650	12,19	2205	76,5	289	240	73,0
29	Finistère	6743	3884	367	372	22400	12,52	4653	73,3	234	1155	137,5
30	Gard	4816	3734	310	327	20740	13,07	3016	71,6	220	668	129,4
31	Haute-Garonne	3060	5839	413	336	23730	14,63	4996	75,5	335	98	233,0
32	Gers	599	900	225	69	21420	14,18	1424	75,6	237	110	30,8
33	Gironde	9755	7331	437	508	23180	14,92	4770	75,3	380	914	169,6
34	Hérault	13424	6275	440	403	21130	16,13	4012	71,4	269	633	202,1
35	Ille-et-Vilaine	4365	4417	365	426	22840	12,49	3282	75	313	1739	165,1
36	Indre	563	946	197	142	20820	12,93	1975	74,3	135	312	31,6
37	Indre-et-Loire	2542	2351	391	235	22450	13,20	2987	74,9	216	1106	100,9
38	Isère	3441	5416	337	524	23580	12,65	4709	75,5	366	1497	174,2
39	Jura	1656	1266	246	126	22460	12,48	1541	76	187	385	51,4
40	Landes	7975	2212	295	199	22100	14,85	1814	75,6	252	519	46,4
41	Loir-et-Cher	2065	1356	231	182	21940	12,51	1562	75,2	212	1180	51,5
42	Loire	827	3407	345	325	21380	12,38	2859	73,4	235	451	161,6
43	Haute-Loire	609	1081	244	212	21470	9,89	1540	75,6	185	98	45,6
44	Loire-Atlantique	5967	5431	355	369	23430	13,12	5111	76,4	322	2064	217,8
45	Loiret	1618	2490	231	242	22480	12,35	2980	75,7	265	1180	101,3
46	Lot	1548	941	260	191	21310	13,65	1098	75,3	110	71	33,6
47	Lot-et-Garonne	714	1610	250	171	20550	14,40	1948	74,4	179	472	61,2
48	Lozère	1005	435	234	81	20940	10,45	912	74,5	220	22	14,8

49	Maine-et-Loire	2194	2889	350	336	21790	13,28	4599	76	240	1142	116,5
50	Manche	2832	2524	254	212	21740	12,17	2245	75,4	200	740	82,8
51	Marne	1827	2350	367	196	22230	11,53	2782	73,7	379	598	69,1
52	Haute-Marne	573	735	233	182	20780	11,87	1122	73,8	251	98	27,0
53	Mayenne	435	1170	191	187	21510	12,52	2089	76,6	202	769	59,0
54	Meurthe-et-Moselle	1167	2580	405	239	22400	11,72	3277	71,6	262	922	139,1
55	Meuse	299	697	200	116	21320	11,59	974	74	135	175	28,7
56	Morbihan	6924	3691	334	336	22270	13,21	3930	73,7	248	1716	113,9
57	Moselle	2060	4159	264	272	22350	11,70	5760	73,4	310	1435	169,1
58	Nièvre	660	898	221	130	20940	12,30	808	70,9	172	406	29,0
59	Nord	3943	9690	357	654	20750	12,20	7298	71,3	378	353	453,9
60	Oise	1484	2762	209	348	22680	11,69	3742	74,9	294	873	142,2
61	Orne	417	1296	235	114	20800	11,94	1589	73,2	185	407	44,7
62	Pas-de-Calais	3534	5984	253	390	20090	11,80	4739	71	442	1400	217,9
63	Puy-de-Dôme	3039	2809	373	371	22510	13,28	3556	73,7	218	421	84,0
64	Pyrénées-Atlantiques	4869	3716	429	255	22550	14,89	3747	75,3	281	706	91,3
65	Hautes-Pyrénées	2941	1242	343	139	20990	13,24	1204	72,8	100	98	51,7
66	Pyrénées-Orientales	9120	2534	337	229	20070	16,65	2370	70	193	751	119,2
67	Bas-Rhin	4645	4664	408	374	23330	12,42	4000	75,2	192	614	245,7
68	Haut-Rhin	3684	3201	302	189	23760	12,00	3056	76	115	519	218,2
69	Rhône	5973	8212	463	493	23690	13,92	4382	74,7	252	1800	591,2
70	Haute-Saône	243	980	223	134	21260	11,54	1529	74,8	182	98	43,2
71	Saône-et-Loire	2015	2504	268	323	21520	12,86	2789	75	271	883	63,8
72	Sarthe	1220	2089	235	192	21630	13,17	4013	75	242	576	91,4

73	Savoie	4526	2888	388	266	23630	12,07	1917	77,4	236	739	74,3
74	Haute-Savoie	6894	4277	320	291	27030	10,50	2865	80,6	200	98	196,5
75	Paris	37850	20366	888	655	28790	13,66	2420	78,3	1	475	19949,2
76	Seine-Maritime	3298	5228	333	434	21700	12,03	4591	72,9	379	1131	199,8
77	Seine-et-Marne	9984	5007	196	389	24000	12,44	3938	76,8	403	1106	245,6
78	Yvelines	2353	5047	267	390	27470	12,24	3916	76,8	252	1179	639,8
79	Deux-Sèvres	611	1288	221	199	21590	13,64	2097	77	192	474	62,3
80	Somme	2341	2263	360	244	20980	11,68	2865	71,8	322	609	91,4
81	Tarn	747	1838	287	132	21080	14,26	2287	74,1	147	432	68,5
82	Tarn-et-Garonne	521	1191	249	124	20860	14,63	1408	74,4	111	284	71,5
83	Var	12898	6487	324	247	22320	16,77	3020	73	200	620	185,9
84	Vaucluse	3907	3261	319	230	20640	15,55	1960	73,8	111	816	158,0
85	Vendée	10832	3052	221	313	22040	13,47	3431	76,9	164	1811	105,6
86	Vienne	2171	1577	352	244	21580	13,14	2051	72,5	220	485	63,2
87	Haute-Vienne	872	1556	395	200	21610	12,62	1974	72,5	264	98	67,0
88	Vosges	1631	1842	248	184	20900	11,87	2018	74,3	201	343	60,6
89	Yonne	948	1392	212	167	21390	12,52	1759	74,6	274	470	44,3
90	Territoire de Belfort	296	523	331	64	22370	11,59	715	73,8	46	284	224,6
91	Essonne	2186	4043	238	322	24410	12,99	3636	76	142	970	729,4
92	Hauts-de-Seine	5294	5614	381	279	28810	12,32	2722	79	55	450	9350,3
93	Seine-Saint-Denis	4239	5947	237	239	18470	12,91	2031	72,8	59	495	7124,5
94	Val-de-Marne	2999	4759	384	255	23540	12,99	2391	75,9	80	258	5822,7
95	Val-d'Oise	3391	3988	228	241	22650	12,52	3077	75,1	118	566	1022,8

C. Script R

#Librairies

```
library(readxl)
```

```
library(FactoMineR)
```

```
library(readxl)
```

```
library(gridExtra)
```

```
library(corrplot)
```

```
library(factoextra)
```

```
library(RColorBrewer)
```

```
library(ggplot2)
```

```
library(tidyverse)
```

```
library(outliers)
```

```
library(reshape2)
```

```
library(EnvStats)
```

#Recuperation base de données

```
base=read_excel("C:/Users/Marie/Documents/Master/M1/Rendu/Excel ADD -  
Tourisme.xlsx",col_names=T)
```

```
View(base)
```

```
str(base)
```

#Création et segmentation de la base de données

```
base1 = base[,3:13] # avec la variable illustrative
```

```
View(base1)
```

```
base2 = base[, 4:13] #sans la variable illustrative
```

```
View(base2)
```

```
summary(base1)
```

#ACP avec la variable illustrative

```
res.pca1=PCA(base1)
```

```
res.pca1
```

ACP sans la variable illustrative

```
res.pca2=PCA(base2)
```

```
res.pca2
```

II – Analyse du sujet

#Plots: variables illustrative et chaque variable explicatives

```
plot(base$Commerces, base$Nuitées, type = "p", col=0, main = "Corrélation entre le  
nombre de nuitées et le nombre de commerces ",
```

```
      xlab = "Nombre de Commerces", ylab = "Nombre de Nuitées", font.lab=3)
```

```
abline(lm(base$Nuitées~base$Commerces), col="orange", lwd = 2)
```

```
text(base$Commerces, base$Nuitées, base$Code, col = "black", cex = .6)
```

```
cor(base$Nuitées,base$Commerces)
```

```
plot(base$`Médecins Généralistes`, base$Nuitées, type = "p", col=0, main = " Corrélation  
entre le nombre de nuitées et l'effectif de médecins généralistes ",
```

```
      xlab = "Effectif de Médecins Généralistes", ylab = "Nombre de Nuitées", font.lab=3)
```

```
abline(lm(base$Nuitées~base$`Médecins Généralistes`), col="orange", lwd = 2)
```

```
text(base$`Médecins Généralistes`, base$Nuitées, base$Code, col = "black", cex = .6)
```

```
cor(base$Nuitées,base$`Médecins Généralistes`)
```

```
plot(base$`Equipements Culturels`, base$Nuitées, type = "p", col=0, main= "Corrélation  
entre le nombre de nuitées et le nombre d'équipements culturels",
```

```

xlab = "Nombre d'équipements Culturels", ylab = "Nombre de Nuitées", font.lab=3)
abline(lm(base$Nuitées~base$`Equipements Culturels`), col="orange", lwd=2)
text(base$`Equipements Culturels`, base$Nuitées, base$Code, col="black", cex=.6)
cor(base$Nuitées,base$`Equipements Culturels`)

```

```

plot(base$`Niveau de vie`, base$Nuitées, type = "p", col=0, main= "Corrélation entre le
nombre de nuitées et le niveau de vie",

```

```

xlab = "Niveau de Vie (en €)", ylab = "Nombre de Nuitées", font.lab=3)
abline(lm(base$Nuitées~base$`Niveau de vie`), col="orange", lwd=2)
text(base$`Niveau de vie`, base$Nuitées, base$Code, col="black", cex=.6)
cor(base$Nuitées,base$`Niveau de vie`)

```

```

plot(base$`Température moyenne`, base$Nuitées, type = "p", col=0, main = "Corrélation
entre le nombre de nuitées et la température ",

```

```

xlab = "Température (en °C)", ylab = "Nombre de Nuitées", font.lab=3)
abline(lm(base$Nuitées~base$`Température moyenne`), col="orange", lwd = 2)
text(base$`Température moyenne`, base$Nuitées, base$Code, col = "black", cex = .6)
cor(base$Nuitées,base$`Température moyenne`)

```

```

plot(base$`Equipements Sportifs`, base$Nuitées, type = "p", col=0, main= "Corrélation
entre le nombre de nuitées et le nombre d'équipements sportifs",

```

```

xlab = "Nombre d'équipements Sportifs", ylab = "Nombre de Nuitées", font.lab=3)
abline(lm(base$Nuitées~base$`Equipements Sportifs`), col="orange", lwd=2)
text(base$`Equipements Sportifs`, base$Nuitées, base$Code, col="black", cex=.6)
cor(base$Nuitées,base$`Equipements Sportifs`)

```

```

plot(base$`Population Active`, base$Nuitées, type = "p", col=0, main= "Corrélation entre
le nombre de nuitées et la population active",

```

```

xlab = "Population Active 15 - 64 ans (en %)", ylab = "Nombre de Nuitées", font.lab=3)
abline(lm(base$Nuitées~base$`Population Active`), col="orange", lwd=2)
text(base$`Population Active`, base$Nuitées, base$Code, col="black", cex=.6)
cor(base$Nuitées,base$`Population Active`)

plot(base$Routes, base$Nuitées, type = "p", col=0, main= "Corrélation entre le nombre
de nuitées et l'infrastructure routière",

xlab = "Infrastructure Routière (en km)", ylab = "Nombre de Nuitées", font.lab=3)
abline(lm(base$Nuitées~base$Routes), col="orange", lwd=2)
text(base$Routes, base$Nuitées, base$Code, col="black", cex=.6)
cor(base$Nuitées,base$Routes)

plot(base$Vélo, base$Nuitées, type = "p", col=0, main = "Corrélation entre le nombre de
nuitées et les aménagements cyclables",

xlab = "Aménagements Cyclables (en km)", ylab = "Nombre de Nuitées", font.lab=3)
abline(lm(base$Nuitées~base$Vélo), col="orange", lwd = 2)
text(base$Vélo, base$Nuitées, base$Code, col = "black", cex = .6)
cor(base$Nuitées,base$Vélo)

plot(base$`Densité Population`, base$Nuitées, type = "p", col=0, main = " Corrélation
entre le nombre de nuitées et la densité de la population",

xlab = "Densité de la population",

ylab = "Nombre de Nuitées", font.lab=3)
abline(lm(base$Nuitées~base$`Densité Population`), col="orange", lwd = 2)
text(base$`Densité Population`, base$Nuitées, base$Code, col = "black", cex = .6)
cor(base$Nuitées,base$`Densité Population`)

```

#Boxplot

```

par(mfrow=c(3,4))

par(cex.axis=1.9, cex.lab=1)

boxplot(base$Nuitées, xlab = "Nuitées", col="lightblue")

boxplot(base$Commerces, xlab = "Commerces", col="orange")

boxplot(base$`Médecins Généralistes`, xlab = "Médecins Généralistes", col="orange")

boxplot(base$`Equipements Culturels`, xlab = "Equipements Culturels", col="orange")

boxplot(base$`Niveau de vie`, xlab = "Niveau de vie", col="orange")

boxplot(base$`Température moyenne`, xlab = "Température", col="orange")

boxplot(base$`Equipements Sportifs`, xlab = "Equipements Sportifs", col="orange")

boxplot(base$`Population Active`, xlab = "Population Active", col="orange")

boxplot(base$Routes, xlab = "Routes", col="orange")

boxplot(base$Vélo, xlab = "Vélo", col="orange")

boxplot(base$`Densité Population`, xlab = "Densité Population", col="orange")

boxplot(base$`Densité Population`, xlab = "Dens Pop sans outliers", col="orange", outline
= FALSE)

```

Statistiques des toutes les variables

```

desc_stats <- sapply(base1, function(col) {

  c(

    Min = formatC(min(col, na.rm = TRUE), format = "f", digits = 0),

    `1st Qu.` = formatC(quantile(col, 0.25, na.rm = TRUE), format = "f", digits = 0),

    Médiane = formatC(median(col, na.rm = TRUE), format = "f", digits = 0),

    Moyenne = formatC(mean(col, na.rm = TRUE), format = "f", digits = 0),

    `3rd Qu.` = formatC(quantile(col, 0.75, na.rm = TRUE), format = "f", digits = 0),

    Max = formatC(max(col, na.rm = TRUE), format = "f", digits = 0),

```

```

`Ecart type` = formatC(sd(col, na.rm = TRUE), format = "f", digits = 0)
)
})

desc_stats_df <- as.data.frame(t(desc_stats))

View(desc_stats_df)

```

Test Rosner pour détecter les outliers

```

rosnerTest(base$Nuitées, k=4, alpha=0.05)

rosnerTest(base$Commerces, k=4, alpha=0.05)

rosnerTest(base$`Médecins Généralistes`, k=2, alpha = 0.05)

rosnerTest(base$`Equipements Culturels`, k=3, alpha=0.05)

rosnerTest(base$`Niveau de vie`, k=5, alpha=0.05)

rosnerTest(base$`Température moyenne`, k=1, alpha=0.05)

rosnerTest(base$`Equipements Sportifs`, k=2, alpha=0.05)

rosnerTest(base$`Population Active`, k=1, alpha=0.05)

rosnerTest(base$Routes, k=1, alpha=0.05)

rosnerTest(base$ Vélo, k=1, alpha=0.05)

rosnerTest(base$`Densité Population`, k=10, alpha=0.05)

```

III – ACP

Création image des valeurs propres

```

round(res.pca2$eig,2)

Valeurs_propres<- round(res.pca2$eig,2)

jpeg("Valeurs_propres.jpg", height=25*nrow(Valeurs_propres),
width=165*ncol(Valeurs_propres))

grid.table(Valeurs_propres)

```

```
dev.off()
```

```
getwd()
```

Graphique Valeurs Propres

```
fviz_eig(res.pca2, addlabels = TRUE, ylim=c(0,45),
```

```
  barfill = "orange",
```

```
  barcolor = "orange",
```

```
  ylab = "Pourcentage de la variance expliquée ",
```

```
  addplot = TRUE) + # courbe cumulative
```

```
  geom_vline(xintercept = 3.5, linetype = 2, linecolor="black") # Ligne d'accentuation  
après la troisième composante
```

#Définition de coordonnées, cosinus carrés et contributions des individus

```
cordonees_var <- round(res.pca2$var$coord,2)
```

```
cosinus2_var <- round(res.pca2$var$cos2,2)
```

```
contribution_var <- round(res.pca2$var$contrib,2)
```

#Coordonnées, cosinus carrés et contributions des variables

```
par(mfrow=c(1,1))
```

```
corrplot(cordonees_var[,1:3], col = "black" ,
```

```
  is.corr=FALSE, method = "number",
```

```
  tl.srt=45, tl.col="black",
```

```
  #col=brewer.pal(n=3, name="OrRd"),
```

```
  addCoef.col="black")
```

```
corrplot(cosinus2_var[,1:3],
```

```
  is.corr=FALSE,
```

```

    tl.srt=45, tl.col="black",

    col=brewer.pal(n=3, name="OrRd"),

    addCoef.col="black")

corrplot(contribution_var[,1:3],

    is.corr=FALSE,

    tl.srt=45,

    tl.col="black",

    col=brewer.pal(n=3, name="OrRd"),

    addCoef.col="black")

# ACP axes 1, 2 - Sans la variable illustrative

par(mfrow=c(1,2))

fviz_pca_var(res.pca2, axes=c(1,2),

    col.var = "cos2",

    gradient.cols = c("#1111fa", "#fae311", "#FF0000"),

    repel = TRUE)+ labs(title = "Analyse en composantes principales sur les axes 1 et
2")

# ACP axes 1, 3 - Sans la variable illustrative

fviz_pca_var(res.pca2, axes=c(1,3),

    col.var = "cos2",

    gradient.cols = c("#1111fa", "#fae311", "#FF0000"),

    repel = TRUE)+ labs(title = "Analyse en composantes principales sur les axes 1 et
3")

#Projection des individus sur le plan (1;2)

fviz_pca_ind(res.pca2,

```



```

axes=c(1,2),

col.ind="cos2",

gradient.cols=c("black", "blue", "red"),

repel=TRUE,

title="Projection des individus sur les axes 1 et 2

30 observations",select.ind = list(cos2 = 30))

```

#Projection des individus sur le plan (1;3)

```

fviz_pca_ind(res.pca2,

axes=c(1,3),

col.ind="cos2",

gradient.cols=c("black", "blue", "red"),

repel=TRUE,

title="Projection des individus sur les axes 1 et 3

30 observations",select.ind = list(cos2 = 30))

```

Problème Graphique R - Pour inverser l'axe 2 - Graphique avec la variable illustrative

```
res.pca1$var$coord[, 2] <- -res.pca1$var$coord[, 2]
```

ACP axes 1, 2 - Avec la variable illustrative

```

fviz_pca_var(res.pca1, axes=c(1,2),

col.var = "cos2",

gradient.cols = c("#1111fa", "#fae311", "#FF0000"),

repel = TRUE)+ labs(title = "Analyse en composantes principales sur les axes 1 et

2

avec la variable illustrative")

```

```
# ACP axes 1, 3 - Avec la variable illustrative
```

```
fviz_pca_var(res.pca1, axes=c(1,3),  
             col.var = "cos2",  
             gradient.cols = c("#1111fa", "#fae311", "#FF0000"),  
             repel = TRUE)+ labs(title = "Analyse en composantes principales sur les axes 1 et  
3")
```

```
# IV – Corrélation entre la variable illustrative et les variables latentes
```

```
# Définition des nom des variables latentes
```

```
Accès_services_consommation_loisir <- res.pca2$ind$coor[,1]
```

```
Pratiques_mobilité_collective <- res.pca2$ind$coor[,2]
```

```
Environnement_travail <- res.pca2$ind$coor[,3]
```

```
# RLM
```

```
modele1 <-
```

```
lm(base$Nuitées~Accès_services_consommation_loisir+Pratiques_mobilité_collective)
```

```
summary(modele1)
```

```
modele2 <-
```

```
lm(base$Nuitées~Accès_services_consommation_loisir+Environnement_travail)
```

```
summary(modele2)
```

```
# Résidus des RLM
```

```
base$residu1=round(residuals(modele1),2)
```

```
base$residu2=round(residuals(modele2),2)
```

```
#Boxplots résidus des RLM
```

```
par(mfrow=c(1,2))
```

```
boxplot(base$residu1,main = "Résidus du modèle 1",col="orange")
```

```
boxplot(base$residu2,main = "Résidus du modèle 2",col="orange")
```

#Histogrammes résidus des RLM

```
hist(base$residu1,ylab="Nombre de départements",xlab = "Résidus du modèle 1",  
freq=TRUE,col="orange",cex.main=0.6,breaks=10)
```

```
hist(base$residu2,ylab="Nombre de départements",xlab = "Résidus du modèle 2",  
freq=TRUE,col="orange",cex.main=0.6,breaks=10)
```

Skewness et Kurtosis de résidus

```
skewness(base$residu1)
```

```
kurtosis(base$residu1)
```

```
skewness(base$residu2)
```

```
kurtosis(base$residu2)
```

Diagrammes QQ des résidus

```
qqnorm(base$residu1)
```

```
qqline(base$residu1,col="orange")
```

```
qqnorm(base$residu2)
```

```
qqline(base$residu2,col="orange")
```

Tests de Kolmogorov-Smirnov des résidus

```
ks.test(base$residu1,"pnorm",mean(base$residu1),sd(base$residu1))
```

```
ks.test(base$residu2,"pnorm",mean(base$residu2),sd(base$residu2))
```

#Création des colonnes de valeur prédite par les modèles

```
base$predict1 = round(predict(modele1), 2)
```

```
base$predict2 = round(predict(modele2), 2)
```

Graphiques entre la variable illustrative observée et prédite par les RLM

```

par(mfrow=c(1,1))

plot(base$predict1, base$Nuitées, xlab="Nuitées prédites", ylab="Nuitées observées",
col=0,main = "Modèle 1", font.lab=3,

      abline(a=0, b=1, col="orange", lty=1))

text(base$predict1, base$Nuitées, base$Code, col="black", cex=.6)

plot(base$predict2, base$Nuitées, xlab="Nuitées prédites", ylab="Nuitées observées",
col=0,main = "Modèle 2", font.lab=3,

      abline(a=0, b=1,col="orange", lty=1))

text(base$predict2, base$Nuitées, base$Code, col="black", cex=.6)

```

#Résumé de la variable illustrative

```
summary(base$Nuitées)
```

Création d'une variable binaire en numérique à partir de la variable illustrative pour différents seuils : 877.2 , 2 281,5 et 879

```

base$Nuitéesbis[base$Nuitées>877.2 ]="1"

base$Nuitéesbis[base$Nuitées<=877.2 ]="0"

baseNuitéesbisN = as.numeric(base$Nuitéesbis)

```

Régressions logistiques binaires

```

modele1bis=glm(baseNuitéesbisN ~
Accès_services_consommation_loisir+Pratiques_mobilité_collective, family=binomial)

summary(modele1bis)

modele2bis=glm(baseNuitéesbisN ~
Accès_services_consommation_loisir+Environnement_travail, family=binomial)

summary(modele2bis)

```

VII. Bibliographie

- Le surtourisme : quel impact sur les villes et sur l'environnement ? : <https://www.vie-publique.fr/eclairage/24088-le-surtourisme-quel-impact-sur-les-villes-et-sur-lenvironnement> (consulté le 22/10/2023)
- La France, toujours première destination touristique au monde : <https://www.gouvernement.fr/actualite/la-france-toujours-premiere-destination-touristique-au-monde> (consulté le 26/09/2023)
- L'essentiel sur... le tourisme : <https://www.insee.fr/fr/statistiques/7653005#graphique-figure4> (consulté le 20/09/2023)
- Fréquentation des hébergements collectifs touristiques en 2022 : https://www.insee.fr/fr/statistiques/2012672#tableau-TCRD_020_tab1_regions2016/ (consulté le 20/09/2023)
- Nombre d'équipements et de services dans le domaine du commerce en 2021: <https://www.insee.fr/fr/statistiques/3568602?sommaire=3568656#dictionnaire> (consulté le 01/10/2023)
- Commerce de détail, loisirs et tourisme : vers un renforcement des liens ? : https://www.persee.fr/docAsPDF/htn_0018-439x_2001_num_2_1_2766.pdf (consulté le 15/10/2023)
- Du tourisme médical à une géographie des interactions entre tourisme et santé : <https://socgeo.com/2017/05/19/du-tourisme-medical-a-une-geographie-des-interactions-entre-tourisme-et-sante/> (consulté le 04/10/2023)
- Professionnels de santé au 1^{er} janvier 2023 : <https://www.insee.fr/fr/statistiques/2012677/> (consulté le 01/10/2023)
- Nombre d'équipements et de services dans le domaine du sport, des loisirs et de la culture en 2021 : <https://www.insee.fr/fr/statistiques/3568599?sommaire=3568656> (consulté le 27 /09 /2023)

- La culture, un véritable atout pour le succès touristique de nos territoires :
<https://www.culture.gouv.fr/Thematiques/Developpement-culturel/Actualites/La-culture-un-veritable-atout-pour-le-succes-touristique-de-nos-territoires/> (consulté le 12 /10 /2023)
- Culture, tourisme et territoire : les apports du tourisme culturel au développement local :
https://memoires.sciencespo-lyon.fr/internet/Masters/SECI/juanchich_l/pdf/juanchich_l.pdf
(consulté le 17 /10 /2023)
- Niveau de vie et pauvreté des territoires en 2020 :
<https://www.insee.fr/fr/statistiques/6692414?sommaire=6692394#documentation> (consulté le 27 /09 /2023)
- Température moyenne mensuelle par département :
<https://www.data.gouv.fr/fr/reuses/temperature-moyenne-mensuelle-par-departement/>
(consulté le 01/10/2023)
- Tourisme et changement climatique, une relation à double sens : <https://www.tec-conseil.com/lab/publications/colloques-conferences-et-ateliers/tourisme-et-changement-climatique-une-relation-a-double-sens.htm?lng=fr> (consulté le 22/10/2023)
- Nombre d'équipements et de services dans le domaine du sport, des loisirs et de la culture en 2021 : <https://www.insee.fr/fr/statistiques/3568599?sommaire=3568656#dictionnaire>
(consulté le 27 /09 /2023)
- Le tourisme sportif, un marché en plein essor : <https://www.sport-et-tourisme.fr/corporate/essor-du-tourisme-sportif/> (consulté le 21 /10 /2023)
- Le tourisme sportif en quête d'identité. La construction identitaire dans les organisations de tourisme sportif, entre idéologies sportives et matérialité professionnelle marchande. Bernard Massiera : <https://theses.hal.science/tel-00482139/document> (consulté le 27/09/2023)
- Population active et taux d'activité au sens du recensement selon le sexe et l'âge en 2020 :
https://www.insee.fr/fr/statistiques/2012710#tableau-TCRD_015_tab1_departements (consulté le 27/09/2023)
- Population active : <https://data.oecd.org/fr/emp/population-active.htm> (consulté le 26/10/2023)
- Pénurie de main-d'œuvre dans le secteur du tourisme :
<https://www.senat.fr/questions/base/2022/qSEQ220628097.html> (consulté le 26/10/2023)

- Pénurie de main-d'œuvre : le couac des vacances? : https://www.lepoint.fr/economie/les-vacances-menacees-par-la-penurie-de-main-d-oeuvre-24-05-2022-2476807_28.php#11 (consulté le 26/10/2023)
- Réseau routier au 31 décembre 2021 : <https://www.insee.fr/fr/statistiques/2012705> (consulté le 28/09/2023)
- Réseaux de transport et mobilités touristiques : https://www.researchgate.net/publication/325256026_Reseaux_de_transport_et_mobilites_touristiques (consulté le 14/10/2023)
- Infrastructures de transport en France : vulnérabilité au changement climatique et possibilités d'adaptation : https://www.caissedesdepots.fr/fileadmin/PDF/finance_carbone/etudes_climat/etude_climat_n18-e.pdf (consulté le 14/10/2023)
- Réseaux de transport et mobilités touristiques. Pratiques, modèles et nouveaux défis : https://www.researchgate.net/publication/325256026_Reseaux_de_transport_et_mobilites_touristiques (consulté le 14/10/2023)
- La tendance de « la sobriété », qui vise à réduire une activité spécifique, dans ce cas, la longueur d'un trajet : https://presse.ademe.fr/2021/06/le-tourisme-durable-en-france-un-levier-de-relance-ecologique.html#_ftnref1 (consulté le 15/10/2023)
- Infrastructure autoroutière et développement du tourisme : https://temis.documentation.developpement-durable.gouv.fr/docs/Temis/0060/Temis-0060963/7664_Rapport.pdf (consulté le 15/10/2023)
- Aménagements cyclables : <https://www.amenagements-cyclables.fr/fr/stats> (consulté le 01/10/2023)
- Hecker Anne. Véloroutes et voies vertes : supports, ou objets touristiques ?. In: *Collection EDYTEM. Cahiers de géographie*, numéro 4, 2006. Transport et tourisme. pp. 199-208 : <https://doi.org/10.3406/edyte.2006.982> www.persee.fr/doc/edyte_1762-4304_2006_num_4_1_982 (consulté le 12/10/2023)
- Estimation de la population au 1^{er} janvier 2023 : <https://www.insee.fr/fr/statistiques/1893198> (consulté le 27/09/2023)

- Superficie : <https://www.observatoire-des-territoires.gouv.fr/superficie> (consulté le 27 /09 /2023)
- Aménagement du territoire : plus que jamais une nécessité : https://www.senat.fr/rap/r16-565/r16-565_mono.html (consulté le 30 /10 /2023)
- I. LA MOBILITÉ ET LES ZONES PEU DENSES : UN PÉRIMÈTRE FLOU :
https://www.senat.fr/rap/r20-313/r20-313_mono.html
- Le tourisme : acteurs, lieux et enjeux » M. Stock. Belin, Paris 2003, p10 (consulté le 08/11 /2023)

VIII. Table des matières

Sommaire.....	2
I. Introduction	3
II. Analyse détaillée du sujet.....	4
A. Pertinence de la variable à expliquer	4
B. Rôle des variables explicatives et corrélation entre la variable illustrative et les variables actives	6
1. Les commerces	6
2. Les médecins généralistes.....	8
3. Les équipements culturels	9
4. Le niveau de vie médian.....	10
5. La température	12
6. Les équipements sportifs	13
7. La population active.....	14
8. Le réseau routier.....	16
9. Les aménagements cyclables	18
10. La densité de population	19
C. Statistiques et valeurs aberrantes des variables	21
III. Analyse en Composantes Principales ACP.....	22
A. Valeurs propres et nombre d'axes	22
B. Variables : coordonnées, contributions, cosinus carrés.....	23
1. Les coordonnées	23
2. Les contributions	25

3.	Les cosinus carrés.....	26
C.	Cercle de corrélation.....	27
D.	Définition des variables latentes.....	29
E.	Projection des individus sur le plan factoriel.....	30
F.	Projection de la variable illustrative sur le cercle de corrélation.....	34
G.	ACP : le cas sans Paris.....	35
IV.	Corrélation entre la variable illustrative et les variables latentes	38
A.	Régression linéaire multiple.....	38
B.	Relation entre la variable illustrative observée et prédite par RLM	42
C.	Régression logistique binaire	44
1.	Le premier quartile.....	44
2.	La médiane	45
3.	Le seuil optimal	45
V.	Conclusion et discussion des résultats	47
A.	Conclusion	47
B.	Discussion	48
VI.	Annexes	50
A.	Tableaux et graphiques supplémentaires.....	50
B.	Base de données	53
C.	Script R.....	57
VII.	Bibliographie.....	69