

PROJET

ETUDE DE POLLUTION AÉRIENNE AUX ABORDS
DES ÉCOLES ET CRÈCHES D'ÎLE-DE-FRANCE

Équipe: CHIRINA Rania, MAKHLOUFI Khalil, VITOFFODJI Adjimon.



Département MIASHS, UFR 6 Informatique, Mathématique et Statistique
Université Paul Valéry, Montpellier 3

Novembre 2023

SOUMIS COMME CONTRIBUTION PARTIELLE
POUR LE COURS DE SCIENCE DES DONNÉES

Déclaration de non plagiat

Nous déclarons que ce rapport est le fruit de notre seul travail, à part lorsque cela est indiqué explicitement.

Nous acceptons que la personne évaluant ce rapport puisse, pour les besoins de cette évaluation:

- la reproduire et en fournir une copie à un autre membre de l'université; et/ou,
- en communiquer une copie à un service en ligne de détection de plagiat (qui pourra en retenir une copie pour les besoins d'évaluation future).

Nous certifions que nous avons lu et compris les règles ci-dessus.

En signant cette déclaration, nous acceptons ce qui précède.

Signature: _____ Date: _____

Signature: _____ Date: _____

Signature: _____ Date: _____

Remerciements

Nos plus sincères remerciements vont à notre encadrant pédagogique pour les conseils avisés sur notre travail.

28/11/2023.

Résumé

L'exploration visuelle de notre jeu de données sur la pollution atmosphérique aux abords des crèches et écoles en Île-de-France a permis de dégager des tendances significatives. Le carte géographique a révélé des zones de concentrations élevées, plus on se rapproche de Paris plus la pollution de l'air est élevée, le graphique à barre a illustré les concentration de l'un des polluants majeures qui impacte la santé des jeunes (dioxyde d'azote), la droite de régression montre la relation entre ce dernier polluant en 2012 et en 2017, les boîtes à moustaches ont fourni des indications sur la dispersion des trois polluants au fil des années, le graphique en radar expose les niveaux moyens des trois polluants atmosphériques qui semblent avoir stagné sur toute la période de 2012 à 2017, et enfin le graphe qui montre que la concentration en pourcentage de NO₂ est largement supérieure par rapport aux autres polluants, cela s'explique par le dégagement du NO₂ produit par les véhicules.

Table des matières

Chapitre 1	Contexte, définition et intérêt du projet	2
1.1	Enoncé du problème	2
1.2	Interêt de l'étude	2
1.3	Composition de l'équipe	3
1.4	Objectif, l'imites de l'étude et Logiciel	3
1.4.1	Objectif de l'étude	3
1.4.2	Limtes de l'étude	3
1.4.3	Logicieles	4
Chapitre 2	Science des données	5
2.1	Nature, Source des données et description des tables	5
2.1.1	Nature et source	5
2.1.2	Description des tables	5
2.1.3	Vision générale des données	6
2.2	PRESENTATION ET INTERPRETATION DES RESULTATS . .	6
2.2.1	Résumé figure 1	6
2.2.2	Résumé figure 2	7
2.2.3	Résumé figure 3	8
2.2.4	Résumé figure 4	9
2.2.5	Résumé figure 5	10
2.2.6	Résumé figure 6	11
2.3	Difficultés rencontrés	12

Introduction

Le changement climatique, phénomène mondial aux répercussions multidimensionnelles, s'impose aujourd'hui comme l'un des défis les plus pressants de notre époque. La planète subit des altérations substantielles de son climat, attribuables en grande partie aux émissions de gaz à effet de serre résultant des activités humaines (GIEC, 2021). En Île-de-France, région métropolitaine au cur de cette dynamique, les conséquences du changement climatique sont de plus en plus tangibles, mettant en lumière la nécessité impérieuse d'agir pour atténuer ces impacts et s'adapter aux transformations inéluctables.

Au-delà des perturbations climatiques elles-mêmes, le changement climatique agit comme un catalyseur, exacerbant d'autres problématiques environnementales, dont l'une des plus préoccupantes est la pollution atmosphérique. Les émissions de polluants, liées aux activités industrielles, au trafic routier, et à d'autres sources anthropiques, contribuent de manière significative à la dégradation de la qualité de l'air (ADEME, 2021). Cet enjeu revêt une importance particulière aux abords des écoles et crèches, où la vulnérabilité des enfants face aux effets nocifs de la pollution atmosphérique est avérée (OMS, 2018 ; American Academy of Pediatrics, 2018).

Ainsi, dans le contexte francilien, La qualité de l'air est cruciale pour la santé, en particulier celle des enfants dont les systèmes respiratoires en pleine croissance sont excessivement sensibles aux agents polluants tels que le dioxyde d'azote (NO₂), les particules fines (PM₁₀) et ultrafines (PM_{2,5}) présents dans l'air. Une exposition prolongée à ces polluants, souvent présents à des niveaux élevés dans les environnements urbains et spécifiquement dans les établissements scolaires, peut avoir un impact dévastateur sur la santé et le bien-être des élèves, des enseignants et du personnel éducatif. Pour ce faire, nous disposons d'un jeux de données prenant source du site internet data.gouv. Nous nous demanderons dans ce contexte:

Quelles sont les concentrations en microgramme par mètre cube de moyenne journalière annuelle pour trois polluants aériens (NO₂, PM₁₀ et PM_{2,5}) relevées aux abords des crèches et établissements scolaires d'Île-de-France pour les années de 2012 à 2017?

CHAPITRE 1

Contexte, définition et intérêt du projet

1.1 Enoncé du problème

L'Île-de-France, située au nord de la France et abritant Paris, la métropole la plus peuplée du pays, se trouve au centre d'un enjeu de santé publique majeur : la qualité de l'air. Cette région dynamique, foisonnante d'activités culturelles et économiques, est confrontée à un défi crucial. En son sein, les écoles, véritables foyers de savoir et d'épanouissement pour les générations à venir, font face quotidiennement aux effets néfastes de la pollution atmosphérique. Cette forme de pollution est liée spécifiquement au trafic routier représente une menace sur la santé de nos enfants qui normalement sont censés étudier dans des environnements d'apprentissage sains. Si l'on s'accorde sur le fait que la pollution atmosphérique aux abords des écoles et crèches est très néfaste pour la bonne santé de nos enfants, il est alors digne d'intérêt de se poser des questions sur la concentration microgramme par mètre cube de moyenne journalière annuelle pour trois polluants aériens (NO₂, PM₁₀ et PM_{2,5}) aux abords des crèches et établissements scolaires d'Île-de-France pour les années de 2012 à 2017. Cela suscite en nous les interrogations suivantes : comment évaluer et atténuer de manière efficace la pollution atmosphérique aux abords des écoles et crèches d'Île-de-France ? Quelles sont les zones à forte concentration de pollution ? Quelle est la pollution moyenne aux abords des crèches et établissement scolaire d'Île-de-France pour les années 2012 à 2017?

Autant de questionnements qui suscitent l'intérêt de notre thème "ETUDE DE POLLUTION AÉRIENNE AUX ABORDS DES ÉCOLES ET CRÈCHES D'ÎLE-DE-FRANCE"

1.2 Intérêt de l'étude

L'importance de l'étude sur la pollution atmosphérique n'est plus à démontrer, tant sur le plan environnemental que sur celui de la santé publique. Tout d'abord, la proximité des établissements scolaires avec des sources potentielles de pollution, telles que les routes à fort trafic, les zones industrielles ou les centres urbains, exposent les enfants à des niveaux préoccupants de contaminants atmosphériques. En raison de leur physiologie en développement et de leur taux de ventilation plus élevé que celui des adultes, les enfants sont particulièrement vulnérables aux effets néfastes de la pollution atmosphérique (American Academy of Pediatrics, 2018).

Ainsi, cette étude vise à évaluer ces risques sanitaires spécifiques pour les enfants et à concevoir des mesures de protection appropriées. Par ailleurs, l'analyse de la qualité de l'air aux abords des écoles et des crèches contribue à sensibiliser la communauté éducative, les autorités locales et les parents aux enjeux environnementaux et sanitaires auxquels les enfants sont régulièrement confrontés. Cette prise de conscience collective peut servir de levier pour promouvoir des initiatives visant à réduire les émissions de polluants, à améliorer l'urbanisme et à promouvoir des modes de transport plus durables dans les zones sensibles. À un niveau plus global, cette étude s'inscrit dans une démarche visant à intégrer les objectifs de développement durable, notamment liés à la santé (OMS, 2018), à la qualité de l'air et à la lutte contre le changement climatique. Cette campagne vise notamment à promouvoir deux mesures immédiates pour protéger les enfants dans les communes de la métropole du Grand Paris : la création de zone piétonnes autour des écoles ; la mise en place de zones à faibles émissions (ZFE).

1.3 Composition de l'équipe

- **CHIRINA Rania :**
- **MAKHOULI KHALIL :**
- **VITOFFODJI Adjimon Jérôme :**

1.4 Objectif, l'impact de l'étude et Logiciel

1.4.1 Objectif de l'étude

L'objectif général de cette étude est de fournir des figures informatives et des données visuelles concernant la pollution aérienne aux abords des écoles et crèches d'Ile-de-France pour les années 2012 à 2017. De façon spécifique il s'agit de mettre particulièrement l'accent sur la présence de dioxyde d'azote (NO₂), de particules fines PM₁₀ et PM_{2,5} en cartographiant ces polluants et en identifiant les zones à risque. Les figures générées vont permettre de sensibiliser le grand public à l'importance de la qualité de l'air près des établissements scolaires.

1.4.2 Limites de l'étude

- Notre étude s'est basée dans un second temps uniquement sur une analyse descriptive à travers l'interprétation des graphes. Ainsi donc, nous nous sommes pas allés en profondeur de l'étude, en faisant une analyse approfondie à travers la modélisation de nos modèles afin de vérifier nos hypothèses.
- Nous nous sommes pas assurés de la distribution des variables de cette étude, ni de vérifier à travers les tests d'hypothèses de la validation de nos modèles.

1.4.3 Logiciels

Dans la réalisation de notre projet, nous avons utilisé le logiciel excel(2016) pour constituer notre base. Ensuite nous avons importé cette base au format csv sous google colab pour la réalisation de nos différents graphs. Nous avons ensuite utilisé les library Numpy, Pandas, Seaborn, Geopandas, matplotlib.pyplot

Pour la rédaction de notre projet, nous avons utiliser la version 28-1-2 du logiciel Rstudio sous Mac pro à l'aide de R Markdown pour générer directement notre document en format pdf. L'analyse descriptive a été faite aussi avec ce dernier.

CHAPITRE 2

Science des données

2.1 Nature, Source des données et description des tables

2.1.1 *Nature et source*

Les données utilisées dans la présente étude proviennent d'Internet en licence libre et exploitable. En effet, ne nous sommes pas par les auteurs de toutes ses données, ainsi donc pour permettre aux autres utilisateurs à consulter ou utiliser ses données dans leurs différents travaux, nous mettrons dans la bibliographie avec plus de précision les différentes sources de ses données.

Ces données proviennent de :

[https://www.data.gouv.fr/fr/datasets/
base-de-donnees-de-la-pollution-aerienne-aux-abords-des-ecoles-et-creches-dil](https://www.data.gouv.fr/fr/datasets/base-de-donnees-de-la-pollution-aerienne-aux-abords-des-ecoles-et-creches-dil)
Cette dernière est constituée de 25 colonnes et 12 196 lignes.

2.1.2 *Description des tables*

Les principales variables de notre jeu de données sont :

- ID : l'identifiant unique de la crèche ou de l'école (**ID**);
- nom : le nom usuel de l'établissement dans les fichiers de référence (**nom**);
- Code postal : le code postal de l'établissement (**CP**);
- ville : la ville suivant la nomenclature de l'Insee (à deux exceptions : Crèvecœur-en-Brie et Vincy-Manuvre, non référencées [dans la dernière base disponible] (**ville**);
- type : la catégorie d'établissement (crèche, primaire ou secondaire). (**type**);
- Les variables correspondant aux niveaux d'exposition suivent la nomenclature suivante : "polluant_AAAA", AAAA figurant l'année civile sur 4 chiffres. Leurs valeurs numériques sont exprimées en microgrammes par mètre cube en moyenne journalière annuelle.

Exemple : la variable "PM25_2014" contient les concentrations moyennes par jour de PM2,5 pour l'année 2014.

- geometry : La variable "geometry" correspond aux coordonnées des adresses géolocalisées des établissements en système EPSG:4326 (WGS-84) (**geometry**)

2.1.3 Vision générale des données

Identifiants	Variables quantitative	Variables qualitative
1- Unique	Discretes : 21	Norminal: 3

Table: Description des types d'attribut.

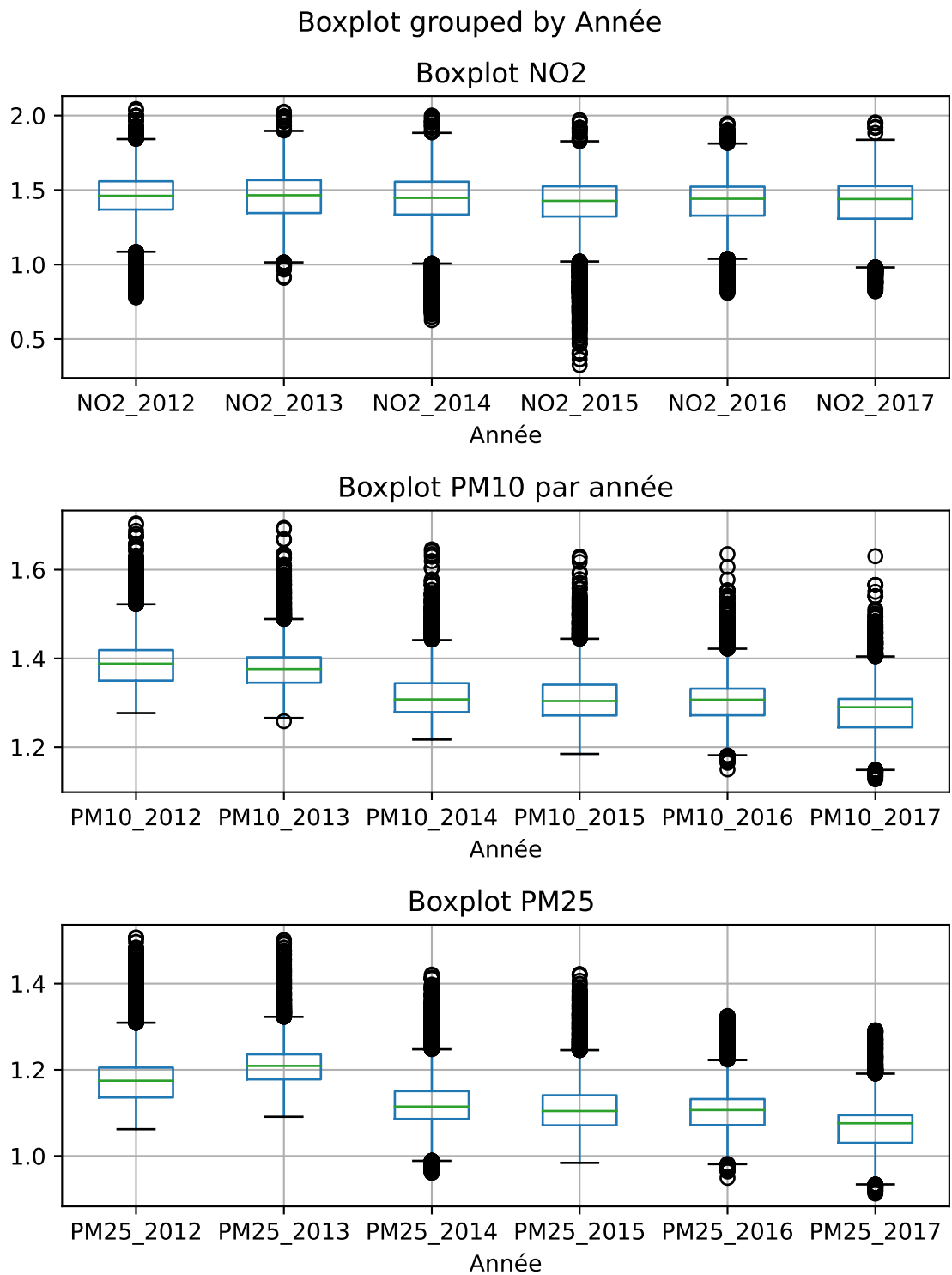
2.2 PRESENTATION ET INTERPRETATION DES RESULTATS

2.2.1 Résumé figure 1

Sur la figure 1, la concentration médiane du polluant diazote (NO₂) au cours de l'année 2016 est d'environ 1.48%. Cette valeur est encadré par le premier quartile et le troisième quartile respectivement 1.30% et 1.53%. cette médiane est presque la même pour les années 2013 à 2017. Plusieurs valeurs atypiques se trouvent au dessus de la frontière haute et basse de chaque Boxplot.

De la sous figure 2, la concentration médiane du polluant PM₁₀ observé en 2012 est de 1.38%. Cette valeur est encadré par le premier quartile et le troisième quartile respectivement 1.35% et 1.42%. Par ailleurs on note la présence de plusieurs valeurs atypiques au dessus de la frontière haute qui est 1.53%. Notons également une diminution au fil des années de la valeur médiane du PM₁₀. Elle est passée de 25 en 2012 à 19 en 2017, avec une valeur atypique en dessous de la frontière basse en 2013 et plusieurs valeurs atypiques en 2016 et 2017.

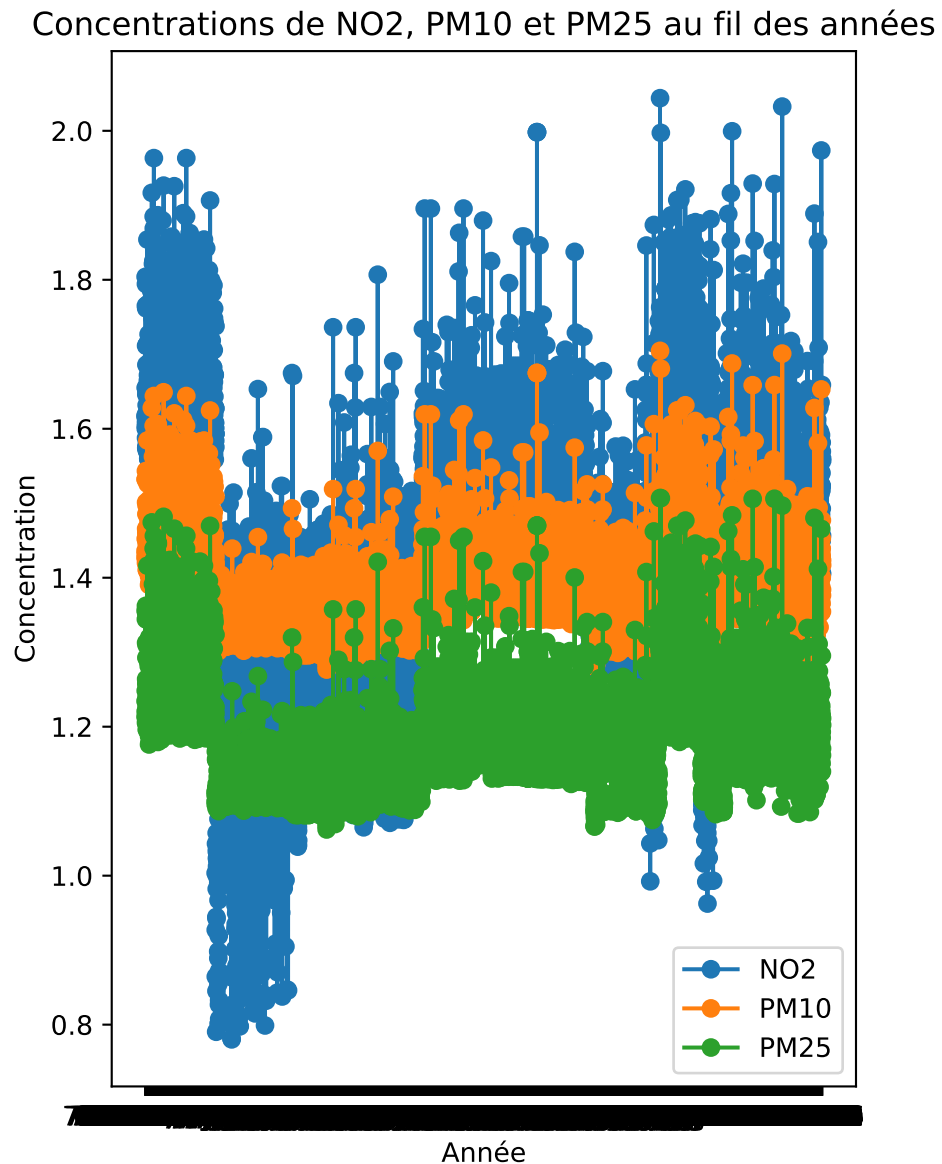
De l'analyse sous figure 3, on note que la concentration médiane de PM₂₅ a atteint son pic en 2013. Cette valeur est de 1.22%. Elle est encadré par le premier quartiles et le troisième quartiles qui sont respectivement 1.18% et 1.25%. Note plusieurs valeurs atypiques de PM₂₅ au dessus de la frontière haute pour toutes les années et des valeurs atypiques en dessous de la frontière basse en 2014 2016 et 2017



2.2.2 Résumé figure 2

Ce type de graphique nous permet de visualiser la tendance temporelle des données et comparer les variations de différentes séries. En l'occurrence, il montre comment les concentrations de NO2, PM10 et PM25 varient d'une année à l'autre. Chaque ligne représente une série temporelle pour une des trois polluants (vaeiable), avec les années sur l'axe des abscisses (X) et les concentrations sur

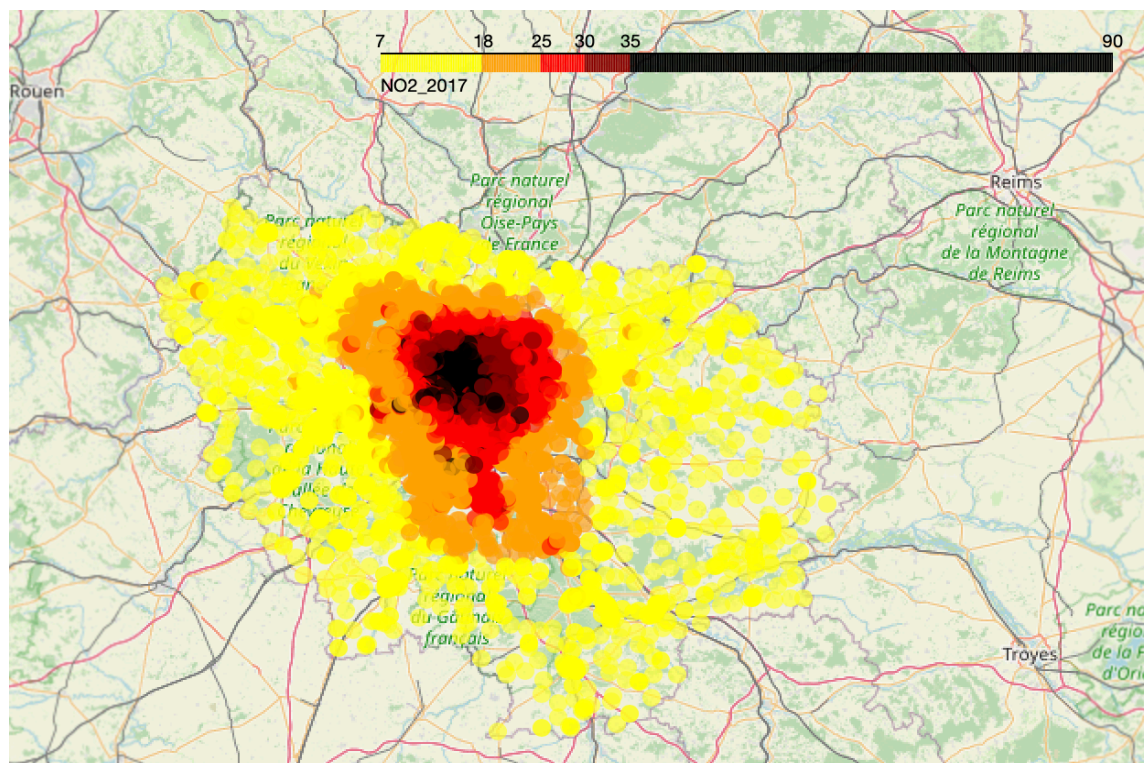
l'axe des ordonnées (Y). Les points marqués avec 'o' indiquent les valeurs spécifiques pour chaque année. La concentration moyenne en pourcentage du diazote (NO₂) aux abords des crèches et établissements scolaires est largement supérieur à celle de PM₁₀ et PM₂₅. Cela s'explique par le dégagement du diazote que dégagent les véhicules à moteurs.



2.2.3 Résumé figure 3

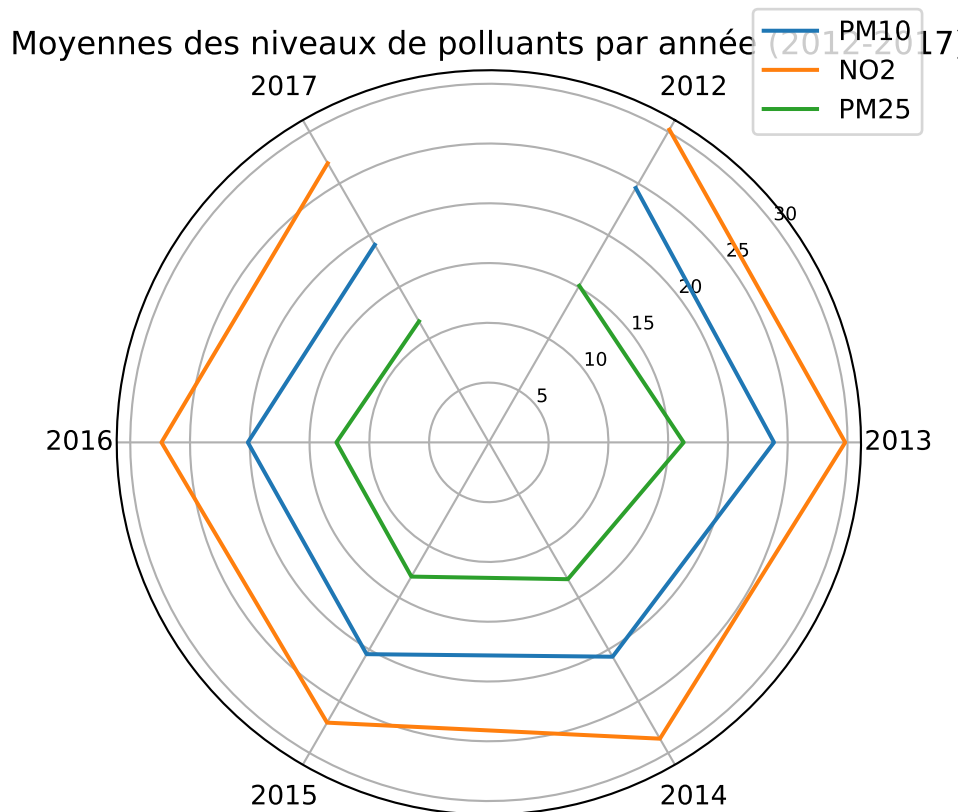
Sur cette carte nous pouvons apercevoir plusieurs cercles représentant chacun un établissement scolaires. Cette analyse cartographique nous montre la présence d'une corrélation. En effet, à mesure que l'on se rapproche de la métropole française (Paris), les niveaux de dioxyde d'azote (NO₂) augmentent

considérablement, comme indiqué par des cercles de couleurs plus foncées. Cette corrélation entre la proximité des écoles et les niveaux élevés de pollution expose donc les élèves à des environnements potentiellement plus pollués. Le NO₂ est un polluant qui est généré notamment à cause du trafic routier. Paris possède une densité de trafic routier élevée en raison de sa densité urbaine importante et de sa population dense. Ces facteurs contribuent à un grand nombre de véhicules circulant dans la ville, générant ainsi des niveaux significatifs de pollution atmosphérique, notamment de dioxyde d'azote (NO₂). Ainsi, cette analyse suggère une relation entre l'emplacement des écoles et les niveaux de pollution atmosphérique, soulignant l'importance de prendre en compte ces aspects pour maintenir la santé et le bien-être des élèves.



2.2.4 Résumé figure 4

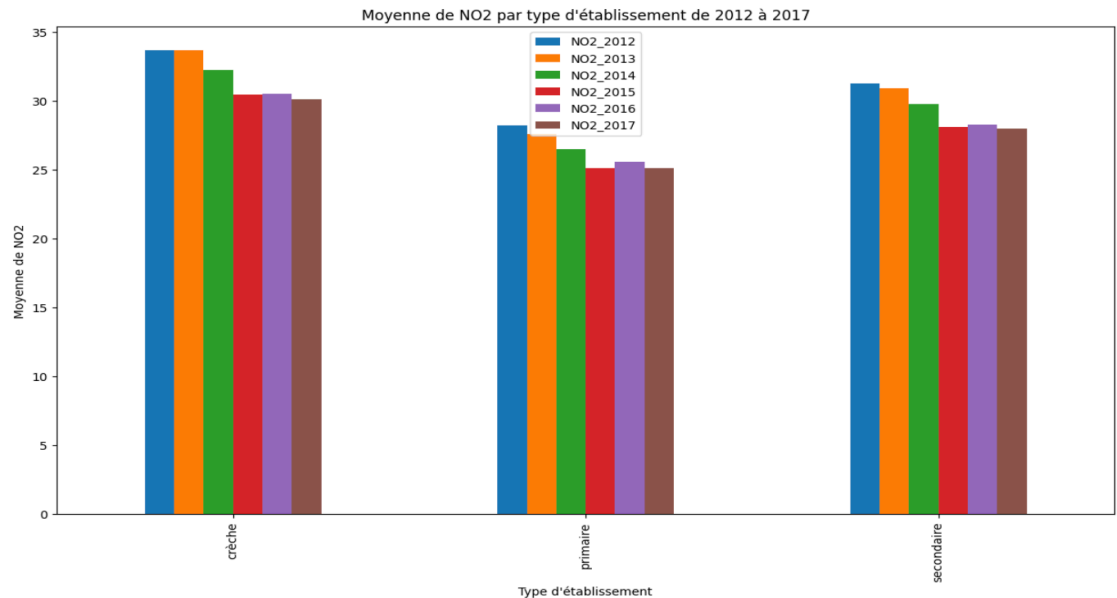
Ce graphique en radar nous expose les niveaux moyens de polluants atmosphériques (NO₂, PM₁₀, PM₂₅). Les trois polluants semblent avoir stagné sur toute la période de 2012 à 2017. Cette absence de variation suggère une relative stabilité dans les niveaux de pollution de l'air dans la région d'Île-de-France au cours de ces années. Cela pourrait indiquer une absence de changement significatif dans les émissions de ces polluants et une constance dans l'exposition potentielle des enfants à ces polluants. Ce graphe souligne l'importance de surveiller de près ces niveaux pour assurer des environnements sains et sûrs pour les jeunes enfants fréquentant les établissements scolaires exposés aux différents polluants.



2.2.5 Résumé figure 5

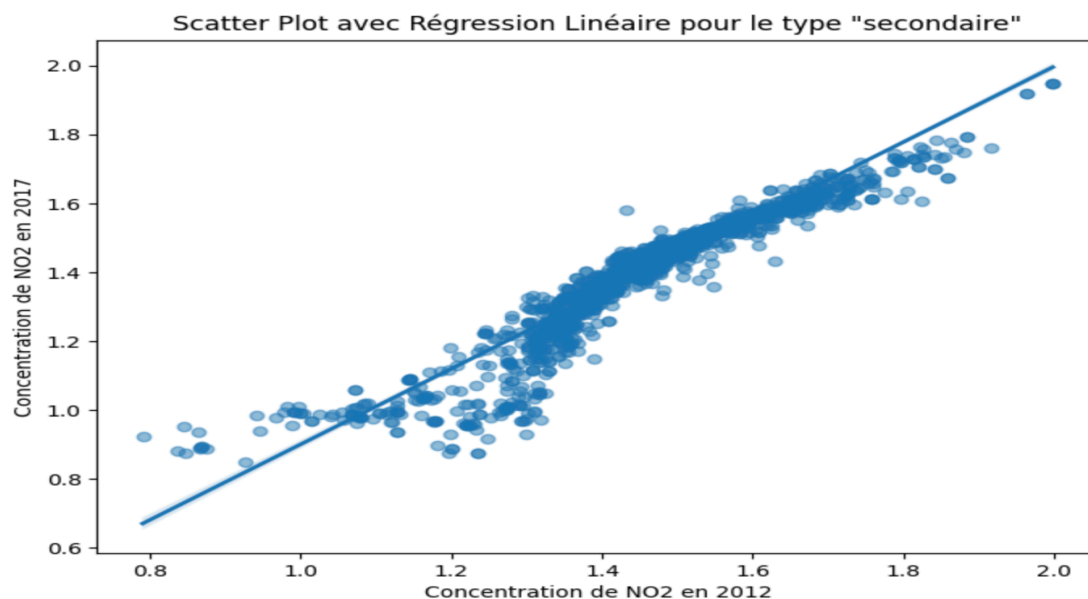
Cette figure représente la moyenne de NO₂ par type d'établissement de l'année 2012 à 2017, elle a sur l'axe y la moyenne de NO₂ et sur l'axe des x le type d'établissement (maternel, primaire, secondaire). Ce diagramme à barres généré montre les moyennes de concentration de NO₂ au fil des années de 2012 à 2017.

pour chaque type d'établissement. On peut observer que la moyenne de NO₂ qui est représentée par une barre varie d'une année à l'autre et d'un établissement à l'autre, on peut voir que la moyenne de la concentration de NO₂ diminue au fil des années pour chaque type d'établissement, et on d'après le graphe on peut constater que la moyenne de concentration de NO₂ est élevée dans le type d'établissement secondaire à par rapport aux autres (crèche et primaire).



2.2.6 Résumé figure 6

Ce scatter plot avec la régression linéaire permet de visualiser la relation entre les concentrations de NO₂ en 2012 et en 2017 pour les établissements du type secondaire, chaque point représente un établissement et la ligne de régression indique la tendance générale. Dans notre nuage de points, les points suivent la droite de régression cela veut dire que les établissements ayant des concentrations plus élevées en 2012 ont tendance à avoir des concentrations plus élevées en 2017 (corrélation positive).



2.3 Difficultés rencontrés

La réalisation du projet a impliqué plusieurs défis. Initialement, la recherche d'une base de données appropriée a demandé du temps. Générer des idées pour créer des figures informatives a été une tâche très réflexive. Connecter de manière fluide toutes les figures pour assurer la cohérence du rapport a été la tâche la plus importante. Sélectionner les variables pertinentes dans la base de données a nécessité une analyse approfondie. Enfin, la manipulation des chunks dans R Markdown pour une présentation adéquate du code et des résultats a été complexe.

CONCLUSION

En conclusion, nos diverses représentations visuelles ont considérablement approfondi notre compréhension de la pollution atmosphérique aux abords des écoles et crèches en Île-de-France, couvrant la période de 2012 à 2017. Ces visualisations se révèlent être des instruments cruciaux pour sensibiliser, analyser et concevoir des solutions ciblées afin de réduire l'impact de la pollution sur la santé des élèves, en répondant ainsi à notre problématique initiale. Néanmoins, cette étude a mis en évidence une augmentation notable de la propagation du dioxyde d'azote en Île-de-France. Cette hausse s'explique notamment par la forte utilisation de carburants liée à la présence de la métropole française, Paris, dans cette région. Ainsi, il est vivement recommandé de privilégier l'utilisation des transports en commun et de considérer l'adoption précoce de moteurs électriques pour préserver la santé de ces élèves.