

Cahier des Charges : Étude de l'empreinte CO2 de l'électricité

SAÉ 2-06 analyse de données, reporting et datavisualisation



⚡ Electricity Maps

Remerciements :

Nous tenons à exprimer notre sincère reconnaissance à l'ensemble des personnes qui ont contribué de près ou de loin à la réalisation de ce projet.

En premier lieu, nous remercions chaleureusement nos tuteurs — Francine EMERY, Agnès FRONT, Sophie LAMBERT-LACROIX et Manon VALLA — pour leur accompagnement bienveillant, leur disponibilité constante, et la qualité de leurs conseils tout au long de cette SAÉ. Leur regard expert et leurs retours constructifs nous ont permis de structurer notre travail, d'ajuster notre méthodologie et d'approfondir nos analyses avec exigence et clarté.

Nous adressons également notre gratitude à l'ensemble de l'équipe pédagogique du département Sciences des Données, dont les enseignements en statistique, gestion de projet, datavisualisation, reporting, ou encore outils de traitement de données ont constitué des fondations solides pour mener à bien ce projet. Leur implication dans notre formation a été un réel atout pour aborder ce travail de manière professionnelle et autonome.

Aussi, nous remercions pleinement nos commanditaires, Claudia Roncancio et Mario Cortes-Cornax, du Laboratoire d'informatique de Grenoble.

Pour leur confiance et pour nous avoir offert l'opportunité de contribuer à une problématique aussi actuelle que cruciale. Leur implication et leur intérêt pour notre travail ont été très motivants, et nous espérons que les résultats présentés répondront à leurs attentes.



Enfin nous souhaitons également souligner l'engagement et la complémentarité de notre équipe projet, composée de Jonathan WILD et Marius DELAVOET. Notre collaboration s'est toujours déroulée dans un esprit de respect mutuel, d'écoute et de persévérance. Grâce à un travail réparti, une bonne communication et une volonté commune de produire un rendu de qualité, nous avons pu mener ce projet à bien dans les délais impartis, avec rigueur et enthousiasme.

À toutes et à tous, un grand merci pour votre soutien, vos conseils et votre engagement à nos côtés.

Table des matières

Introduction :	4
Objectifs :	4
I.Récolte des données :	5
1. Identification des sources de données	5
2. Méthodologie de collecte	5
3. Sélection des pays et modélisation complémentaire	6
II.Transformation des données :	7
1. Nettoyage des données	7
2. Harmonisation et structuration	7
3. Normalisation et création de nouvelles variables	8
4. Modélisation de la consommation domestique	8
5. Préparation à la visualisation	8
III.Analyse des données	9
1. Exploration univariée	9
2. Tendances annuelles	12
3. Analyse saisonnière (moyenne mensuelle)	13
4. Corrélations et comparaisons croisées	15
5.Repérer les pays atypiques (résidus extrêmes)	19
7. Différence LCA – Direct par pays	20
8. Histogrammes des résidus (modèle Low Carbon ~ Direct)	21
IV.Importance de l'impact du cloud	23
1. Croissance du cloud	23
2. Explosion de l'IA	23
3. Perspectives d'évolution	24
4. Emissions de CO2 et consommation énergétique	26
5. Consommation d'eau	26
6. Pression sur les ressources et déchets électroniques	26
7. Zoom : Le rôle central des data centers	27
8. Le cas du cloud public : illusion verte ?	27
9. Conclusion	28
V.Partie foyers	29
1. Modélisation des foyers	29
2. Consommation des foyers par pays (officiel)	34

3. Consommation des foyers par pays (calculé)	35
VI.Impact et mise en relation des foyers et du cloud.....	38
1.Comparaison directe des impacts : foyers, cloud et datacenters	38
2.Interprétation générale	38
Conclusion	39
Annexes.....	40

Introduction :

La question environnementale est aujourd'hui au cœur des préoccupations sociétales, économiques et politiques à l'échelle mondiale. Parmi les principaux leviers d'action pour limiter le réchauffement climatique figure la réduction des émissions de gaz à effet de serre, en particulier celles liées à la production et à la consommation d'électricité. En effet, le secteur de l'énergie constitue l'un des plus gros contributeurs aux émissions mondiales de CO₂, et son impact varie considérablement selon les sources utilisées pour produire l'électricité dans chaque pays.

Dans ce contexte, notre projet de SAÉ 2.06 intitulé « Construction et présentation d'indicateurs de performance » s'inscrit dans une démarche d'analyse approfondie de l'intensité carbone du mix électrique de sept pays (France, Allemagne, Pologne, Norvège, Inde, Brésil et États-Unis) entre 2021 et 2024. L'objectif de ce projet est double : d'une part, mesurer et comparer les émissions de CO₂ générées par la production électrique dans chacun de ces pays ; d'autre part, identifier les facteurs explicatifs de ces variations, en s'appuyant sur des analyses statistiques, des représentations graphiques pertinentes et des modèles de régression.

Pour cela, nous avons exploité des données horaires issues d'Electricity Maps, que nous avons nettoyées, transformées et enrichies grâce à plusieurs logiciels, principalement R et Excel. Nous avons ensuite effectué une série d'analyses : calculs de corrélations, modélisation de la relation entre l'intensité carbone et la part d'énergies fossiles, étude des résidus et des écarts typiques ou atypiques entre pays.

Une dimension importante de notre projet réside dans l'étude des usages numériques. Nous avons cherché à quantifier l'impact carbone du cloud et des datacenters selon les pays, en croisant données d'intensité horaire et besoins estimés en kWh par utilisateur ou service (stockage, requêtes, streaming...). Nous avons également modélisé des profils types de foyers domestiques (faible, moyen, gros consommateurs) afin de comparer concrètement leur empreinte carbone électrique dans chaque pays étudié.

Ce rapport est structuré de la manière suivante : nous commencerons par décrire les jeux de données et les traitements appliqués, puis nous présenterons nos analyses statistiques détaillées, nos cas d'usage et nos modélisations, avant de conclure sur les limites de notre étude et nos recommandations pour une production et une consommation électrique plus durables.

Objectifs :

Comparer la consommation d'électricité et l'intensité carbone de la production électrique entre les pays étudiés.

Étudier la composition de leur mix énergétique (renouvelable, bas-carbone, fossile).

Analyser l'évolution de ces indicateurs sur la période 2021-2024.

Examiner en détail les tendances de consommation électrique en France (régions et variations quotidiennes).

Fournir des insights basés sur les données pour orienter les politiques énergétiques durables.

I. Récolte des données :

Dans le cadre de notre projet portant sur l'analyse de l'empreinte carbone de la production d'électricité à l'échelle internationale, une attention particulière a été portée à la rigueur méthodologique dès l'étape de la recherche des données. La pertinence de notre analyse repose en effet directement sur la qualité, la précision et la fiabilité des données collectées.

1. Identification des sources de données

La principale source utilisée pour ce projet a été la plateforme Electricity Maps, recommandée dès le début du projet par notre commanditaire, Mme Claudia Roncancio, lors de la réunion de cadrage. Electricity Maps propose des données en temps réel ou historiques sur la production d'électricité dans de nombreux pays, en précisant l'intensité carbone (en gCO_2/kWh) selon l'origine de l'énergie (renouvelable, fossile, nucléaire, etc.). Cette plateforme constitue donc une référence incontournable pour suivre les variations de l'empreinte carbone horaire, journalière ou annuelle des mix électriques nationaux. Les données y sont téléchargeables sous format CSV, ce qui nous a permis une récupération structurée sur les années 2021 à 2024.

Parallèlement, nous avons enrichi notre base d'analyse avec des données issues d'organismes complémentaires afin de croiser les informations et d'assurer une contextualisation économique et environnementale plus fine, surtout pour ce qui est de la modélisation des foyers ainsi que l'impact du cloud et des datacenters. L'ADEME (Agence de la transition écologique) a été mobilisée pour obtenir des données et rapports spécifiques à la France concernant l'impact environnemental de l'énergie. La plateforme Boavizta a été consultée pour simuler l'impact carbone des infrastructures numériques.

2. Méthodologie de collecte

Les données ont été récupérées, dans la majorité des cas, sous forme de fichiers CSV (toutes les données sur lesquelles nous avons fait des calculs et transformations). Ce format tabulaire a facilité leur manipulation dans Microsoft Excel, et leur intégration dans R.

Pour chaque pays ciblé, nous avons veillé à collecter uniquement les fichiers correspondant à nos variables-clés : intensité carbone (gCO_2/kWh), part des différentes sources d'électricité (renouvelable, fossile, nucléaire), et données de consommation si disponibles. Dans certains cas, l'accès aux données a nécessité l'utilisation de filtres spécifiques sur les plateformes (comme les menus par pays sur Electricity Maps).

Enfin, pour limiter les risques de perte ou de corruption des données, nous avons centralisé l'ensemble des fichiers bruts dans un dossier partagé en ligne sécurisé (Google Drive), organisé par source et par pays. Cela nous a également permis de travailler à plusieurs sur les fichiers sans duplication excessive, et de garantir la continuité du projet même en cas d'imprévu technique.

3. Sélection des pays et modélisation complémentaire

Sept pays ont été retenus comme cibles d'analyse principale : la France, l'Allemagne, la Pologne, la Norvège, l'Inde, le Brésil et les États-Unis. Ce choix a été motivé par leur diversité énergétique : certains sont très carbonés (comme la Pologne), d'autres majoritairement renouvelables (comme la Norvège), et d'autres mixtes ou en transition (comme la France ou l'Inde). Ce panel a permis d'étudier une large gamme de situations représentatives au niveau mondial.

Par ailleurs, nous avons introduit une modélisation complémentaire visant à estimer l'impact de la consommation électrique domestique, via la création de profils types de foyers. Pour cela, nous nous sommes appuyés sur notre propre expérience et les données fournies par Boavizta, mais aussi un large panel de données trouvables sur internet que ce soient les sites des états concernés mais aussi wikipedia qui regroupe un grand nombre de données sur la consommation énergétique par pays. Globalement nous avons créé nos différents profils de pays en considérant les usages numériques (ordinateurs, box, téléviseurs, etc.) et leur durée d'utilisation. Ces profils ont été utiles pour illustrer les effets concrets de l'intensité carbone selon les usages quotidiens.

II. Transformation des données :

Après avoir terminé la phase de collecte, nous avons entrepris un travail rigoureux de transformation de nos données afin de les rendre pleinement exploitables dans le cadre de notre analyse. Cette étape s'est révélée essentielle pour garantir la qualité, la comparabilité et la lisibilité des résultats que nous souhaitions produire. Les transformations ont concerné à la fois la structure des données, leur contenu, et leur adaptation aux objectifs statistiques et visuels du projet.

1. Nettoyage des données

Dès la récupération des fichiers bruts, nous avons procédé à une vérification systématique de leur intégrité. Nous avons commencé par identifier et supprimer tous les doublons, notamment dans les jeux de données horaires où certaines valeurs apparaissaient deux fois suite à des extractions multiples. Ensuite, nous avons traité les valeurs manquantes : lorsqu'il s'agissait de trous isolés, nous avons eu recours à une interpolation linéaire. Dans les cas où les lacunes étaient plus nombreuses ou situées sur de longues périodes, nous avons préféré les exclure de nos calculs pour éviter toute distorsion.

Nous avons également standardisé tous les formats. Les dates ont été unifiées au format AAAA-MM-JJ HH:MM, et les unités ont été harmonisées. Par exemple, toutes les émissions ont été converties en grammes de CO₂ par kilowattheure produit (gCO₂/kWh), afin de maintenir une cohérence d'ensemble entre les pays et les périodes.

2. Harmonisation et structuration

Une fois les données nettoyées, nous les avons restructurées pour garantir leur compatibilité entre elles. Nous avons adopté un format tabulaire uniforme, avec les mêmes noms de colonnes pour chaque fichier, indépendamment du pays. Nous avons également regroupé certaines données selon différents pas de temps : horaires, journaliers, mensuels ou annuels. Cette agrégation nous a permis de passer d'une lecture brute à une lecture plus synthétique des tendances.

Nous avons filtré les colonnes non nécessaires pour nous concentrer uniquement sur les variables utiles à notre étude : intensité carbone, type de production, volume total produit, part d'énergies renouvelables, etc. Cela a aussi facilité le passage vers les outils de visualisation et réduit le risque d'erreurs d'interprétation.

3. Normalisation et création de nouvelles variables

Afin de pouvoir comparer des pays très différents en taille, nous avons introduit plusieurs transformations supplémentaires. En particulier, nous avons rapporté les émissions totales par habitant à partir des données de population issues de la Banque mondiale. Cela nous a permis de produire des indicateurs pertinents d'un point de vue environnemental et de justice énergétique.

Nous avons également calculé des écarts interannuels (par exemple entre 2021 et 2024) pour identifier les dynamiques de transition ou de régression selon les contextes nationaux. Une autre transformation a concerné la création de variables de type "part indirecte" liée à l'usage de l'électricité dans les secteurs numériques : nous avons estimé, à partir du mix énergétique, la part des émissions attribuables aux usages informatiques, y compris les datacenters.

Pour aller plus loin, nous avons conçu des indicateurs sur les heures dites "critiques", en sélectionnant les moments où l'intensité carbone dépassait un seuil que nous avons défini à 500 gCO₂/kWh. Ces heures ont été identifiées dans chaque pays, ce qui nous a permis de comparer les pics d'émission et de comprendre les risques liés aux périodes de forte demande ou de faible disponibilité renouvelable.

4. Modélisation de la consommation domestique

Dans le but de rendre nos résultats plus concrets, nous avons construit plusieurs profils types de foyers. Pour chaque modèle, nous avons estimé la consommation électrique journalière moyenne, puis nous avons croisé cette consommation avec les données d'intensité carbone afin de calculer une empreinte quotidienne, mensuelle et annuelle pour chaque pays. Ce travail nous a permis de relier des données abstraites à des situations concrètes, proches du quotidien.

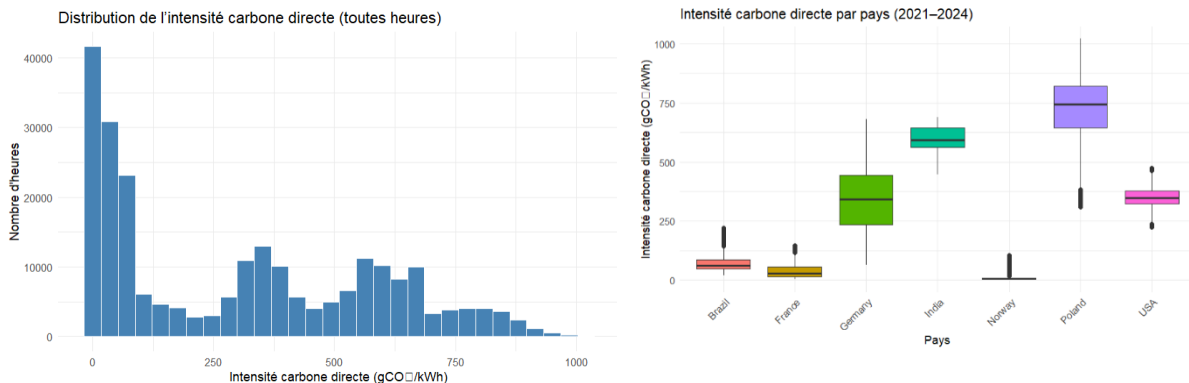
5. Préparation à la visualisation

Enfin, nous avons organisé nos fichiers pour faciliter leur utilisation dans l'outil Tableau. Nous avons structuré les données sous un format long (type "tidy data"), ce qui permet de créer facilement des graphiques dynamiques. Nous avons aussi produit des jeux de données distincts par thème (émissions, mix énergétique, consommation par foyer, heures extrêmes) et ajouté des métadonnées explicites à chaque fichier (définitions des variables, unités, pays concernés, période couverte). Ce travail préparatoire a largement simplifié la phase de construction des visualisations interactives.

III. Analyse des données

Notre analyse porte sur l'intensité carbone horaire de la production d'électricité dans 7 pays (France, Norvège, Brésil, Allemagne, Pologne, Inde, USA) entre 2021 et 2024. L'objectif est de comprendre les profils énergétiques, les variations temporelles, les relations entre indicateurs et les différences d'impact carbone. Cette analyse a été faite sous R, car c'est l'outil statistique le plus approprié du fait de notre accoutumance avec ce dernier lors de notre formation.

1. Exploration univariée



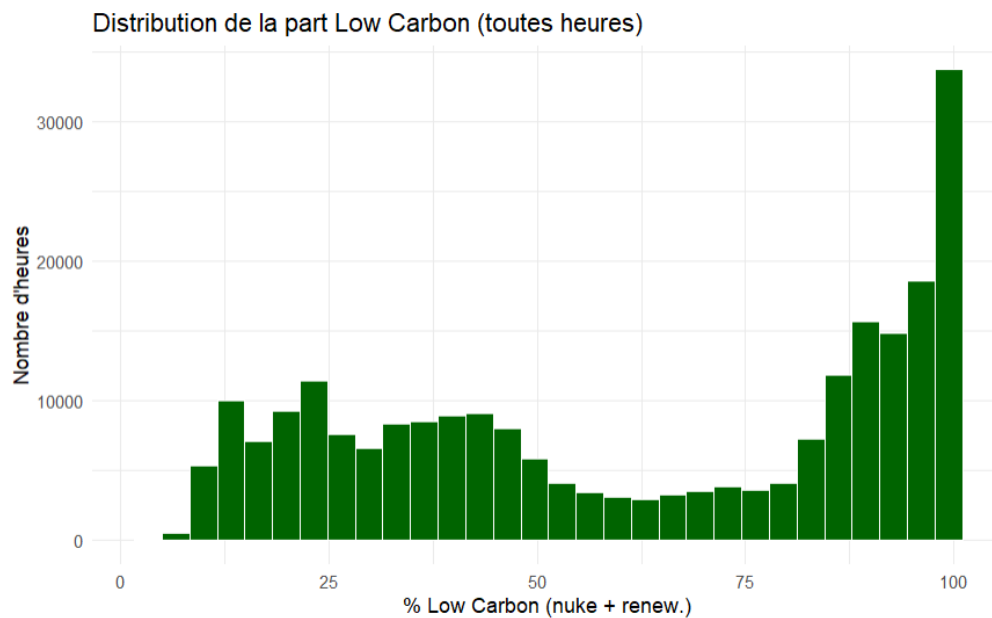
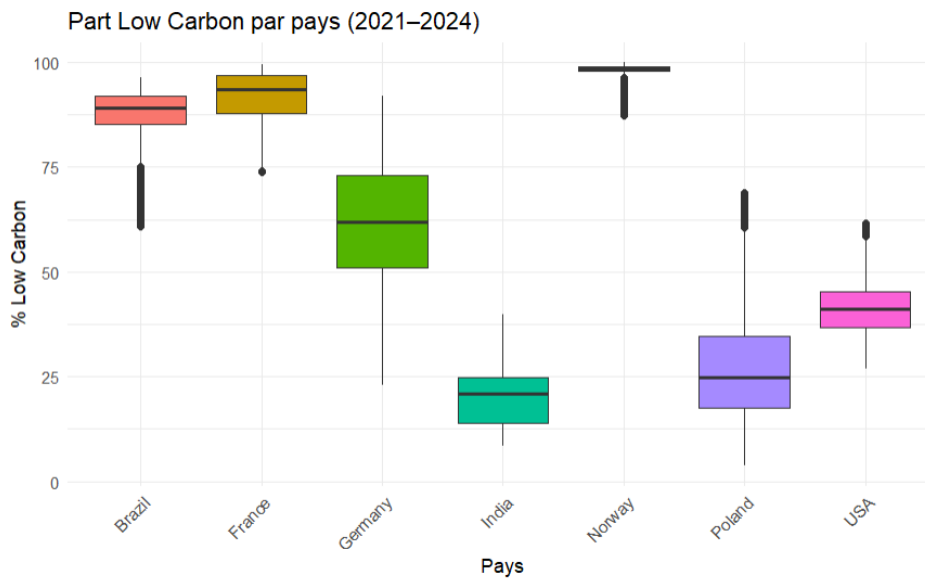
Le premier graphique (histogramme) révèle une répartition très étalée de l'intensité carbone :

- De nombreuses heures sont concentrées autour de 0 à 100 gCO₂/kWh, correspondant à des périodes à forte part d'électricité décarbonée (nucléaire, hydraulique).
- Plusieurs pics sont également visibles entre 300 et 800 gCO₂/kWh, traduisant des périodes dominées par les énergies fossiles (charbon, gaz).

Le second graphique (boxplot par pays) permet de comparer les différents mix électriques :

- **Norvège** : intensité extrêmement faible et très stable, reflet d'un mix dominé par l'hydroélectricité.
- **France** : intensité également très faible et peu variable, due à une forte part de nucléaire.
- **Brésil** : valeurs faibles mais plus dispersées, traduisant un mix composé d'hydroélectricité et de renouvelables.
- **Allemagne** et **Inde** : intensités moyennes élevées et forte dispersion, témoins d'un mix carboné et instable.
- **États-Unis** : intensité modérée, relativement stable.
- **Pologne** : intensité très élevée, conséquence d'une forte dépendance au charbon.

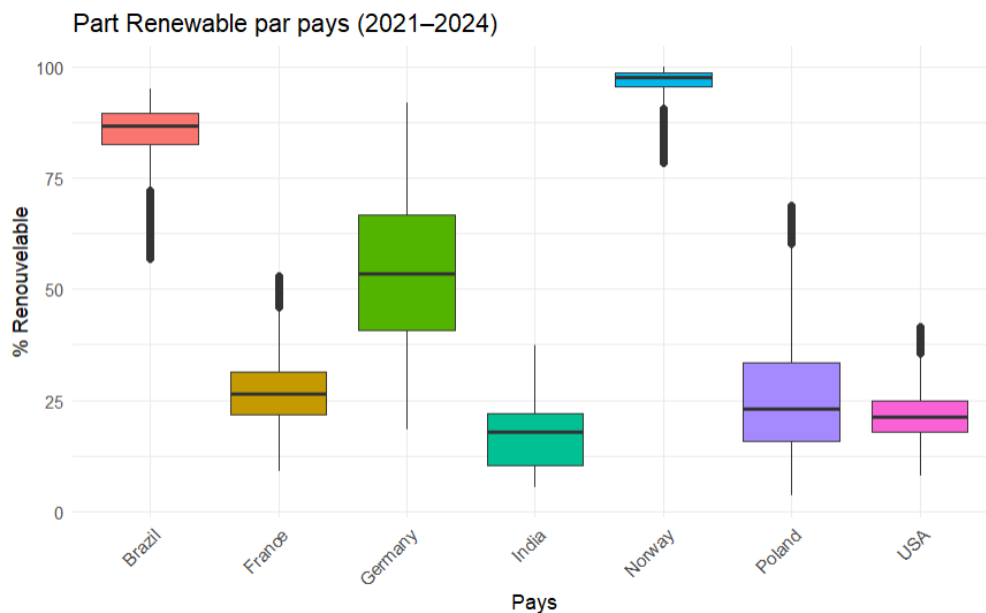
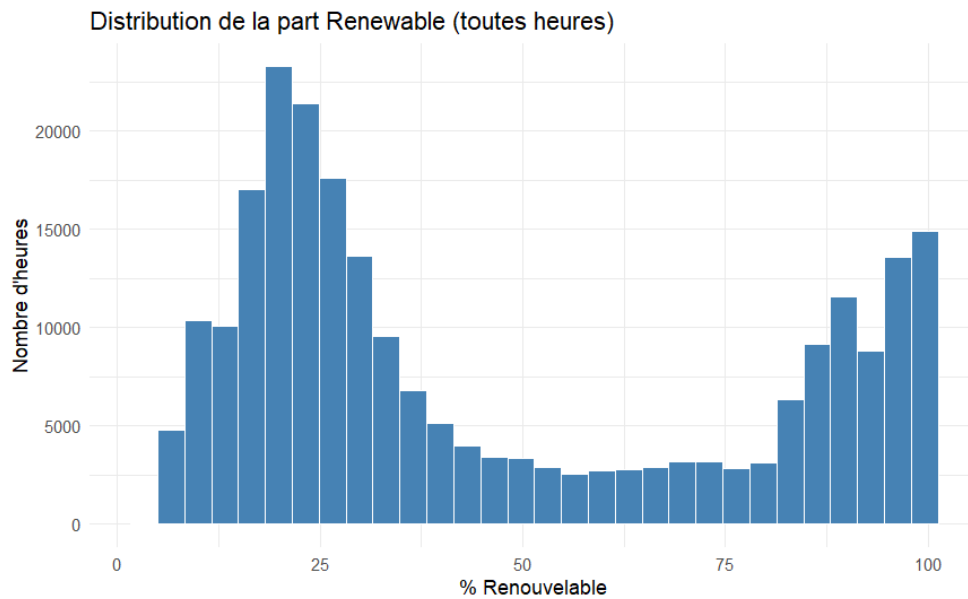
Part Low Carbon (nucléaire + renouvelable)



La distribution montre une bimodalité marquée :

- Un pic autour de 0–20 %, correspondant à un usage élevé d'énergies fossiles.
- Un autre pic proche de 100 %, caractéristique de systèmes très décarbonés (Norvège, France).
- Le boxplot par pays confirme cette hétérogénéité :
- Norvège, France et Brésil présentent une part Low Carbon très élevée et stable.
- Inde, Pologne et États-Unis ont des parts faibles à modérées, avec une forte variabilité.
- Allemagne se situe en position intermédiaire, avec une médiane autour de 70 %.

Part Renewable (partie d'énergie renouvelable)



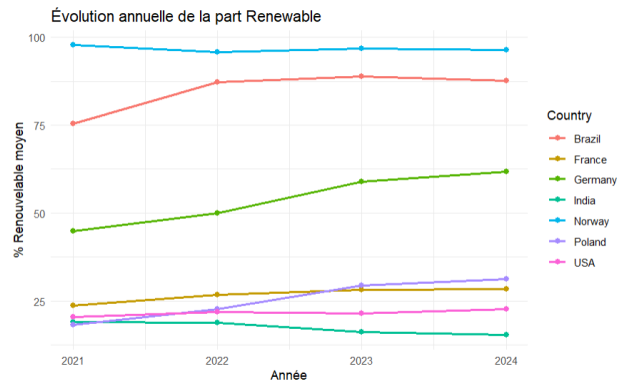
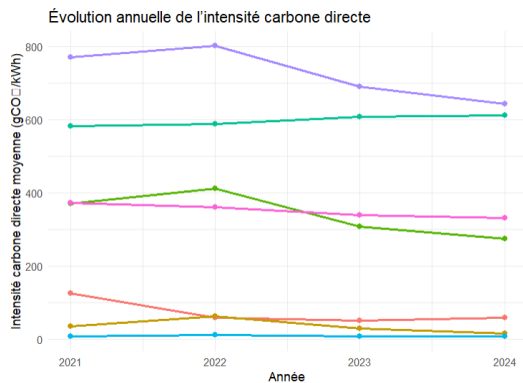
La distribution globale se concentre autour de deux pôles : 20–30 % et près de 100 %, ce qui suggère deux profils distincts :

- Pays très dépendants des renouvelables (Norvège, Brésil).
- Pays avec des parts plus faibles mais relativement stables (France, États-Unis).

Les boxplots par pays indiquent :

- Norvège : près de 100 % d'électricité renouvelable de manière constante.
- Brésil : part très élevée mais plus variable.
- Inde et États-Unis : parts faibles et stables.
- Allemagne : forte variabilité horaire.
- France : position médiane, mais avec une faible dépendance aux renouvelables.

2. Tendances annuelles



Évolution annuelle de l'intensité carbone directe :

Ce graphique montre l'évolution des émissions de CO₂ pour produire un kilowattheure d'électricité, entre 2021 et 2024.

La Pologne reste le pays avec l'électricité la plus polluante. Son intensité carbone passe de 780 à 660 gCO₂/kWh. L'Inde reste stable autour de 600 gCO₂/kWh. L'Allemagne diminue fortement, de plus de 400 à environ 310 gCO₂/kWh. Les États-Unis baissent aussi légèrement, de 380 à 330 gCO₂/kWh. La France est toujours très faible, autour de 90 à 100 gCO₂/kWh. La Norvège reste proche de zéro grâce à l'hydroélectricité. Le Brésil reste aussi relativement bas, autour de 100 à 120 gCO₂/kWh.

En résumé, plus un pays utilise du charbon ou du gaz, plus l'intensité carbone est élevée. Les pays qui utilisent surtout des énergies renouvelables ont une électricité plus propre.

Évolution annuelle de la part d'énergie renouvelable :

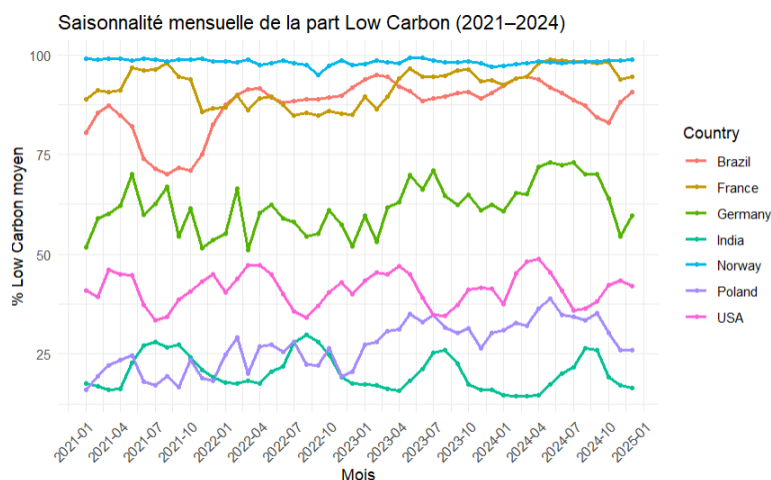
L'autre graphique montre l'évolution de la part des énergies renouvelables dans la production d'électricité entre 2021 et 2024.

La Norvège reste proche de 100 % d'électricité renouvelable. Le Brésil augmente jusqu'à environ 90 %. L'Allemagne progresse fortement, de 45 % à plus de 60 %. La France reste stable autour de 25 %. Les États-Unis, la Pologne et l'Inde restent en dessous de 30 %, mais la Pologne montre une hausse.

Cela montre que plus un pays développe les énergies renouvelables, plus il peut réduire son intensité carbone. La Norvège et le Brésil sont déjà très avancés. L'Allemagne et la Pologne montrent des efforts. La France reste stable mais peut encore progresser.

3. Analyse saisonnière (moyenne mensuelle)

Intensité carbone directe (mensuelle)



Le graphique révèle une variabilité saisonnière marquée pour certains pays.

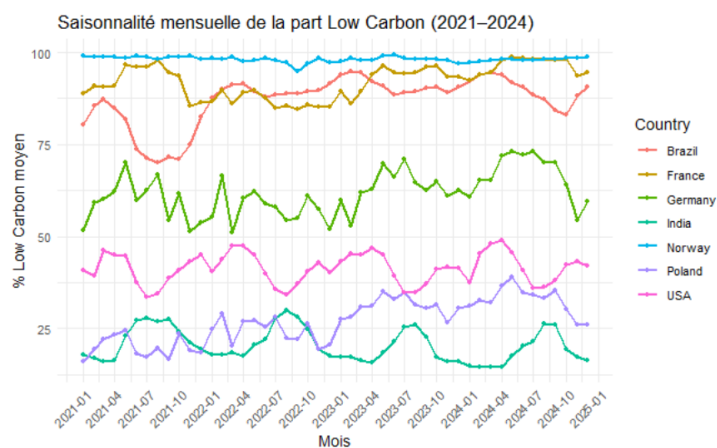
En Allemagne et en Pologne, on observe des hausses hivernales régulières, suggérant un recours accru aux énergies fossiles pour le chauffage.

L'Inde montre une stabilité mensuelle, sans vraie saisonnalité, à un niveau élevé.

Les USA ont une saisonnalité modérée, mais globalement stable autour de 350–400 gCO₂/kWh.

La France, le Brésil et surtout la Norvège affichent une intensité très basse et relativement constante.

Low Carbon Percentage (mensuelle)



Le graphique montre la part moyenne mensuelle de production bas carbone (nucléaire + renouvelable).

La Norvège reste proche de 100 % toute l'année, sans variation notable.

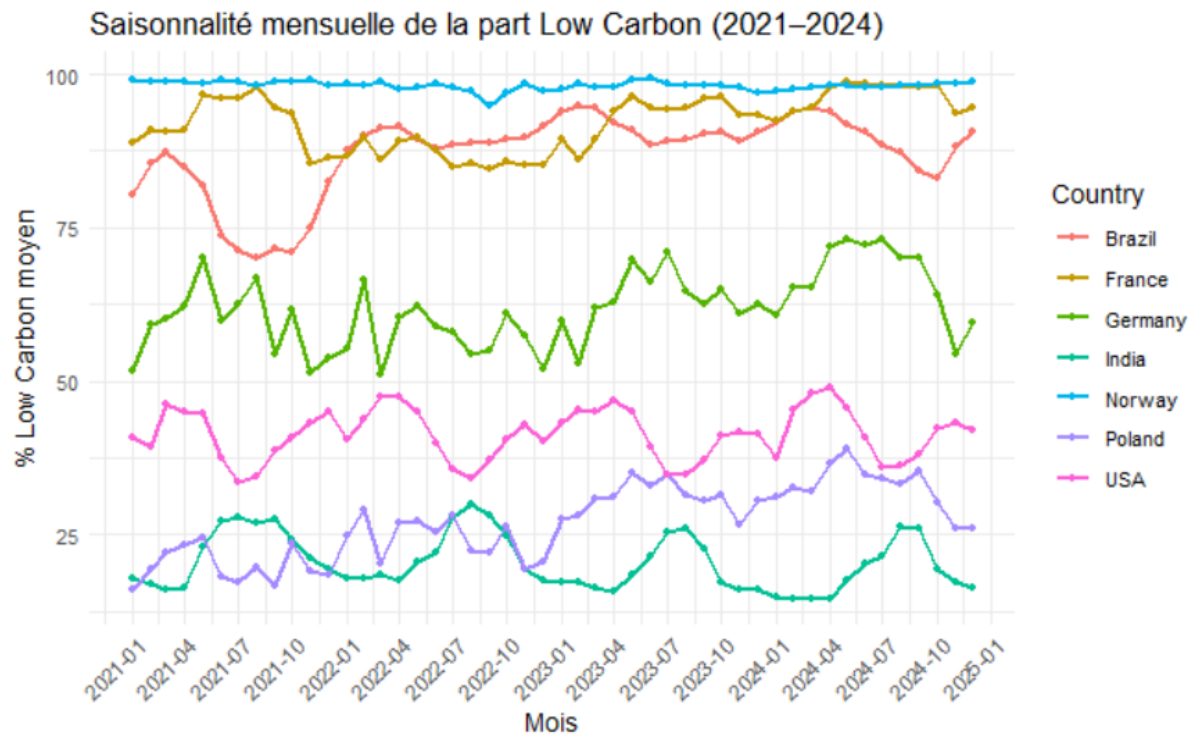
France et Brésil maintiennent également une part élevée (plus de 85 %) avec quelques fluctuations saisonnières.

L'Allemagne montre une saisonnalité visible, avec des creux en hiver et des pics en été.

L'Inde et les États-Unis ont des niveaux faibles et stables, entre 40 % et 50 %.

La Pologne reste à un niveau très bas (moins de 30 %), bien en dessous des autres pays.

Renewable Percentage (mensuelle)



Le graphique montre l'évolution mensuelle moyenne de la part d'électricité produite à partir de sources renouvelables.

La Norvège reste proche de 100 % toute l'année, confirmant une électricité quasi intégralement renouvelable.

Le Brésil affiche également une part très élevée (souvent plus de 85 %), avec une certaine saisonnalité.

L'Allemagne montre une tendance croissante, avec des pics en été, probablement liés au solaire.

L'Inde a une part renouvelable faible et relativement stable.

La Pologne et les États-Unis restent autour de 20–30 %, avec peu de variation.

La France a une part modérée (environ 25–35 %) de renouvelable, l'essentiel de son bas carbone venant du nucléaire.

4. Corrélations et comparaisons croisées

Matrice de corrélation entre indicateurs carbone

Modèle : Carbon.Intensity.gCO.eq.kWh..direct. vs Carbon.Intensity.gCO.eq.kWh..LCA.

$R^2 = 0.999$

EQM = 117.058

Intervalles de confiance (95%) des coefficients :

	2.5 %	97.5 %
(Intercept)	28.041608	28.170502
Carbon.Intensity.gCO.eq.kWh..direct.	1.064039	1.064354

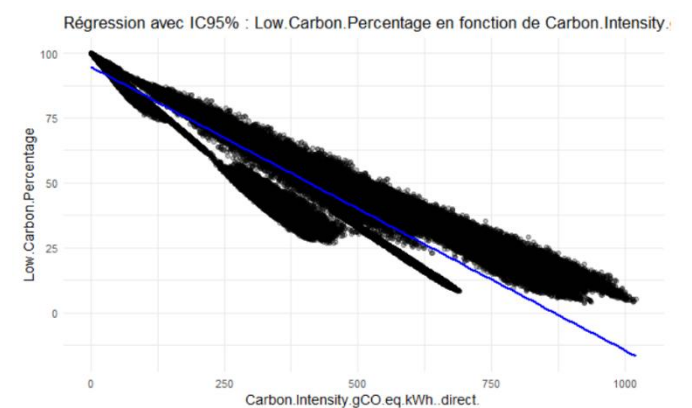
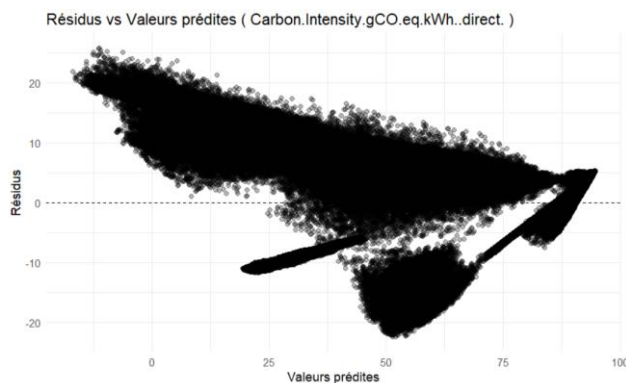
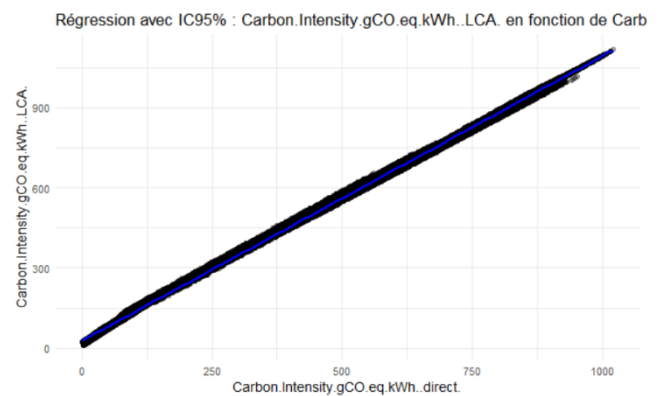
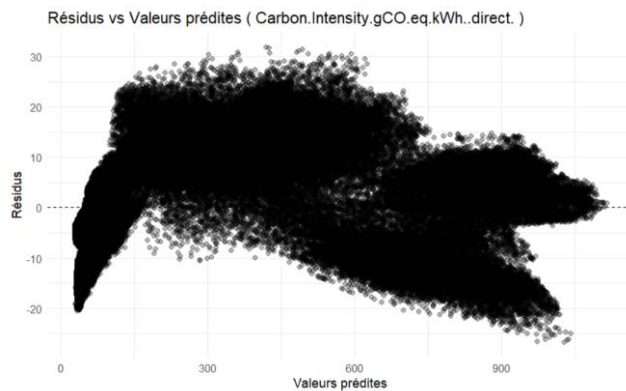
Modèle : Carbon.Intensity.gCO.eq.kWh..direct. vs Low.Carbon.Percentage

$R^2 = 0.924$

EQM = 74.197

Intervalles de confiance (95%) des coefficients :

	2.5 %	97.5 %
(Intercept)	94.6174537	94.7200723
Carbon.Intensity.gCO.eq.kWh..direct.	-0.1092977	-0.1090471



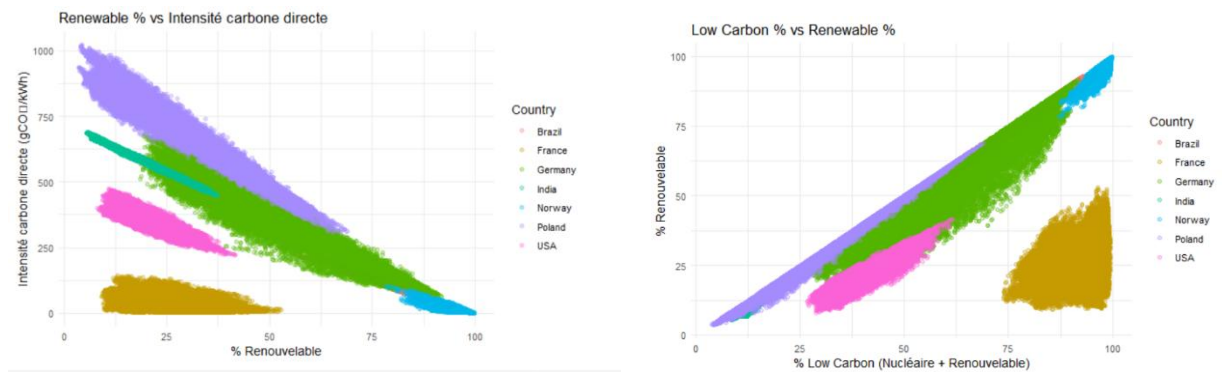
Modèle 1 : Intensité carbone LCA ~ Intensité directe

- Le modèle présente une excellente qualité d'ajustement ($R^2 \approx 0,999$), ce qui montre une relation quasi linéaire parfaite entre l'intensité directe et celle fondée sur l'analyse du cycle de vie (LCA).
- Le coefficient de régression est d'environ 1,064 : en moyenne, l'intensité LCA est 6,4 % plus élevée que l'intensité directe.
- L'intervalle de confiance (IC95 %) du coefficient est très resserré (entre 1,0640 et 1,0644), traduisant une grande précision du modèle.

Modèle 2 : Part Low Carbon ~ Intensité directe

- Ce modèle présente aussi de très bonnes performances ($R^2 \approx 0,924$).
- La relation est négative : plus l'intensité carbone augmente, plus la part d'électricité bas carbone diminue.
- Le coefficient de régression est proche de $-0,109$, ce qui signifie qu'une hausse de 100 gCO₂/kWh entraîne une baisse d'environ 11 points de la part Low Carbon.
- L'IC95 % est plus large ($-0,1093$ à $-0,1090$), indiquant une dispersion légèrement plus importante des données autour de la droite de régression.

Scatter plots



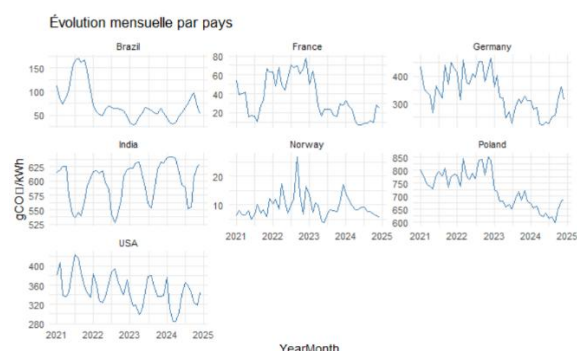
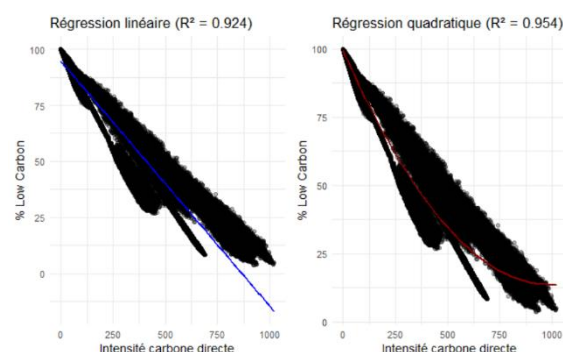
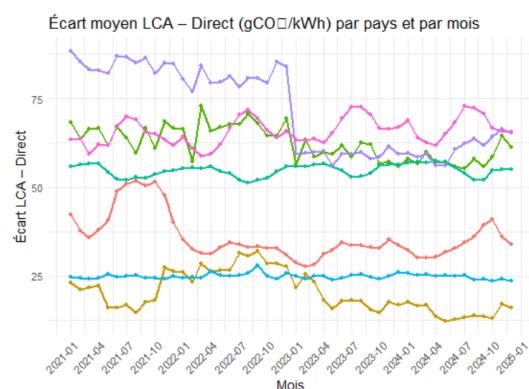
Renouvelable (%) vs Intensité carbone directe :

- Relation inverse marquée : plus la part d'énergies renouvelables est élevée, plus l'intensité carbone est faible.
- Les pays forment des groupes distincts : la Norvège et le Brésil apparaissent à haute part renouvelable et faible intensité, à l'opposé de la Pologne.

Low Carbon (%) vs Renouvelable (%) :

- Relation linéaire croissante.
- La France se distingue : part renouvelable modérée mais part Low Carbon élevée (effet du nucléaire).
- La Norvège et le Brésil sont proches de la diagonale, leur mix étant essentiellement renouvelable.

Écarts mensuels



[1] " Matrice de corrélation entre indicateurs :"

	Carbon.Intensity.gCO ₂ .eq.kWh..direct.	
Carbon.Intensity.gCO ₂ .eq.kWh..direct.	1.000	
Carbon.Intensity.gCO ₂ .eq.kWh..LCA.	0.999	
Low.Carbon.Percentage	-0.961	
Renewable.Percentage	-0.686	
	Renewable.Percentage	
Carbon.Intensity.gCO ₂ .eq.kWh..direct.	-0.686	
Carbon.Intensity.gCO ₂ .eq.kWh..LCA.	-0.682	
Low.Carbon.Percentage	0.752	
Renewable.Percentage	1.000	

Les graphiques montrent des dynamiques contrastées de l'intensité carbone selon les pays entre 2021 et 2025.

L'écart entre les intensités LCA et directe est particulièrement élevé en Pologne et aux États-Unis, traduisant une forte présence d'émissions indirectes, souvent liées au charbon. À l'inverse, il reste faible et stable en Norvège, dont le mix est quasi exclusivement hydraulique.

La comparaison des modèles de régression entre l'intensité directe et la part Low Carbon révèle qu'un modèle quadratique ($R^2 = 0,954$) décrit mieux la relation qu'un modèle linéaire ($R^2 = 0,924$), en capturant les effets de seuils aux extrémités.

La matrice de corrélation confirme une très forte interdépendance entre les indicateurs : l'intensité directe est négativement corrélée à la part Low Carbon (-0,961) et à la part renouvelable (-0,686), tandis que la part Low Carbon est positivement liée à la part renouvelable (0,752).

Enfin, l'évolution mensuelle de l'intensité carbone met en évidence :

- Une stabilité élevée en Inde (~600 gCO₂/kWh),
- Des baisses marquées en Allemagne et Pologne à partir de 2023,
- Une intensité très faible et stable en Norvège et France.

Ces résultats illustrent la diversité des transitions énergétiques nationales et l'importance de considérer les émissions indirectes dans l'analyse carbone.

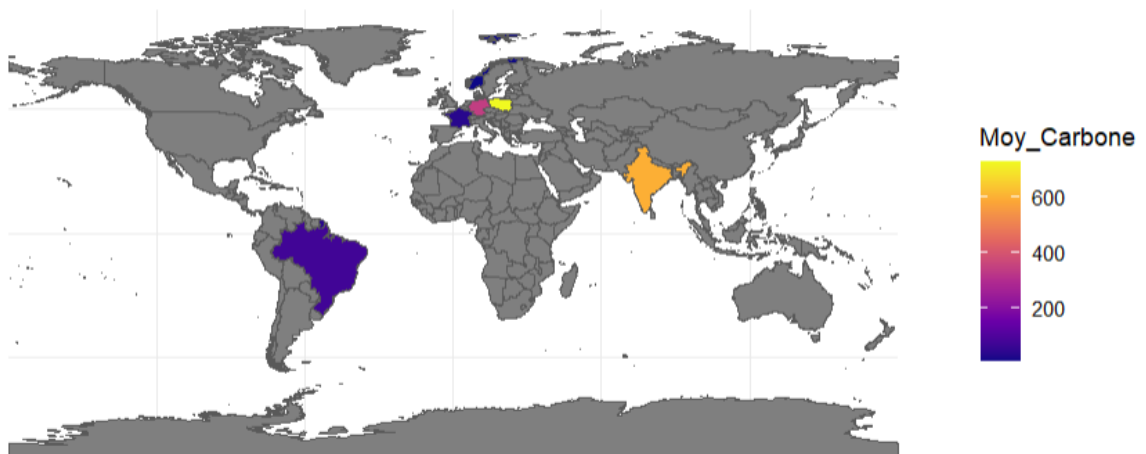
5. Repérer les pays atypiques (résidus extrêmes)

Country <chr>	Residus_Moy <dbl>	Residus_AbsMoy <dbl>	Explication <chr>
USA	-15.2878613	15.287861	Le modèle surestime fortement la part Low Carbon des USA. Erreur moyenne la plus élevée.
Poland	11.4646236	11.464624	Le modèle sous-estime la part Low Carbon en Pologne. Les valeurs observées sont plus élevées.
India	-9.1264955	9.126495	Le modèle surestime la part Low Carbon en Inde, dans une moindre mesure que pour les USA.
Norway	4.5460983	4.546102	Le modèle sous-estime légèrement la Norvège. Prédiction assez proches.
Germany	4.3393137	4.674583	Erreur modérée pour l'Allemagne. Certains points mal ajustés.
France	1.4374423	2.685835	Bon ajustement en France. Faible biais et erreur modérée.
Brazil	0.5223128	2.155947	Très bon ajustement au Brésil. Prédiction proches des observations.

- USA : forte surestimation du modèle (résidu ≈ -15) → inadéquation marquée.
- Pologne : sous-estimation importante (résidu $\approx +11$).
- Inde : surestimation modérée (résidu ≈ -9).
- Norvège : légère sous-estimation (résidu $\approx +4,5$).
- Allemagne : erreur modérée (résidu $\approx +4,3$).
- France : faible biais (résidu $\approx +1,4$) → bon ajustement.
- Brésil : excellent ajustement (résidu $\approx +0,5$).

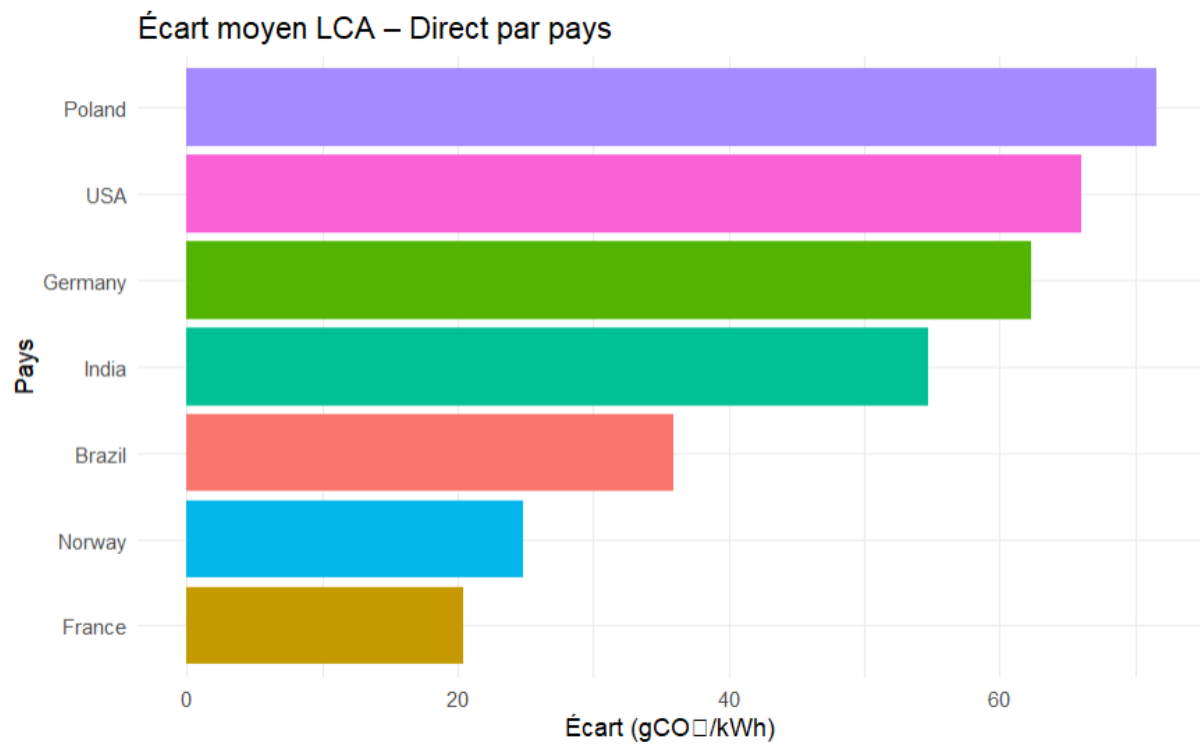
6. Carte des pays avec indicateur carbone moyen

Carte : Intensité carbone moyenne (2021–2024)



- Représente l'intensité moyenne directe en gCO₂/kWh par pays.
- Plus la teinte est foncée, plus l'intensité est faible.
- Pologne, Inde : valeurs très élevées (> 600 gCO₂/kWh).
- France, Norvège, Brésil : intensité faible grâce au nucléaire ou à l'hydroélectricité.

7. Différence LCA – Direct par pays



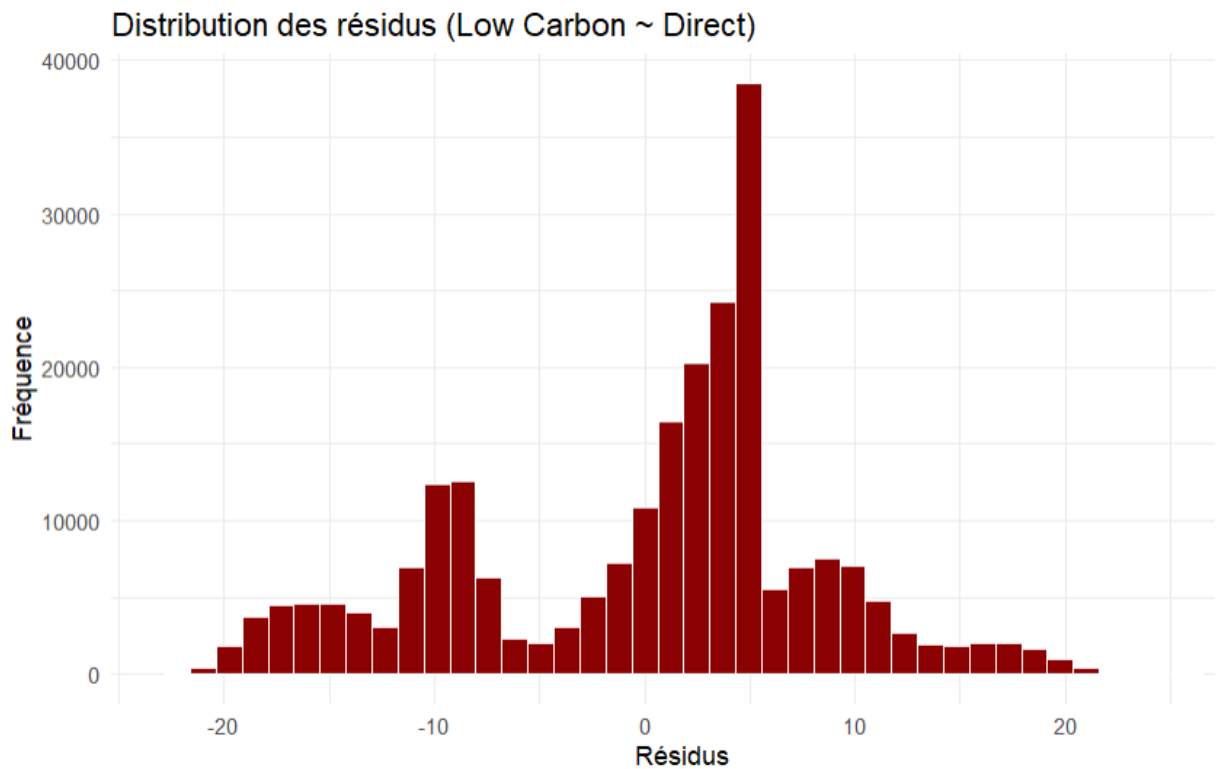
Cet écart indique les émissions indirectes (construction, extraction, transport).

Pologne, USA, Allemagne : écarts élevés → mix carboné, impact indirect fort.

France, Norvège : écarts faibles → production très propre même en cycle de vie complet.

Brésil, Inde : écarts intermédiaires, influencés par les conditions saisonnières.

8. Histogrammes des résidus (modèle Low Carbon ~ Direct)



Les résidus sont bien centrés autour de zéro, ce qui est bon signe.

Distribution légèrement asymétrique :

- Queue à gauche : surestimation du modèle.
- Queue à droite : sous-estimation.

Conclusion :

Le modèle reste robuste malgré une relation légèrement non linéaire donc la qualité globale de la prédiction est satisfaisante.

Conclusion partielle – Ce que révèlent les données

L'analyse des données d'intensité carbone directe de la production d'électricité met en évidence une forte hétérogénéité entre les pays. Deux groupes principaux se distinguent clairement : d'un côté, la Norvège, la France et le Brésil, qui affichent des intensités très faibles, grâce à un mix énergétique fortement décarboné (hydroélectricité pour la Norvège et le Brésil, nucléaire pour la France) ; de l'autre, la Pologne et l'Inde, qui présentent les intensités les plus élevées, dépassant fréquemment les 600 gCO₂/kWh, en raison de leur dépendance persistante au charbon.

Les graphiques montrent aussi que les profils d'évolution varient fortement selon les pays. Par exemple, la Pologne et l'Allemagne ont amorcé une baisse notable de leur intensité depuis 2022, traduisant un recul progressif du charbon. En revanche, l'Inde reste stable à un niveau élevé, sans amélioration visible.

D'autres aspects confirment cette diversité : la distribution bimodale de la part d'électricité bas carbone reflète une fracture énergétique entre les pays très décarbonés et ceux encore dépendants des énergies fossiles. Enfin, les données révèlent une saisonnalité marquée dans certains pays (notamment en Europe) et des extrêmes horaires très contrastés, allant de moins de 1 gCO₂/kWh en Norvège à plus de 1000 en Pologne.

En somme, ces constats soulignent que le niveau d'émissions directes est un indicateur très sensible à la structure du mix énergétique. Ils illustrent aussi le rôle essentiel que peuvent jouer les politiques publiques et les choix technologiques dans la réduction de l'empreinte carbone de l'électricité.

IV.Importance de l'impact du cloud

1. Croissance du cloud

Le cloud a connu une ascension fulgurante ces dernières années, au point de devenir une pierre angulaire des infrastructures informatiques modernes. En 2022, le marché mondial était estimé à 446,51 milliards de dollars. Cette croissance ne s'est pas ralentie : il a atteint 500 milliards en 2023, puis 676,29 milliards en 2024. Selon les projections, le marché du cloud pourrait dépasser les 1 300 milliards de dollars d'ici 2025, avec un taux de croissance annuel de 16,6 % jusqu'en 2032 (Fortune Business Insights).

Cette dynamique s'explique par l'adoption massive des services cloud : en 2023, 94 % des entreprises déclaraient les utiliser. Dans le même temps, les dépenses en infrastructures cloud ont bondi de 22 % en un an, pour atteindre 91 milliards USD au 4e trimestre 2024 (EdgeDelta).

2. Explosion de l'IA

L'intelligence artificielle suit une trajectoire similaire. Elle devrait contribuer à hauteur de 15 700 milliards de dollars à l'économie mondiale d'ici 2030, soit une hausse de 23 % du PIB global (Greenly, 2024). Environ 49 % des dirigeants du secteur technologique considèrent l'IA comme un pilier de leur stratégie, et 37 % des salariés travaillent déjà dans des structures utilisant activement cette technologie.

L'utilisation de l'IA s'est généralisée : assistants virtuels, outils de création de contenu, prédictions financières, ou encore automatisation de processus dans les secteurs de la santé, de la logistique ou du marketing. Des outils comme ChatGPT, qui comptait déjà 300 millions d'utilisateurs en 2024, illustrent cette adoption de masse.

3. Perspectives d'évolution

L'expansion continue des technologies numériques telles que le cloud et l'intelligence artificielle s'accompagne d'une transformation en profondeur de l'infrastructure énergétique mondiale. Depuis le début des années 2020, les besoins en calcul intensif ont explosé sous l'effet de la généralisation des services à haute disponibilité, du télétravail, de la vidéo à la demande et, plus récemment, de l'adoption massive de modèles d'intelligence artificielle générative. Cette demande ne montre aucun signe de ralentissement.

Les centres de données, éléments clefs de cette révolution, voient leurs capacités augmenter à un rythme sans précédent. D'après les projections de l'Agence Internationale de l'Énergie (IEA), l'empreinte énergétique des data centers devrait bondir de près de 60 % entre 2022 et 2030, pour atteindre environ 1 065 térawattheures (TWh) par an. Cela représente plus que la consommation électrique annuelle de pays entiers comme le Japon ou l'Allemagne.

Aux États-Unis, qui concentrent une part importante des infrastructures numériques mondiales, la consommation des data centers atteignait déjà 176 TWh en 2023, soit 4,4 % de la consommation nationale. En Chine, les investissements massifs dans les technologies IA s'accompagnent d'une explosion du nombre de centres de calcul, souvent alimentés par des centrales à charbon, ce qui alourdit considérablement leur empreinte carbone.

En parallèle, l'Europe connaît une croissance forte mais plus maîtrisée, grâce à une régulation plus stricte et à un mix énergétique souvent plus vert. La France, par exemple, bénéficie d'un avantage comparatif en raison de son recours massif au nucléaire et aux énergies renouvelables, ce qui limite l'intensité carbone du kilowattheure consommé. Toutefois, cette situation est fragile : les besoins croissants en puissance de calcul pourraient mettre sous tension les réseaux nationaux d'ici la fin de la décennie.

La montée en puissance des applications d'IA génère une demande sans précédent en ressources matérielles et énergétiques. Selon l'IEA, l'intelligence artificielle pourrait représenter entre 3 et 4 % de la consommation électrique mondiale d'ici 2030, un chiffre susceptible de grimper plus encore avec l'essor des modèles multi-agents et des applications décentralisées. Cette trajectoire suggère que sans rupture technologique ou régulation drastique, le numérique pourrait devenir l'un des secteurs les plus émissifs d'ici la fin du siècle.

La demande n'est pas uniquement tirée par les entreprises du secteur technologique. Les gouvernements, les établissements d'enseignement supérieur, les banques, les compagnies d'assurance et les services de santé investissent eux aussi massivement dans le cloud et l'IA pour optimiser leurs opérations, réduire les coûts ou moderniser leurs infrastructures. Chaque secteur apporte ainsi sa part à l'intensification de la pression énergétique et environnementale.

Enfin, l'hyperscalabilité – cette capacité à étendre les services numériques à l'échelle planétaire – rend la maîtrise de cette croissance encore plus complexe. À mesure que les plateformes numériques deviennent omniprésentes, la demande en calcul et en stockage suit une courbe exponentielle. Le développement de l'IA générative, en particulier, impose des charges de calculs considérables, tant pour l'entraînement que pour l'inférence, engendrant une consommation continue de ressources énergétiques et matérielles.

Les besoins en calcul intensif et en stockage associés à ces technologies ont entraîné une croissance exponentielle des centres de données. Leur expansion est estimée à 28 % d'ici 2030. Cette progression est directement liée à la multiplication des applications basées sur l'IA, dont la consommation électrique pourrait atteindre jusqu'à 4 % de la demande mondiale d'ici la fin de la décennie (IEA).

Aux États-Unis, la consommation des data centers était déjà de 176 TWh en 2023, représentant 4,4 % de la consommation nationale. À l'échelle mondiale, ce chiffre pourrait grimper à 1 065 TWh d'ici 2030, équivalant à la consommation annuelle de certains pays industrialisés.

4. Emissions de CO2 et consommation énergétique

Le secteur numérique représente aujourd'hui entre 3 et 4 % des émissions mondiales de gaz à effet de serre (GES). Les centres de données à eux seuls sont responsables de 1 à 2 % de ces émissions, soit l'équivalent du secteur aérien. À l'échelle de l'Union européenne, le numérique pourrait consommer jusqu'à 7 % de l'électricité totale d'ici 2030.

Les serveurs dédiés à l'IA sont particulièrement gourmands : leur consommation peut atteindre jusqu'à 100 kW, contre 7 kW pour les serveurs classiques. Cette consommation est majoritairement liée à l'inférence (60 à 70 %), c'est-à-dire l'exécution continue des modèles, plutôt qu'à leur entraînement (20 à 40 %) (Greenly, 2024).

Un exemple emblématique : l'entraînement du modèle GPT-4 génère à lui seul 7 138 tonnes de CO₂e, soit plus de 4 300 vols aller-retour Paris-New York. En phase d'utilisation, la réponse à 1 million de mails par mois via GPT-4 ajoute 42 900 kg de CO₂e.

La répartition géographique de ces émissions révèle des disparités notables : les États-Unis concentrent une grande part des infrastructures, avec un mix électrique encore largement carboné. À l'inverse, des pays comme la France, dont l'énergie est majoritairement d'origine nucléaire et renouvelable, affichent une empreinte par calcul nettement réduite.

En Chine, l'expansion rapide des data centers dans des régions alimentées par le charbon fait exploser l'empreinte carbone du cloud : les émissions peuvent y dépasser 700 g de CO₂ par kWh, contre seulement 50 g en France.

5. Consommation d'eau

Au-delà du carbone, les technologies IA sont également gourmandes en ressources hydriques. Chaque requête de 100 mots via ChatGPT-4 utilise en moyenne 519 mL d'eau, nécessaire au refroidissement des serveurs (U. of California, 2024). D'ici 2030, les data centers pourraient absorber jusqu'à 9 % de l'eau industrielle dans certaines régions du globe.

6. Pression sur les ressources et déchets électroniques

L'accélération de l'IA s'accompagne d'une hausse de la demande en composants spécialisés (GPU, ASIC), intensifiant l'extraction de ressources rares. En 2022, le monde a produit 53 millions de tonnes de DEEE, un chiffre appelé à croître avec l'obsolescence rapide des équipements.

7. Zoom : Le rôle central des data centers

Les data centers sont devenus les véritables épicentres du cloud et de l'IA. Ils stockent les données, alimentent les algorithmes et assurent le fonctionnement de nos outils numériques quotidiens.

Un centre de données standard consomme entre 1 et 5 MW ; les plus grands, dits "hyperscale", peuvent dépasser 100 MW. En 2030, leur consommation mondiale pourrait atteindre 945 TWh (IEA, 2024).

L'efficacité énergétique d'un data center est mesurée via l'indicateur PUE (Power Usage Effectiveness). Si un PUE de 1,1 est excellent, la moyenne mondiale reste autour de 1,6. Des solutions d'amélioration existent : refroidissement par air extérieur, construction dans les zones froides, optimisation logicielle. Mais la croissance de la demande contrebalance souvent ces efforts.

D'autres enjeux s'ajoutent : les fuites de réfrigérants, l'artificialisation des sols pour construire les centres, et l'accroissement des investissements matériels liés à l'IA (serveurs, batteries, onduleurs).

8. Le cas du cloud public : illusion verte ?

Selon une étude Boavizta de 2024, si 64 % des entreprises pensent que le cloud réduit leur empreinte carbone, seules 33 % utilisent effectivement les outils de mesure des fournisseurs. De plus, 58 % ne savent pas quelle part de leur impact est liée au cloud. Les mécanismes de calcul fournis par AWS, Azure ou GCP ne prennent généralement pas en compte le scope 3, soit les émissions liées à la fabrication du matériel.

Boavizta propose des outils plus transparents (Boaviztapi, Datavizta, Cloud Scanner) pour combler ce manque de visibilité.

9. Conclusion

Cette analyse permet donc de mettre en lumière les dynamiques impressionnantes qui façonnent le développement du cloud et de l'intelligence artificielle, tout en soulignant leurs effets collatéraux sur l'environnement. Ce qui apparaissait initialement comme une révolution technologique libératrice s'avère être, aussi, une source majeure d'externalités négatives.

Le cloud, devenu omniprésent, structure aujourd'hui l'architecture des systèmes d'information dans le monde entier. Son adoption généralisée s'est faite au nom de la flexibilité, de la scalabilité et de la réduction des coûts, mais elle s'accompagne d'une consommation énergétique exponentielle. L'IA, quant à elle, redéfinit les usages et les métiers, tout en entraînant des besoins gigantesques en traitement de données, ce qui alimente un cercle vicieux d'extension continue des infrastructures matérielles.

Les données présentées révèlent une réalité préoccupante : les technologies numériques, pourtant souvent perçues comme propres ou immatérielles, génèrent une empreinte environnementale équivalente ou supérieure à celle de secteurs historiquement émissifs comme le transport aérien ou l'industrie lourde. Si des écarts existent entre pays, ils sont loin d'être marginaux : un même traitement algorithmique peut engendrer une empreinte carbone dix fois plus importante selon qu'il est exécuté en France, en Chine ou aux États-Unis.

Les impacts ne se limitent pas aux émissions de CO₂. La consommation d'eau, les déchets électroniques et la surexploitation des ressources minières sont autant de signaux d'alarme. En d'autres termes, ce sont toutes les dimensions de la durabilité – environnementale, énergétique, mais aussi sociale et géopolitique – qui sont concernées.

Le cloud et l'IA ne sont pas intrinsèquement incompatibles avec les objectifs climatiques. Mais leur trajectoire actuelle, non régulée, non optimisée et peu transparente, laisse peu de place à une transition véritablement soutenable. L'urgence est donc à la prise de conscience, à la transparence des données d'impact, et à l'invention de nouvelles approches numériques plus sobres, résilientes et respectueuses des limites planétaires.

Le cloud et l'intelligence artificielle sont des leviers majeurs de transformation numérique, mais ils engendrent un coût environnemental considérable. Leur consommation énergétique explose, leurs émissions de CO₂ atteignent des niveaux comparables aux plus gros secteurs émetteurs, et leur demande en eau et en matières premières est croissante.

Sources : Greenly (2024), Boavizta (2023-2024), IEA, Fortune Business Insights, EdgeDelta, U.C. Riverside, Pew Research Center.

V.Partie foyers

1. Modélisation des foyers

Nous avons créé cinq profils de foyers (studio, couple, couple senior, famille et grande famille) pour représenter la majorité des logements et comprendre d'un coup d'œil où part l'électricité. Pour chaque appareil, nous avons pris les consommations moyennes publiées par l'ADEME ou les fabricants, puis nous avons additionné les valeurs. Seul le chauffage a été ajusté : nous avons réduit ou augmenté ses kWh afin que la moyenne de l'ensemble des profils atteigne environ 5 000 kWh par foyer, qui est la consommation réelle observée en France. Ainsi, les tableaux restent simples à lire tout en reflétant la réalité et en permettant de tester facilement des scénarios (par exemple : ajouter une pompe à chaleur ou de nouveaux appareils).

Studio (1 adulte)

Catégorie	Appareil (quantité)	kWh/an
Chauffage	Radiateur 1 kW (x1)	1400
Froid & cuisson	Réfrigérateur 1 porte	174
Froid & cuisson	Plaque vitrocéramique	159
Froid & cuisson	Micro-ondes	39
Lavage	Lave-linge	90
Boissons	Bouilloire	60
TIC / média	Téléviseur LED	187
TIC / média	Box Internet	97
TIC / média	Ordinateur portable	146
	TOTAL	2352

Couple (1–2 pièces)

Catégorie	Appareil (quantité)	kWh/an
Chauffage	Radiateurs 1 kW (x2)	2300
Froid & entretien	Réfrigérateur combiné	346
Froid & entretien	Lave-vaisselle	192
Froid & entretien	Lave-linge	101
Froid & entretien	Sèche-linge	301
Cuisson	Four électrique	146
Cuisson	Plaque vitrocéramique	159
Cuisson	Micro-ondes	39
Boissons	Bouilloire	60
TIC / média	Téléviseur LED	187
TIC / média	Box Internet	97
TIC / média	Console de jeux	103
TIC / média	Ordinateurs portables (x2)	292
	TOTAL	4323

Famille (2 enfants)

Catégorie	Appareil (quantité)	kWh/an
Chauffage	Radiateurs 1 kW (x3)	3360
Froid & entretien	Réfrigérateur combiné	346
Froid & entretien	Congélateur	308
Froid & entretien	Lave-vaisselle	192
Froid & entretien	Lave-linge	150
Froid & entretien	Sèche-linge	301
Cuisson	Four électrique	146
Cuisson	Plaque vitrocéramique	159
Cuisson	Micro-ondes	39
Boissons	Bouilloire	80
TIC / média	Téléviseur LED	187
TIC / média	Box Internet	97
TIC / média	Console de jeux	103
TIC / média	Ordinateur de bureau	292
TIC / média	Ordinateurs portables (x3)	438
	TOTAL	6199

Grande famille + bureau

Catégorie	Appareil (quantité)	kWh/an
Chauffage	Radiateurs 1 kW (x4)	3120
Chauffage	PAC / clim centralisée	500
Froid & entretien	Réfrigérateur combiné	346
Froid & entretien	Congélateur	308
Froid & entretien	Lave-vaisselle	192
Froid & entretien	Lave-linge	200
Froid & entretien	Sèche-linge	400
Cuisson	Four électrique	146
Cuisson	Plaque vitrocéramique	159
Cuisson	Micro-ondes	39
Boissons	Bouilloire	100
TIC / média	Téléviseurs LED (x2)	374
TIC / média	Box Internet	97
TIC / média	Consoles de jeux (x2)	206
TIC / média	Ordinateurs de bureau (x2)	584
TIC / média	Ordinateurs portables (x4)	584
	TOTAL	7606

Couple senior (maison)

Catégorie	Appareil (quantité)	kWh/an
Chauffage	PAC air/eau 2 kW	3000
Ventilation / confort	Ventilateurs colonne (x2)	60
Froid & entretien	Réfrigérateur combiné	346
Froid & entretien	Congélateur coffre	308
Froid & entretien	Lave-vaisselle	192
Froid & entretien	Lave-linge	101
Cuisson	Four électrique	146
Cuisson	Plaque induction	190
Cuisson	Micro-ondes	39
Boissons	Bouilloire	60
TIC / média	Téléviseur LED	187
TIC / média	Box Internet	97
TIC / média	Ordinateur portable	146
TIC / média	Radios numériques (x2)	30
	TOTAL	4906

2. Consommation des foyers par pays (officiel)

Pays	Consommation moyenne par foyer (kWh/an)	Année de référence
Norvège	15 722	2023
États-Unis	10 791	2022
Inde	≈ 5 200	2022-23
France	4 679	2024
Allemagne	3 383	2021
Brésil	≈ 2 590	2025
Pologne	≈ 1 840	2023

Comment ces chiffres ont été obtenus :

Lorsque la statistique “kWh par foyer” existe déjà

- **Norvège et Allemagne** : les instituts statistiques (SSB et Destatis) publient directement la consommation annuelle moyenne par ménage. Nous avons repris ces valeurs brutes (15 722 kWh et 3 383 kWh).

Lorsque seul le “kWh par client résidentiel” est publié

- **États-Unis** : la FAQ de l’EIA donne la consommation annuelle moyenne d’un client résidentiel (10 791 kWh). Aucune opération supplémentaire n’est nécessaire.

Cas où l’on ne dispose que de données agrégées

- **France** : la Commission de régulation de l’énergie (CRE) communique chaque année la consommation moyenne par compteur résidentiel (4 679 kWh / foyer en 2024).
- **Inde** : seules les données par habitant (1 331 kWh / hab.) sont publiques. Nous avons multiplié par la taille moyenne des ménages issue du ministère des Statistiques (≈ 3,9 pers./foyer) pour obtenir ≈ 5 200 kWh/foyer.
- **Brésil** : CEIC donne la consommation résidentielle mensuelle (15 637 GWh en janv. 2025). Supposant qu’elle reflète l’année, on annualise (× 12) puis on divise par le nombre de ménages IBGE (72,5 millions) ⇒ ≈ 2 590 kWh/foyer/an.
- **Pologne** : Enerdata indique la consommation totale (150 TWh) et la part “ménages” (20 %). On obtient 30 TWh résidentiels. Divisé par ~16,3 millions de ménages (GUS), cela donne ≈ 1 840 kWh/foyer/an.

En regroupant ces sources officielles – ou, à défaut, des estimations transparentes basées sur des données publiques récentes – on obtient un panorama cohérent des écarts de consommation électrique à l’échelle des ménages dans les sept pays demandés.

3. Consommation des foyers par pays (calculé)

Pour conserver nos **cinq archétypes** (studio, couple, couple senior, famille, grande famille) **tout en collant à la consommation réelle mesurée dans chaque pays**, j'ai appliqué la démarche suivante :

- **Répartition démographique** — on utilise les parts officielles de ménages : 1 pers., 2 pers., 3 pers., ≥ 4 pers. (plus « couple senior » pour la France).
- **Moyenne "modèle"** — on calcule la consommation moyenne par foyer avec les valeurs-type (2 352 kWh pour un studio, 7 355 kWh pour une grande famille, etc.).
- **Facteur d'ajustement** — on compare cette moyenne au **kWh/foyer officiel** (ou la meilleure estimation publique) ; le rapport « officiel \div modèle » sert de coefficient multiplicateur pour **tous les archétypes** du pays.
- **Tableaux** — on reproduit, pour chaque pays, le pourcentage de ménages, le **kWh/foyer ajusté** (après coefficient) et le **kWh pondéré**. Le total correspond exactement au chiffre national.

France

Échelle de foyer	Part des ménages (%)	kWh/foyer ajusté	kWh pondéré
Studio	37.0	2 538	939.1
Couple	11.0	4 664	513.0
Couple senior	22.0	4 642	1 021.2
Famille	14.0	6 687	936.2
Grande famille	16.0	7 935	1 269.6
Moyenne (kWh/foyer)			4 679

Allemagne

Échelle de foyer	Part des ménages (%)	kWh/foyer ajusté	kWh pondéré
Studio	40.9	1 875	766.9
Couple	30.0	3 447	1 034.1
Famille	13.5	4 942	667.2
Grande famille	15.6	5 864	914.8
Moyenne (kWh/foyer)			3 383

Inde

Échelle de foyer	Part des ménages (%)	kWh/foyer ajusté	kWh pondéré
Studio	11.0	2 027	223.0
Couple	17.0	3 725	633.2
Famille	22.0	5 341	1 175.0
Grande famille	50.0	6 338	3 169.0
Moyenne (kWh/foyer)			5 200

Brésil

Échelle de foyer	Part des ménages (%)	kWh/foyer ajusté	kWh pondéré
Studio	15.0	1 085	162.8
Couple	25.0	1 994	498.5
Famille	20.0	2 859	571.8
Grande famille	40.0	3 393	1 357.2
Moyenne (kWh/foyer)			2 590

États-Unis

Échelle de foyer	Part des ménages (%)	kWh/foyer ajusté	kWh pondéré
Studio	28.0	5 344	1 496.3
Couple	34.0	9 822	3 339.5
Famille	15.0	14 081	2 112.2
Grande famille	23.0	16 710	3 843.3
Moyenne (kWh/foyer)			10 791

Norvege

Échelle de foyer	Part des ménages (%)	kWh/foyer ajusté	kWh pondéré
Studio	42.0	8 842	3 713.6
Couple	30.0	16 251	4 875.3
Famille	14.0	23 300	3 262.0
Grande famille	14.0	27 650	3 871.0
Moyenne (kWh/foyer)			15 722

Pologne

Échelle de foyer	Part des ménages (%)	kWh/foyer ajusté	kWh pondéré
Studio	24.0	862	206.9
Couple	29.0	1 584	459.4
Famille	22.0	2 272	499.8
Grande famille	25.0	2 696	674.0
Moyenne (kWh/foyer)			1 840

Conclusion

En conservant nos cinq profils-type mais en appliquant un coefficient d'ajustement propre à chaque pays, on obtient des tableaux qui :

- Respectent la distribution réelle des tailles de ménages (sources statistiques nationales).
- Reproduisent exactement la consommation moyenne officielle (ou l'estimation publique la plus récente) par foyer.

Cette approche garde la lisibilité pédagogique des archétypes tout en les ancrant dans la réalité de chaque pays ; elle peut servir de base à des scénarios "et si" (ex. : conversion des radiateurs polonais au tout-électrique, généralisation des pompes à chaleur en France, etc.) en modifiant simplement le coefficient d'ajustement ou la répartition démographique.

VI. Impact et mise en relation des foyers et du cloud

Dans le contexte de l'urgence climatique et de la transition énergétique, il devient crucial de mieux comprendre l'impact environnemental de nos usages électriques. Ce rapport vise à comparer l'empreinte carbone des foyers domestiques avec celle des infrastructures informatiques à grande échelle, telles que le cloud computing et les datacenters. L'analyse repose sur des modélisations de consommation énergétique croisées avec l'intensité carbone de la production électrique selon les pays.

1. Comparaison directe des impacts : foyers, cloud et datacenters

Un foyer minimaliste avec peu d'appareils électroniques consomme environ 1 000 kWh par an. En France, cela représente environ 60 kg de CO₂ émis. Le même foyer en Pologne, où l'électricité est produite principalement à partir du charbon, peut atteindre 700 kg de CO₂ à consommation égale.

Un foyer étudiant avec des usages modérés (PC, box, TV) consomme entre 2 000 et 2 500 kWh par an, soit de 120 à 150 kgCO₂ en France, et jusqu'à 1 750 kgCO₂ en Pologne.

Un foyer urbain moyen consomme autour de 3 500 kWh/an. Cela génère 210 kgCO₂ en France contre 2 450 kgCO₂ en Pologne. Le même foyer en Allemagne produit environ 1 400 kgCO₂/an.

Les foyers tout-électrique peuvent consommer jusqu'à 15 000 kWh/an, ce qui correspond à 900 kgCO₂ en France, 6 000 kgCO₂ en Allemagne, et jusqu'à 10 500 kgCO₂ en Pologne.

De leur côté, les services cloud personnels comme le stockage et les e-mails représentent une consommation estimée de 200 kWh/an par utilisateur. Cela correspond à environ 12 kgCO₂ en France, 80 kgCO₂ en Allemagne, et 140 kgCO₂ en Pologne.

Un usage intensif de streaming vidéo (3 heures par jour) représente une consommation d'environ 300 kWh par an. Cela entraîne 18 kgCO₂ en France, 120 kgCO₂ en Allemagne et 210 kgCO₂ en Pologne.

Une seule instance de serveur dans un datacenter peut consommer jusqu'à 8 000 kWh/an. En France, cela représente 480 kgCO₂. En Allemagne, on atteint 3 200 kgCO₂, et en Pologne jusqu'à 5 600 kgCO₂.

2. Interprétation générale

Un datacenter hébergé en Pologne peut émettre autant de CO₂ qu'une trentaine de foyers français classiques. De même, le simple fait d'utiliser des services numériques dans un pays à forte intensité carbone peut faire doubler ou tripler son empreinte écologique.

Chaque geste numérique a un coût invisible : un mail avec pièce jointe = 10 gCO₂, une heure de visio ou de streaming en 5G = jusqu'à 400 gCO₂ selon le pays. Ainsi, un foyer numérique intensif dans un pays fortement carboné peut émettre autant qu'une voiture effectuant plusieurs centaines de kilomètres par an.

Conclusion :

Ces comparaisons soulignent que l'origine de l'électricité est aussi importante que la quantité consommée. La réduction de l'empreinte carbone passe autant par la sobriété numérique que par un choix de localisation énergétiquement propre des infrastructures. Cette analyse peut être intégrée à toute réflexion plus globale sur les politiques énergétiques et les usages individuels.

Conclusion

Ce projet, mené dans le cadre de la SAÉ 2.06 « Analyse de données, reporting et datavisualisation », nous a permis d'approfondir une problématique environnementale majeure : celle de l'empreinte carbone de l'électricité et de ses usages numériques. En combinant une analyse rigoureuse des données de production électrique dans sept pays avec une modélisation concrète des consommations domestiques et numériques, nous avons pu mettre en lumière les disparités énergétiques mondiales, les dynamiques de transition et les zones de tension potentielles.

L'analyse des données horaires issues d'Electricity Maps a révélé de fortes hétérogénéités dans les mix énergétiques nationaux. Des pays comme la Norvège ou la France bénéficient d'un mix historiquement décarboné, tandis que d'autres, comme la Pologne ou l'Inde, restent très dépendants des énergies fossiles, notamment du charbon. Ces différences ont des conséquences directes sur l'intensité carbone de l'électricité, et donc sur l'impact environnemental des usages électriques courants.

Notre étude a également démontré que la simple mesure de la consommation électrique ne suffit pas à évaluer l'impact écologique : à consommation identique, l'empreinte carbone peut être multipliée par dix selon le pays. Ce constat est particulièrement frappant lorsqu'on compare les usages numériques — en apparence dématérialisés — tels que le cloud computing, les services de streaming ou l'IA, dont les besoins en électricité sont croissants et massifs. L'essor des datacenters et des technologies d'intelligence artificielle redéfinit les équilibres énergétiques, en exerçant une pression inédite sur les ressources mondiales, aussi bien en électricité qu'en eau et en matières premières.

En modélisant différents types de foyers selon les pays, nous avons pu estimer de manière concrète les émissions associées à leur consommation électrique. Cela a permis de contextualiser l'impact du numérique au regard des usages domestiques et d'illustrer, de manière pédagogique, les ordres de grandeur impliqués.

Notre démarche a combiné des outils variés (Excel, R, Tableau), une méthodologie rigoureuse (nettoyage, harmonisation, modélisation) et une volonté constante de rendre les résultats accessibles et utiles pour la prise de décision. En croisant les indicateurs statistiques avec des considérations concrètes, nous avons mis en évidence les leviers les plus prometteurs pour une électricité plus propre et une consommation plus responsable.

Malgré les limites liées aux sources de données ou à certaines hypothèses de modélisation, nos résultats fournissent un socle solide pour de futures analyses. Ils soulignent la nécessité d'une transition énergétique accélérée, mais aussi d'une régulation plus transparente et exigeante du secteur numérique. L'optimisation des infrastructures, le développement de technologies sobres, l'éducation aux enjeux numériques et énergétiques sont autant de pistes qu'il convient d'explorer.

En conclusion, l'électricité n'est pas un bien neutre. Sa provenance, sa temporalité, son usage ont un poids réel sur notre avenir collectif. À travers cette étude, nous espérons avoir contribué à une meilleure compréhension de ces enjeux, et ouvert la voie à des politiques plus éclairées, intégrant pleinement la dimension carbone de nos choix technologiques et énergétiques.

Annexes

```

17 # 1.2. Création de full_data : stack de tous les data.frames
18 objets <- ls(pattern = "(DE|BR|FR|IN|NO|PL|US)_[0-9]{4}")
19 liste_dfs <- mget(objets)
20
21 full_data <- bind_rows(liste_dfs, .id = "source")
22
23 # Vérifie que la colonne Datetime..UTC. existe
24 if (!"Datetime..UTC." %in% names(full_data)) {
25   stop("La colonne 'Datetime..UTC.' est absente de full_data.")
26 }
27
28 # Sauvegarde de l'original (optionnel mais utile pour vérification)
29 full_data$Datetime_Original <- as.character(full_data$Datetime..UTC.)
30
31 # Nettoyage : supprimer les millisecondes si présentes (ex : "2022-01-01 00:00:00.000")
32 full_data$Datetime_Clean <- gsub("\\.\\d+$", "", full_data$Datetime_Original)
33
34 # Détection des formats :
35 # - Format ISO : "2021-01-01 13:00:00"
36 # - Format Européen : "01/01/2021 13:00"
37 format_iso <- grepl("(^\\d{4}-\\d{2}-\\d{2})\\s*", full_data$Datetime_Clean)
38 format_euro <- grepl("(^\\d{2}/\\d{2}/\\d{4})\\s*", full_data$Datetime_Clean)
39
40 # Initialiser la nouvelle colonne de datetime propre
41 full_data$Datetime.UTC <- NA
42

```

```

21 # 1.3. Installation des bibliothèques
22 # - ALTERNATIVE
23 DE_2022 <- read.table(file.path(data_dir, "DE_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
24 BR_2022 <- read.table(file.path(data_dir, "BR_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
25 IN_2022 <- read.table(file.path(data_dir, "IN_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
26 NO_2022 <- read.table(file.path(data_dir, "NO_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
27 PL_2022 <- read.table(file.path(data_dir, "PL_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
28 US_2022 <- read.table(file.path(data_dir, "US_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
29
30 # BRESIL
31 BR_2021 <- read.table(file.path(data_dir, "BR_2021_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
32 BR_2022 <- read.table(file.path(data_dir, "BR_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
33 BR_2023 <- read.table(file.path(data_dir, "BR_2023_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
34 BR_2024 <- read.table(file.path(data_dir, "BR_2024_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
35
36 # FRANCE
37 FR_2021 <- read.table(file.path(data_dir, "FR_2021_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
38 FR_2022 <- read.table(file.path(data_dir, "FR_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
39 FR_2023 <- read.table(file.path(data_dir, "FR_2023_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
40 FR_2024 <- read.table(file.path(data_dir, "FR_2024_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
41
42 # INDIA
43 IN_2021 <- read.table(file.path(data_dir, "IN_2021_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
44 IN_2022 <- read.table(file.path(data_dir, "IN_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
45 IN_2023 <- read.table(file.path(data_dir, "IN_2023_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
46 IN_2024 <- read.table(file.path(data_dir, "IN_2024_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
47

```

```

128 # 1.4. Vérification des données
129 full_data$Country <- case_when(
130   str_detect(full_data$source, "ADE") ~ "Germany",
131   str_detect(full_data$source, "ABR") ~ "Brazil",
132   str_detect(full_data$source, "AFR") ~ "France",
133   str_detect(full_data$source, "AIN") ~ "India",
134   str_detect(full_data$source, "ANO") ~ "Norway",
135   str_detect(full_data$source, "APL") ~ "Poland",
136   str_detect(full_data$source, "AUS") ~ "USA",
137   TRUE ~ "Autre"
138 )
139
140 # Pour les graphiques mensuels
141 full_data$YearMonth <- as.Date(floor_date(full_data$Datetime.UTC, "month"))
142
143 # Vérifie que la colonne Datetime..UTC. existe
144 if (!"Datetime..UTC." %in% names(full_data)) {
145   stop("La colonne 'Datetime..UTC.' est absente de full_data.")
146 }
147
148 # Sauvegarde de l'original (optionnel mais utile pour vérification)
149 full_data$Datetime_Original <- as.character(full_data$Datetime..UTC.)
150
151 # Nettoyage : supprimer les millisecondes si présentes
152 full_data$Datetime_Clean <- gsub("\\.\\d+$", "", full_data$Datetime_Original)
153

```

```

180 nb_na <- sum(is.na(full_data$Datetime.UTC))
181 cat("Nombre de valeurs manquantes dans Datetime.UTC :", nb_na, "\n")
182
183 cat("Plage de dates dans les données :\n")
184 print(range(full_data$Datetime.UTC, na.rm = TRUE))
185
186 # Extraire l'année depuis 'source' (ex : "DE_2022" → 2022)
187 full_data$Year <- as.integer(str_extract(full_data$source, "[0-9]{4}"))
188
189 # Extraire le pays depuis 'source' (ex : "DE_2022" → "Germany")
190 full_data$Country <- case_when(
191   str_detect(full_data$source, "ADE") ~ "Germany",
192   str_detect(full_data$source, "ABR") ~ "Brazil",
193   str_detect(full_data$source, "AFR") ~ "France",
194   str_detect(full_data$source, "AIN") ~ "India",
195   str_detect(full_data$source, "ANO") ~ "Norway",
196   str_detect(full_data$source, "APL") ~ "Poland",
197   str_detect(full_data$source, "AUS") ~ "USA",
198   TRUE ~ "Autre"
199 )
200
201 # Créer une variable 'YearMonth' utilisable pour les graphiques mensuels
202 full_data$YearMonth <- as.Date(floor_date(full_data$Datetime.UTC, "month"))
203

```

```

241 # 2.2. Part Low Carbon (nucléaire + renouvelables)
242 # Distribution globale
243 full_data %>%
244   filter(!is.na(Low.Carbon.Percentage)) %>%
245   ggplot(aes(x = Low.Carbon.Percentage)) +
246     geom_histogram(bins = 30, fill = "darkgreen", color = "white") +
247     labs(
248       title = "Distribution de la part Low Carbon (toutes heures)",
249       x = "% Low Carbon (nuke + renew.)",
250       y = "Nombre d'heures"
251     ) +
252     theme_minimal()
253
254 # Boxplot par pays
255 full_data %>%
256   filter(!is.na(Low.Carbon.Percentage)) %>%
257   ggplot(aes(x = Country, y = Low.Carbon.Percentage, fill = Country)) +
258     geom_boxplot() +
259     labs(
260       title = "Part Low Carbon par pays (2021-2024)",
261       x = "Pays",
262       y = "% Low Carbon"
263     ) +
264     theme_minimal() +
265     theme(axis.text.x = element_text(angle = 45, hjust = 1),
266           legend.position = "none")
267

```

```

128 # 1.3. Installation des bibliothèques
129 # - ALTERNATIVE
130 DE_2022 <- read.table(file.path(data_dir, "DE_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
131 BR_2022 <- read.table(file.path(data_dir, "BR_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
132 IN_2022 <- read.table(file.path(data_dir, "IN_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
133 NO_2022 <- read.table(file.path(data_dir, "NO_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
134 PL_2022 <- read.table(file.path(data_dir, "PL_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
135 US_2022 <- read.table(file.path(data_dir, "US_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
136
137 # BRESIL
138 BR_2021 <- read.table(file.path(data_dir, "BR_2021_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
139 BR_2022 <- read.table(file.path(data_dir, "BR_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
140 BR_2023 <- read.table(file.path(data_dir, "BR_2023_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
141 BR_2024 <- read.table(file.path(data_dir, "BR_2024_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
142
143 # FRANCE
144 FR_2021 <- read.table(file.path(data_dir, "FR_2021_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
145 FR_2022 <- read.table(file.path(data_dir, "FR_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
146 FR_2023 <- read.table(file.path(data_dir, "FR_2023_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
147 FR_2024 <- read.table(file.path(data_dir, "FR_2024_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
148
149 # INDIA
150 IN_2021 <- read.table(file.path(data_dir, "IN_2021_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
151 IN_2022 <- read.table(file.path(data_dir, "IN_2022_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
152 IN_2023 <- read.table(file.path(data_dir, "IN_2023_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
153 IN_2024 <- read.table(file.path(data_dir, "IN_2024_hourly.csv"), header = TRUE, sep = ";", dec = ".", encoding = "UTF-8")
154

```

```

103 # Conversion format ISO
104 full_data$Datetime.UTC[format_iso] <- as.character(
105   as.POSIXct(full_data$Datetime_Clean[format_iso], format = "%Y-%m-%d %H:%M:%S", tz = "UTC")
106 )
107
108 # Conversion format Européen
109 full_data$Datetime.UTC[format_euro] <- as.character(
110   as.POSIXct(full_data$Datetime_Clean[format_euro], format = "%d/%m/%Y %H:%M", tz = "UTC")
111 )
112
113 # Conversion finale en POSIXct (timezone UTC)
114 full_data$Datetime.UTC <- as.POSIXct(full_data$Datetime.UTC, tz = "UTC")
115
116 # Vérifications utiles
117 cat("Résumé des dates (Datetime.UTC) :\n")
118 print(summary(full_data$Datetime.UTC))
119
120 nb_na <- sum(is.na(full_data$Datetime.UTC))
121 cat("Nombre de valeurs manquantes dans Datetime.UTC :", nb_na, "\n")
122
123 cat("Plage de dates dans les données :\n")
124 print(range(full_data$Datetime.UTC, na.rm = TRUE))
125
126 # Extraction des composants utiles
127 full_data$Year <- as.integer(str_extract(full_data$source, "[0-9]{4}"))
128

```

```

154 # Détection des formats :
155 # - Format ISO : "2021-01-01 13:00:00"
156 # - Format Européen : "01/01/2021 13:00"
157 format_iso <- grepl("(^\\d{4}-\\d{2}-\\d{2})\\s*", full_data$Datetime_Clean)
158 format_euro <- grepl("(^\\d{2}/\\d{2}/\\d{4})\\s*", full_data$Datetime_Clean)
159
160 # Initialiser la nouvelle colonne
161 full_data$Datetime.UTC <- NA
162
163 # Conversion format ISO
164 full_data$Datetime.UTC[format_iso] <- as.character(
165   as.POSIXct(full_data$Datetime_Clean[format_iso], format = "%Y-%m-%d %H:%M:%S", tz = "UTC")
166 )
167
168 # Conversion format Européen
169 full_data$Datetime.UTC[format_euro] <- as.character(
170   as.POSIXct(full_data$Datetime_Clean[format_euro], format = "%d/%m/%Y %H:%M", tz = "UTC")
171 )
172
173 # Conversion finale de toute la colonne en POSIXct
174 full_data$Datetime.UTC <- as.POSIXct(full_data$Datetime.UTC, tz = "UTC")
175
176 # Vérifications
177 cat("Résumé des dates (Datetime.UTC) :\n")
178 print(summary(full_data$Datetime.UTC))
179

```

```

209 # 2. Exploration univariée
210 # 2.1. Intensité carbone directe
211 # Distribution globale
212 full_data %>%
213   filter(!is.na(Carbon.Intensity.gco.eq.kwh.direct)) %>%
214   ggplot(aes(Carbon.Intensity.gco.eq.kwh.direct)) +
215     geom_histogram(bins = 30, fill = "darkgreen", color = "white") +
216     labs(
217       title = "Distribution de l'intensité carbone directe (toutes heures)",
218       x = "Intensité carbone directe (gCO2/kwh)",
219       y = "Nombre d'heures"
220     ) +
221     theme_minimal()
222
223 # Boxplot par pays
224 full_data %>%
225   filter(!is.na(Carbon.Intensity.gco.eq.kwh.direct)) %>%
226   ggplot(aes(x = Country, y = Carbon.Intensity.gco.eq.kwh.direct, fill = Country)) +
227     geom_boxplot() +
228     labs(
229       title = "Intensité carbone directe par pays (2021-2024)",
230       x = "Pays",
231       y = "Intensité carbone directe (gCO2/kwh)"
232     ) +
233     theme_minimal() +
234     theme(axis.text.x = element_text(angle = 45, hjust = 1),
235           legend.position = "none")
236

```



```
303- # 3. Tendances annuelles
304- ```{r}
305- # 3.1. Intensité carbone directe annuelle
306- df_int_ann <- full_data %>%
307-   filter(!is.na(Carbon.Intensity.gCO.eq.kwh..direct.)) %>%
308-   mutate(Year = as.integer(str_extract(source, "[0-9]{4}"))) %>%
309-   group_by(Country, Year) %>%
310-   summarise(
311-     Moy_Int_Directe = mean(Carbon.Intensity.gCO.eq.kwh..direct., na.rm = TRUE),
312-     .groups = "drop"
313-   )
314-
315- ggplot(df_int_ann, aes(x = Year, y = Moy_Int_Directe, color = Country)) +
316-   geom_line(linewidth = 1) + # Remplacé size par linewidth
317-   geom_point(size = 2) +
318-   labs(
319-     title = "Evolution annuelle de l'intensité carbone directe",
320-     x = "Année",
321-     y = "Intensité carbone directe moyenne (gCO2/kwh)"
322-   ) +
323-   theme_minimal()
324-
325- # 3.2. Part Renewable annuelle
326- df_ren_ann <- full_data %>%
327-   filter(!is.na(Renewable.Percentage)) %>%
328-   mutate(Year = as.integer(str_extract(source, "[0-9]{4}"))) %>%
329-   group_by(Country, Year) %>%
330-   summarise(
331-     Moy_Renew = mean(Renewable.Percentage, na.rm = TRUE),
332-     .groups = "drop"
333-   )
334-
335- ggplot(df_ren_ann, aes(x = Year, y = Moy_Renew, color = Country)) +
336-   geom_line(linewidth = 1) + # idem ici
337-   geom_point(size = 2) +
338-   labs(
339-     title = "Evolution annuelle de la part Renewable",
340-     x = "Année",
341-     y = "Part Renewable moyenne"
342-   ) +
343-   theme_minimal()
344-
345- 
```

```
348- # 4. Analyse saisonnière (moyenne mensuelle)
349- ## 4.1. Intensité carbone directe (mensuelle)
350- ```{r}
351- df_int_mens <- full_data %>%
352-   filter(!is.na(Carbon.Intensity.gCO.eq.kwh..direct.),
353-          !is.na(Datetime.UTC)) %>%
354-   mutate(YearMonth = as.Date(Floor_date(Datetime.UTC, "month"))) %>%
355-   group_by(Country, YearMonth) %>%
356-   summarise(
357-     Moy_Mens_Int = mean(Carbon.Intensity.gCO.eq.kwh..direct., na.rm = TRUE),
358-     .groups = "drop"
359-   )
360-
361- ggplot(df_int_mens, aes(x = YearMonth, y = Moy_Mens_Int, color = Country)) +
362-   geom_line(size = 1) +
363-   geom_point(size = 1) +
364-   scale_x_date(date_labels = "%Y-%m", date_breaks = "3 months") +
365-   labs(
366-     title = "Saisonnalité mensuelle de l'intensité carbone directe (2021-2024)",
367-     x = "Mois",
368-     y = "Intensité carbone moyenne (gCO2/kwh)"
369-   ) +
370-   theme_minimal() +
371-   theme(axis.text.x = element_text(angle = 45, hjust = 1))
372-
373- 
```

```
404- # 4.3. Renewable Percentage (mensuelle)
405- ```{r}
406- # 4.2. Part Low Carbon (mensuelle)
407- df_low_mens <- full_data %>%
408-   filter(!is.na(Low.Carbon.Percentage),
409-          !is.na(Datetime.UTC)) %>%
410-   mutate(YearMonth = as.Date(Floor_date(Datetime.UTC, "month"))) %>%
411-   group_by(Country, YearMonth) %>%
412-   summarise(
413-     Moy_Mens_Low = mean(Low.Carbon.Percentage, na.rm = TRUE),
414-     .groups = "drop"
415-   )
416-   filter(!is.na(Moy_Mens_Low)) # Enlève les moyennes manquantes après agrégation
417-
418- ggplot(df_low_mens, aes(x = YearMonth, y = Moy_Mens_Low, color = Country)) +
419-   geom_line(linewidth = 1) + # Remplace size par linewidth
420-   geom_point(size = 1) +
421-   scale_x_date(date_labels = "%Y-%m", date_breaks = "3 months") +
422-   labs(
423-     title = "Saisonnalité mensuelle de la part Low Carbon (2021-2024)",
424-     x = "Mois",
425-     y = "Part Low Carbon moyen"
426-   ) +
427-   theme_minimal() +
428-   theme(axis.text.x = element_text(angle = 45, hjust = 1))
429-
430- 
```

```
458- # Intervalles de confiance des coefficients
459- cat("Intervalles de confiance (95%) des coefficients :\n")
460- print(confint(mod))
461-
462- # Graphe 1 : Résidus vs valeurs prédites
463- df_pred <- data.frame(
464-   pred = predict(mod),
465-   resid = residuals(mod)
466- )
467- p1 <- ggplot(df_pred, aes(x = pred, y = resid)) +
468-   geom_point(alpha = 0.3) +
469-   geom_hline(yintercept = 0, linetype = "dashed") +
470-   labs(
471-     title = "Résidus vs Valeurs prédites",
472-     x = "Valeurs prédites",
473-     y = "Résidus"
474-   ) +
475-   theme_minimal()
476- print(p1)
477-
478- # Graphe 2 : Régression avec IC95%
479- p2 <- ggplot(df_xy, aes_string(x = xvar, y = yvar)) +
480-   geom_point(alpha = 0.3) +
481-   geom_smooth(method = "lm", level = 0.95, color = "blue") +
482-   labs(
483-     title = "Régression avec IC95%",
484-     yvar, "en fonction de", xvar
485-   ) +
486-   theme_minimal()
487- print(p2)
488- 
```

```
272- # 2.3. Part Renewable (renouvelables seules)
273- ```{r}
274- # Distribution globale
275- full_data %>%
276-   filter(!is.na(Renewable.Percentage)) %>%
277-   ggplot(aes(x = Renewable.Percentage)) +
278-     geom_histogram(bins = 30, fill = "steelblue", color = "white") +
279-     labs(
280-       title = "Distribution de la part Renewable (toutes heures)",
281-       x = "% Renouvelable",
282-       y = "Nombre d'heures"
283-     ) +
284-     theme_minimal()
285-
286- # Boxplot par pays
287- full_data %>%
288-   filter(!is.na(Renewable.Percentage)) %>%
289-   ggplot(aes(x = Country, y = Renewable.Percentage, fill = Country)) +
290-     geom_boxplot() +
291-     labs(
292-       title = "Part Renewable par pays (2021-2024)",
293-       x = "Pays",
294-       y = "% Renouvelable"
295-     ) +
296-     theme_minimal() +
297-     theme(axis.text.x = element_text(angle = 45, hjust = 1),
298-           legend.position = "none")
299-
300- 
```

```
325- # 3.2. Part Renewable annuelle
326- df_ren_ann <- full_data %>%
327-   filter(!is.na(Renewable.Percentage)) %>%
328-   mutate(Year = as.integer(str_extract(source, "[0-9]{4}"))) %>%
329-   group_by(Country, Year) %>%
330-   summarise(
331-     Moy_Renew = mean(Renewable.Percentage, na.rm = TRUE),
332-     .groups = "drop"
333-   )
334-
335- ggplot(df_ren_ann, aes(x = Year, y = Moy_Renew, color = Country)) +
336-   geom_line(linewidth = 1) + # idem ici
337-   geom_point(size = 2) +
338-   labs(
339-     title = "Evolution annuelle de la part Renewable",
340-     x = "Année",
341-     y = "Part Renewable moyenne"
342-   ) +
343-   theme_minimal()
344-
345- 
```

```
377- # 4.2. Low Carbon Percentage (mensuelle)
378- ```{r}
379- df_low_mens <- full_data %>%
380-   filter(!is.na(Low.Carbon.Percentage), !is.na(Datetime.UTC)) %>%
381-   mutate(YearMonth = as.Date(Floor_date(Datetime.UTC, "month"))) %>%
382-   group_by(Country, YearMonth) %>%
383-   summarise(
384-     Moy_Mens_Low = mean(Low.Carbon.Percentage, na.rm = TRUE),
385-     .groups = "drop"
386-   )
387-   filter(!is.na(Moy_Mens_Low)) # double filtre
388-
389- ggplot(df_low_mens, aes(x = YearMonth, y = Moy_Mens_Low, color = Country)) +
390-   geom_line(size = 1) +
391-   geom_point(size = 1) +
392-   scale_x_date(date_labels = "%Y-%m", date_breaks = "3 months") +
393-   labs(
394-     title = "Saisonnalité mensuelle de la part Low Carbon (2021-2024)",
395-     x = "Mois",
396-     y = "Part Low Carbon moyen"
397-   ) +
398-   theme_minimal() +
399-   theme(axis.text.x = element_text(angle = 45, hjust = 1))
400-
401- 
```

```
432- # 5. Corrélations et comparaisons croisées
433- ## 5.1. Matrice de corrélation entre indicateurs carbone
434- ```{r corr-scatter-reg, message=FALSE, warning=FALSE}
435- library(broom)
436- library(ggplot2)
437- library(dplyr)
438-
439- # Fonction d'analyse complète (sans test Shapiro-Wilk)
440- analyser_modele <- function(df, xvar, yvar) {
441-   # Nettoyage des données
442-   df_xy <- df %>%
443-     filter(!is.na(.data[[xvar]]), !is.na(.data[[yvar]]))
444-
445-   # Régression linéaire
446-   formule <- as.formula(paste0("~", yvar, "~", xvar, ""))
447-   mod <- lm(formule, data = df_xy)
448-
449-   # R² et EQM
450-   r2 <- summary(mod)$r.squared
451-   eqm <- mean(mod$residuals^2)
452-
453-   cat(paste0("Modèle : ", xvar, " vs ", yvar, "\n"))
454-   cat(paste0("R² = ", round(r2, 3), "\n"))
455-   cat(paste0("EQM = ", round(eqm, 3), "\n\n"))
456- }
457- 
```

```
# Appliquer sur les deux couples très corrélés
analyser_modele(full_data,
  "Carbon.Intensity.gCO.eq.kwh..direct.",
  "Carbon.Intensity.gCO.eq.kwh..LCA.")

analyser_modele(full_data,
  "Carbon.Intensity.gCO.eq.kwh..direct.",
  "Low.Carbon.Percentage")

```



```
505- # 3.2. Scatter plots
506- {r}
507- # (a) Renewable % vs Intensité carbone directe
508- full_data %>%
509-   filter(
510-     !is.na(Renewable.Percentage),
511-     !is.na(Carbon.Intensity.gCO.eq.kwh..direct.)
512-   ) %>%
513-   ggplot(aes(
514-     x = Renewable.Percentage,
515-     y = Carbon.Intensity.gCO.eq.kwh..direct.,
516-     color = Country
517-   )) +
518-   geom_point(alpha = 0.3) +
519-   labs(
520-     title = "Renewable % vs Intensité carbone directe",
521-     x = "% Renouvelable",
522-     y = "Intensité carbone directe (gCO2/kwh)"
523-   ) +
524-   theme_minimal()
525-
526- # (b) Low Carbon % vs Renewable %
527- full_data %>%
528-   filter(
529-     !is.na(Low.Carbon.Percentage),
530-     !is.na(Renewable.Percentage)
531-   ) %>%
532-   ggplot(aes(
533-     x = Low.Carbon.Percentage,
534-     y = Renewable.Percentage,
535-     color = Country
536-   )) +
537-   geom_point(alpha = 0.3) +
538-   labs(
539-     title = "Low Carbon % vs Renewable %",
540-     x = "% Low Carbon (Nucléaire + Renouvelable)",
541-     y = "% Renouvelable"
```

```
548- # 3.3. Ecart mensuel = LCA - Direct =
549- {r}
550- library(dplyr)
551- library(ggplot2)
552- library(lubridate)
553-
554- df_ecart <- full_data %>%
555-   filter(
556-     !is.na(Carbon.Intensity.gCO.eq.kwh..direct.),
557-     !is.na(Carbon.Intensity.gCO.eq.kwh..LCA.)
558-   ) %>%
559-   mutate(
560-     YearMonth = as.Date(floor_date(Datetime.UTC, "month")),
561-     Diff_LCA_Direct = Carbon.Intensity.gCO.eq.kwh..LCA. -
562-       Carbon.Intensity.gCO.eq.kwh..direct.
563-   ) %>%
564-   group_by(Country, YearMonth) %>%
565-   summarise(
566-     Moy_Ecart = mean(Diff_LCA_Direct, na.rm = TRUE),
567-     groups = "drop"
568-   ) %>%
569-   filter(
570-     !is.na(YearMonth), # on retire les mois NA
571-     !is.na(Moy_Ecart) # on retire les écarts NA
572-   )
573-
574- ggplot(df_ecart, aes(x = YearMonth, y = Moy_Ecart, color = Country)) +
575-   geom_line(size = 1) +
576-   geom_point(size = 1) +
577-   scale_x_date(date_labels = "%Y-%m", date_breaks = "3 months") +
578-   labs(
579-     title = "Ecart moyen LCA - Direct (gCO2/kwh) par pays et par mois",
580-     x = "Mois",
581-     y = "Ecart LCA - Direct"
582-   ) +
583-   theme_minimal() +
584-   theme(axis.text.x = element_text(angle = 45, hjust = 1))
585-
```

```
{r}
ggplot(df_int_mens, aes(x = YearMonth, y = Moy_Mens_Int)) +
  geom_line(color = "steelblue") +
  facet_wrap(~ Country, scales = "free_y") +
  labs(title = "Evolution mensuelle par pays", y = "gCO2/kwh") +
  theme_minimal()
...
```

```
{r}
# Sélection des colonnes pertinentes
cor_data <- full_data %>%
  select(
    Carbon.Intensity.gCO.eq.kwh..direct.,
    Carbon.Intensity.gCO.eq.kwh..LCA.,
    Low.Carbon.Percentage,
    Renewable.Percentage
  ) %>%
  filter(complete.cases())

# Calcul de la matrice de corrélation
cor_matrix <- round(cor(cor_data), 3)
print("Matrice de corrélation entre indicateurs :")
cor_matrix
```

```
518- {r}
519- # Comparaison pour: Low.Carbon.Percentage ~ Carbon.Intensity.gCO.eq.kwh..direct.
520-
521- df_xy <- full_data %>%
522-   filter(
523-     !is.na(Carbon.Intensity.gCO.eq.kwh..direct.),
524-     !is.na(Low.Carbon.Percentage)
525-   )
526-
527- # Modèle linéaire
528- mod_lm <- lm(Low.Carbon.Percentage ~ Carbon.Intensity.gCO.eq.kwh..direct., data = df_xy)
529- r2_lm <- summary(mod_lm)$r.squared
530-
531- # Modèle quadratique
532- mod_quad <- lm(Low.Carbon.Percentage ~ poly(Carbon.Intensity.gCO.eq.kwh..direct., 2), data = df_xy)
533- r2_quad <- summary(mod_quad)$r.squared
534-
535- # Affichage côte à côte
536- p_lm <- ggplot(df_xy, aes(x = Carbon.Intensity.gCO.eq.kwh..direct., y = Low.Carbon.Percentage)) +
537-   geom_point(alpha = 0.3) +
538-   geom_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "blue") + # R [5]
539-   labs(title = "pas de régression quadratique (R2 = ", round(r2_lm, 3), ")")
540-   x = "Intensité carbone directe", y = "% Low Carbon") +
541-   theme_minimal()
542-
543- p_quad <- ggplot(df_xy, aes(x = Carbon.Intensity.gCO.eq.kwh..direct., y = Low.Carbon.Percentage)) +
544-   geom_point(alpha = 0.3) +
545-   geom_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "blue") + # R [5]
546-   labs(title = "pas de régression quadratique (R2 = ", round(r2_quad, 3), ")")
547-   x = "Intensité carbone directe", y = "% Low Carbon") +
548-   theme_minimal()
549-
550- library(ggExtra)
551- grid.arrange(p_lm, p_quad, ncol = 2)
552-
553-
554-
```

```
650- # 2. Regarder les pays atypiques (résidus extrêmes)
651- {r}
652- # Sélection du modèle Low Carbon ~ Direct
653- df_xy <- full_data %>%
654-   filter(!is.na(Carbon.Intensity.gCO.eq.kwh..direct.), !is.na(Low.Carbon.Percentage))
655-
656- mod <- lm(Low.Carbon.Percentage ~ Carbon.Intensity.gCO.eq.kwh..direct., data = df_xy)
657-
658- df_res <- df_xy %>%
659-   mutate(resid = residuals(mod)) %>%
660-   group_by(Country) %>%
661-   summarise(
662-     Resid_Moy = mean(resid),
663-     Resid_Abs_Moy = mean(abs(resid)),
664-     groups = "drop"
665-   ) %>%
666-   mutate(
667-     Country = case_when(
668-       Country == "USA" ~ "Le modèle surestime fortement la part Low Carbon des USA, erreur moyenne la plus élevée.",
669-       Country == "Poland" ~ "Le modèle sous-estime la part Low Carbon en Pologne. Les valeurs observées sont plus élevées.",
670-       Country == "India" ~ "Le modèle surestime la part Low Carbon en Inde, dans une moindre mesure que pour les USA.",
671-       Country == "Germany" ~ "Le modèle sous-estime légèrement la part Low Carbon en Allemagne. Certains points mal ajustés.",
672-       Country == "France" ~ "Bon ajustement en France. Faibles biais et erreurs modérées.",
673-       Country == "Brazil" ~ "Très bon ajustement au Brésil. Prédiction proche des observations.",
674-       TRUE ~ "Pays non interprété automatiquement."
675-     )
676-   ) %>%
677-   arrange(desc(abs(Resid_Moy)))
678-
679- print(df_res)
680- knitr::kable(df_res, digits = 2, caption = "Analyse des résidus par pays (modèle Low Carbon ~ Direct)")
681-
682-
683-
```

```
694- # 3. (Facultatif) Carte des pays avec indicateur carbone moyen
695- {r}
696- library(ggplot2)
697- library(naturalearth)
698- library(sf)
699-
700- # Moyenne par pays
701- carbone_moy <- full_data %>%
702-   group_by(Country) %>%
703-   summarise(Moy_Carbone = mean(Carbon.Intensity.gCO.eq.kwh..direct., na.rm = TRUE))
704-
705- # Préparer la carte
706- world <- ne_countries(scale = "medium", returnclass = "sf")
707- carte_df <- left_join(world, carbone_moy, by = c("name" = "country"))
708-
709- ggplot(carte_df) +
710-   geom_sf(aes(fill = Moy_Carbone)) +
711-   scale_fill_viridis(option = "C") +
712-   labs(title = "Carte : Intensité carbone moyenne (2021-2024)" +
713-     theme_minimal()
714-
715-
716-
```

```
718- # 4. Regarder les extrêmes horaires
719- {r}
720- full_data %>%
721-   filter(!is.na(Carbon.Intensity.gCO.eq.kwh..direct.)) %>%
722-   arrange(desc(Carbon.Intensity.gCO.eq.kwh..direct.)) %>%
723-   select(Datetime.UTC, Country, Carbon.Intensity.gCO.eq.kwh..direct.) %>%
724-   head(10)
725-
726- full_data %>%
727-   filter(!is.na(Carbon.Intensity.gCO.eq.kwh..direct.)) %>%
728-   arrange(Carbon.Intensity.gCO.eq.kwh..direct.) %>%
729-   select(Datetime.UTC, Country, Carbon.Intensity.gCO.eq.kwh..direct.) %>%
730-   head(10)
731-
732-
733-
```

```
737- # 6. Différence LCA - Direct par pays
738- {r}
739- df_ecart_moy <- full_data %>%
740-   filter(!is.na(Carbon.Intensity.gCO.eq.kwh..direct.),
741-     !is.na(Carbon.Intensity.gCO.eq.kwh..LCA.)) %>%
742-   mutate(Diff = Carbon.Intensity.gCO.eq.kwh..LCA. - Carbon.Intensity.gCO.eq.kwh..direct.) %>%
743-   group_by(Country) %>%
744-   summarise(Ecart_Moy = mean(Diff, na.rm = TRUE), groups = "drop")
745-
746- ggplot(df_ecart_moy, aes(x = reorder(Country, Ecart_Moy, fill = Country)) +
747-   geom_col() +
748-   coord_flip() +
749-   labs(title = "Ecart moyen LCA - Direct par pays", x = "Pays", y = "Ecart (gCO2/kwh)") +
750-   theme_minimal() +
751-   theme(legend.position = "none")
752-
753-
754-
```

```
755- # 7. Histogrammes des résidus (modèle Low Carbon ~ Direct)
756- {r}
757- df_residues <- data.frame(resid = residuals(mod))
758-
759- ggplot(df_residues, aes(x = residues)) +
760-   geom_histogram(bins = 40, fill = "darkred", color = "black") +
761-   labs(title = "Distribution des résidus (Low Carbon ~ Direct)", x = "Résidus", y = "Fréquence") +
762-   theme_minimal()
763-
764-
765-
```