

HEC Lausanne – été 1999

Cas de « Statistique : Analyse de données »



• • • • • • • • • •

La pollution mondiale

Table des matières

Le choix du sujet – source des données	3
Méthodes utilisées	3
Introduction à la régression linéaire	4
Résultats de la régression linéaire	6
Conclusion de la régression linéaire	8
Objectifs de l'analyse en composantes principales	9
La technique de l'Analyse en Composantes Principales	10
Processus de calcul	10
Conclusion de l'analyse en composantes principales	15
Introduction à la classification	16
Les classifications effectuées	17
Classification en 5 classes (clusters)	17
Classification en 10 classes (clusters)	19
Commentaires sur la classification	20
Conclusion sur le sujet choisi	21
Conclusions sur le travail effectué	21
Les méthodes utilisées	22
Régression Linéaire	22
Analyse factorielle	22
Classification	22
Annexe 1 - Selected Energy Statistics for 1996	23
Annexe 2 - Selected Energy Indicators for 1996	26

Introduction

Présentation générale du choix du sujet, ainsi que de la source des données et les méthodes statistiques utilisées

Le choix du sujet – source des données

Le but de notre cas est de tenter d'expliquer les émissions de gaz carbonique (CO₂) dans le monde entier. Nous avons choisi ce thème car il s'agit d'un sujet d'actualité (tout comme nos données) et qu'il est au centre des préoccupations des gouvernements. Nous avons puisé nos données dans le « KEY WORLD ENERGY STATISTICS » de l'Agence Internationale de l'Énergie (IEA) disponible à l'adresse http://www.iea.org/stats/files/keystats/stats_98.htm afin d'obtenir des valeurs actuelles. Nous avons utilisé le logiciel SPSS 8.0 (et accessoirement Excel) pour toute la partie « calculs » lors de nos analyses car il se prête idéalement à cet exercice.

Méthodes utilisées

Pour analyser les données brutes, nous avons tout d'abord procédé à une régression simple permettant d'avoir une idée générale de la situation. Cette méthode nous a rapidement mise en garde contre le fait qu'une donnée qui possède un poids important peut influencer énormément notre droite de régression.

Ensuite, nous avons procédé à une analyse en composantes principales pour expliquer au mieux les émissions de CO₂. Comme nous disposons d'une grande quantité de données quantitatives, l'analyse factorielle nous a permis de simplifier le tableau principal en regroupant les variables. Le choix des composantes principales s'est effectué principalement sur le critère de la valeur propre (on retient les composantes dont la variance est supérieure à un car elles apportent plus d'informations que les variables de départ). Cette analyse devrait également confirmer la segmentation effectuée avec la méthode précédente.

Nous avons finalement entrepris une classification car nous possédons un nombre extrêmement élevé de pays. Le but de cette démarche était d'essayer de construire des groupes de pays le plus hétérogène entre les groupes (between) et le plus homogène à l'intérieur (within) de ceux-ci. Il est vrai que nous avions auparavant déjà une idée sur l'existence de différents groupes et nous avons voulu la démontrer. Une des principales difficultés a été de déterminer le nombre de nuages que nous voulions obtenir. Comme la méthode du « K-mean cluster » est très volatile (les différences suivant le nombre de nuages choisis sont importantes), nous avons effectués de multiples itérations nous permettant de rechercher le nombre de classes « optimal » permettant de découvrir une tendance et avons décidé de présenter la version à cinq groupes.

La régression linéaire

Nous allons dans cette partie nous intéresser aux deux facteurs principaux relatifs à la pollution mondiale

Introduction à la régression linéaire

Pour mettre en évidence les facteurs jouant un rôle déterminant dans la quantité de CO₂ émis, nous effectuons tout d'abord une analyse de régression linéaire simple. Le facteur le plus important est sans doute le produit intérieur brut, qui sera notre variable explicative.

Pour commencer, nous avons choisi les 131 pays de notre base de données desquels nous en avons retiré quatre (la Belgique, l'Espagne, la Russie et le Brésil) afin de tester la droite de régression ultérieurement. SPSS nous a donné les résultats suivants :

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.919 ¹	0.845	0.844	199.5815

Coefficients²

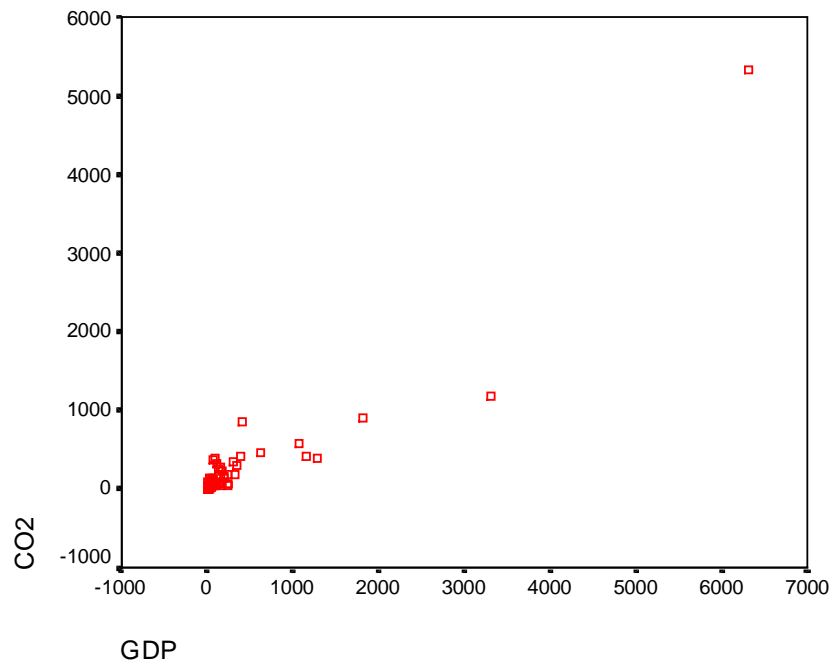
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
1 (Constant)	20.343	18.079		0.263	0.263	-15.426	56.113
GDP	0.705	0.027	0.919	0.000	0.000	0.652	0.757

Le coefficient de corrélation de 0.919 nous montre qu'il existe une liaison très forte et positive entre l'émission de CO₂ des pays observés et leur niveau de PIB. Une régression linéaire décrirait donc bien ce lien. Le coefficient de détermination est également très élevé, et nous montre que déjà seule notre variable PIB permet d'expliquer 84.4% de la variable expliquée.

Toutefois, nous ne restons pas à cette solution. Un test de Student ne nous confirme pas la signification du lien et le graphique nous donne une autre information très intéressante.

¹ Predictors: (Constant), GDP

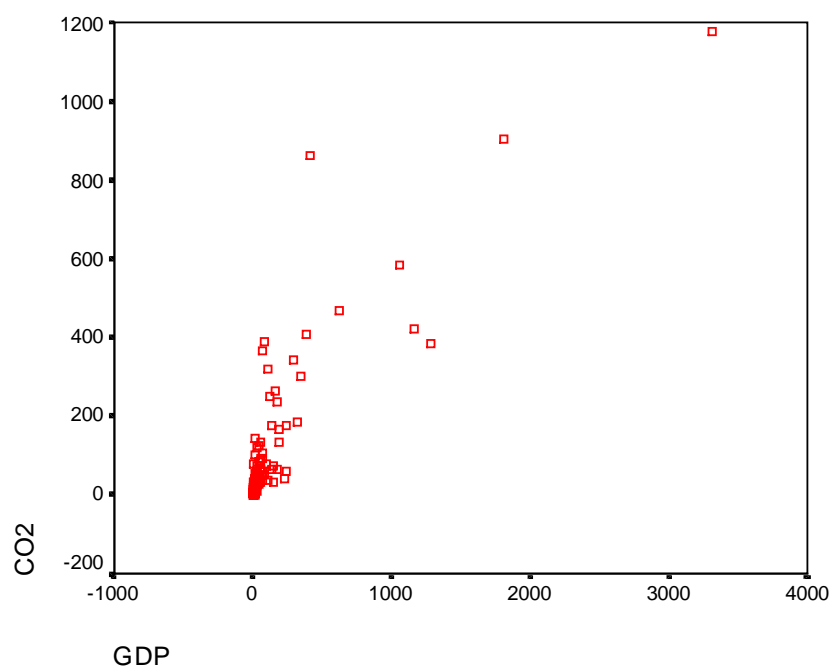
² Dependent Variable: CO₂



En effet, comme on voit sur le graphique, un seul pays (les États-Unis) influence trop notre droite de régression. Avec un niveau de PIB de plus de 6300 milliards de dollars, le pays a une émission de CO₂ de 5300 millions de tonnes. Sa structure de consommation d'énergie est totalement différente de celle des autres pays, ce qui nous empêche de le comparer avec ceux-ci sur les mêmes bases.

Pour éviter une régression trop biaisée, nous préférons donc abandonner les États-Unis et nous concentrer sur les 130 pays restants (dont nous en gardons toujours quatre pour nos tests ultérieurs).

Les couples de données PIB - CO₂ des 126 pays finalement choisis sont présentés par le graphique suivant.



La forme d'ellipse du nuage des points nous permet d'avoir une première analyse visuelle. Nous voyons qu'une droite d'ajustement peut être superposée, mais la qualité d'ajustement ne peut pas être évaluée avec précision. Pour cela, nous allons examiner de nouveau le tableau de résultat fourni par le logiciel.

Résultats de la régression linéaire

Variables Entered/Removed³

Model	Variables Entered	Variables Removed	Method
1	GDP ⁴	.	Enter

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.85 ⁵	0.722	0.72	95.2126

ANOVA⁶

Model		Sum of Square	df	Mean Square	F	Sig.
1	Regression	2924758.4	1	2924758.4	322.628	0.000 ⁷
	Residual	1124113.5	124	9065.432		
	Total	4048871.9	125			

Coefficients⁸

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		
	B	Std. Error	Beta			Lower Bound	Upper Bound	
1	(Constant)	43.51	8.923		4.873	0.000	25.85	61.171
	GDP	0.403	0.022	0.85	17.962	0.000	0.359	0.447

Nous trouvons la droite de régression linéaire suivante : $\hat{CO}_2 = 43.510 + 0.403 \cdot GDP$

Nous allons faire un test d'ajustement de χ^2 pour s'assurer de la normalité des distributions individuelles. Si le test est positif nous pourrions calculer le r.

Nous comparons une distribution observée (les données brutes sur l'émission de CO_2) avec une distribution théorique (l'émission de CO_2 calculé avec la régression linéaire simple).

³ Dependent Variable: CO_2

⁴ All requested variables entered

⁵ Predictors: (Constant), GDP

⁶ Dependent Variable: CO_2

⁷ Predictors: (Constant), GDP

⁸ Dependent Variable: CO_2

L'hypothèse nulle postule que la variable observée suit une loi théorique donnée. Si H_0 est vraie, alors chaque valeur observée doit évaluer une valeur correspondante de la distribution théorique.

$$\chi^2 = \sum \frac{(Obs - Th)^2}{Th}$$

Le χ^2 calculé nous donne 105.09

Le χ^2 théorique avec 125 degrés de liberté (k-1) et un seuil de risque de 5% est de 152.09

χ^2 théorique > χ^2 calculé, nous acceptons H_0 , ce qui signifie que les écarts observés sont dus au hasard.

Le coefficient de corrélation linéaire de 0.85 indique une liaison forte et positive. Pour vérifier cette corrélation, nous posons les hypothèses suivantes, que nous analyserons à l'aide d'un test de Student:

- **$H_0: r = 0$** , donc il n'y a pas de liaison entre les deux variables
- **$H_1: r \neq 0$**

Le test de Student à n-2 degrés de liberté avec une intervalle de confiance de 95% donne une valeur de 17.962. Le logiciel fournit une p-value de 0.000; nous rejetons donc H_0 et concluons que la corrélation observée entre le PIB et les rejets de CO_2 est significativement différente de 0, il y a donc bien une liaison entre les deux variables. De plus, les coefficients de la droite de régression sont bien dans l'intervalle de confiance, ce qui nous montre également que la relation est significative.

Nous remarquons également que la p-value sur la constante est significative.

Le coefficient de détermination de 0.722 indique une bonne qualité de régression. Le modèle explique 72.2% de l'émission de CO_2 . Pour évaluer notre modèle, nous utilisons le test de Fisher suivant:

- **$H_0: R^2 = 0$** , donc la qualité de l'ajustement n'est pas significative (égalité de la variance expliquée et de la variance résiduelle)
- **$H_1: R^2 \neq 0$**

La valeur de F calculée étant de 322.628, avec une p - value de 0.000, nous rejetons l'hypothèse 0. Notre modèle est donc vraiment pertinent.

L'équation de régression est : $\hat{CO}_2 = 43.510 + 0.403 * GDP$

Nous avons par conséquent constaté que le modèle était intéressant. Nous allons prédire l'émission de CO_2 pour les pays que nous avons préalablement retirés, c'est à dire la Belgique, l'Espagne, le Brésil et la Russie.

Belgique

- $43.51 + 0.403 * 211.15 = 128.6$ millions de tonnes de CO_2
- L'émission de CO_2 réel est de 124.61 millions de tonnes de CO_2

Espagne

- $43.51 + 0.403 \cdot 540.93 = 261.5$ millions de tonnes de CO_2
- L'émission de CO_2 réel est de 235.57 millions de tonnes de CO_2

Brésil

- $43.51 + 0.403 \cdot 557.75 = 268.28$ millions de tonnes de CO_2
- L'émission de CO_2 réel est de 285.6 millions de tonnes de CO_2

Russie

- $43.51 + 0.403 \cdot 342.08 = 181.37$ millions de tonnes de CO_2
- L'émission de CO_2 réel est de 1516.67 millions de tonnes de CO_2

Conclusion de la régression linéaire

Nous remarquons que pour les 3 premiers pays, l'estimation est relativement bonne. Nous pouvons en conclure que le produit intérieur brut explique en grande partie l'émission de CO_2 . Cela confirme notre première intuition.

Par contre, nous voyons que pour la Russie, le produit intérieur brut n'est pas forcément le meilleur facteur expliquant l'émission de CO_2 , il y en a sûrement d'autres. Notre modèle n'est donc pas complet et nous allons essayer de l'affiner en faisant une analyse factorielle.

Analyse en composantes Principales (ACP)

Nous avons vu précédemment que le PIB ne suffisait pas à expliquer suffisamment bien la quantité d'émission de CO₂. C'est la raison pour laquelle nous faisons ici usage de l'ACP.

Objectifs de l'analyse en composantes principales

La partie de régression simple nous a montré que nous devons apparemment aussi prendre en considération d'autres variables que le PIB pour mieux expliquer les quantités d'émissions de CO₂ des divers pays.

Intuitivement, nous comprenons la nécessité de la prise en considération la taille de la population (consommation de base d'énergie), la production indigène (par exemple d'électricité; le cycle énergétique est responsable pour 75% des émissions d'activités humaines), les importations nettes d'énergie (ce qui explique une quantité moins élevée d'émissions dans un pays importateur net), la consommation d'électricité, etc. pour expliquer les émissions de CO₂ dans un pays.

Dans la publication du « International Energy Agency » (Key World Energy Statistics, 1998 edition (édition la plus récente)) utilisée ici, nous avons obtenu des données (Energy Indicators), qui pourraient nous aider à mieux expliquer ces émissions. Mais ces données sont probablement très corrélées. Une nation industrialisée, avec une (relativement) grande population, a besoin d'une « Total Primary Energy Supply » élevée, a une consommation d'électricité élevée, etc.

Pour trouver des facteurs variables non corrélées, nous devons donc simplifier ce tableau par regroupement des variables initiales. Cette structuration nous permettra de mettre à jour les composantes cachées, ce qui simplifiera l'interprétation.

En plus, la première composante principale nous donne immédiatement la régression multiple, plus complète que la régression simple (si celle-ci est validée naturellement).

La technique de l'Analyse en Composantes Principales

Nous allons chercher des combinaisons linéaires des 8 variables de départ qui contribuent, chacune pour une part décroissante, à la variance des données de départ. Les coefficients constituent le maximum de variance contenu dans le tableau de données initiales. Le logiciel SPSS fait automatiquement les calculs, mais ce sera à nous de spécifier les tableaux que nous voulons avoir (s'ils sont différents de ce que le programme nous donne comme tableaux standards), le nombre et les noms des nouveaux composants.

Processus de calcul

Il est intéressant de voir initialement les statistiques descriptives des variables dans notre analyse factorielle.

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
Population	111.4963	518.9459	140
GDP	513.6055	2667.0129	140
GDP (PPP)	636.4199	3097.6908	140
Energy Prod.	194.5152	882.2574	140
Net Imports	352,6786	15.680,0127	140
TPES	193.0129	924.1162	140
Elec. Cons.	262.0264	1295.2032	140
CO2 Emissions	457.5786	2221.6552	140

Nous travaillons sur des données centrées réduites parce que nous avons des données de départ avec des unités de mesure extrêmement hétérogènes, ce qui nous donne des moyennes et des variances très différentes.

Avant de commencer les analyses, nous avons vérifié que les données soient factorisables. L'ACP a un sens ici, car les corrélations entre un grand nombre de variables sont très élevées. D'un autre côté, toutes les variables ne sont pas tellement corrélées, donc nous n'allons pas trouver un nuage à une seule dimension. Nous remarquons que les importations nettes forment l'exception dans notre tableau.

Ceci nous suggère deux dimensions.

Correlation Matrix⁹

		Popula- tion	GDP	GDP (PPP)	Energy Prod.	Net Imports	TPES	Elec. Cons.	CO2 Emissions
Correlation	Population	1.000	.814	.902	.934	.116	.912	.863	.905
	GDP	.814	1.000	.984	.930	.515	.972	.990	.971
	GDP (PPP)	.902	.984	1.000	.971	.413	.995	.993	.993
	Energy Prod.	.934	.930	.971	1.000	.200	.986	.967	.985
	Net Imports	.116	.515	.413	.200	1.000	.356	.431	.357
	TPES	.912	.972	.995	.986	.356	1.000	.994	.999
	Elec. Cons.	.863	.990	.993	.967	.431	.994	1.000	.994
	CO2 Emissions	.905	.971	.993	.985	.357	.999	.994	1.000
Sig. (1-tailed)	Population		.000	.000	.000	.086	.000	.000	.000
	GDP	.000		.000	.000	.000	.000	.000	.000
	GDP (PPP)	.000	.000		.000	.000	.000	.000	.000
	Energy Prod.	.000	.000	.000		.009	.000	.000	.000
	Net Imports	.086	.000	.000	.009		.000	.000	.000
	TPES	.000	.000	.000	.000	.000		.000	.000
	Elec. Cons.	.000	.000	.000	.000	.000	.000		.000
	CO2 Emissions	.000	.000	.000	.000	.000	.000	.000	

Le tableau suivant nous donne d'abord les différentes composantes trouvées (première colonne), la variance de chacune des composantes (deuxième colonne), c'est-à-dire la valeur propre, la part de variance initiale représentée par chacune de celles-ci (troisième colonne) et la valeur cumulative de cette variance expliquée.

Total Variance Explained¹⁰

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.875	85.938	85.938	6.875	85.938	85.938
2	.992	12.397	98.335	.992	12.397	98.335
3	.118	1.478	99.813			
4	1.196E-02	.149	99.962			
5	1.470E-03	1.837E-02	99.981			
6	9.507E-04	1.188E-02	99.993			
7	4.881E-04	6.101E-03	99.999			
8	9.672E-05	1.209E-03	100.000			

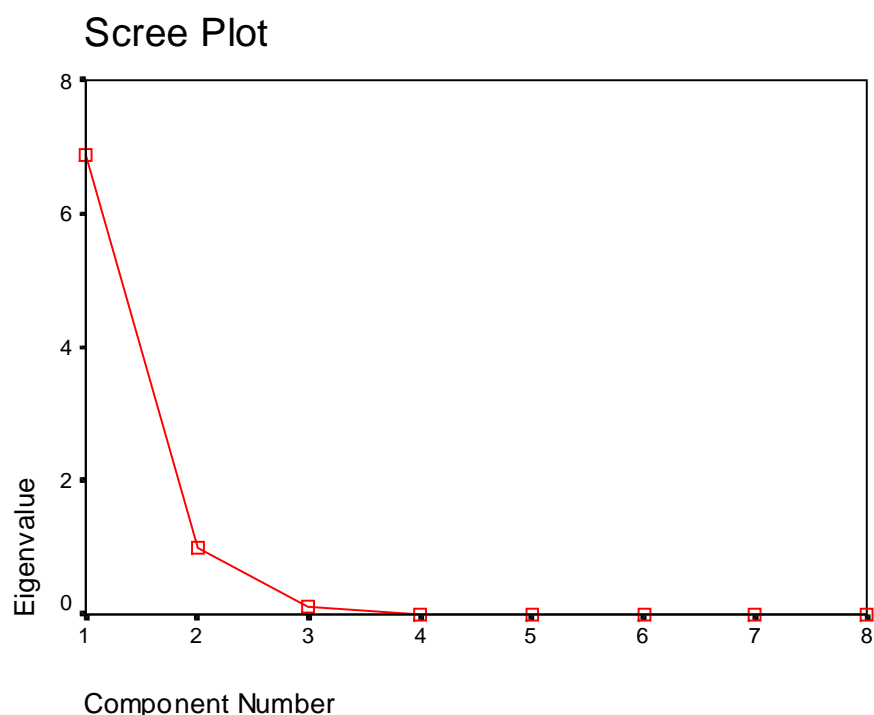
Pour la première composante calculée, nous avons donc une valeur propre (variance) de 6.875, ce qui représente 85.938% de la variance des variables centrées réduites initiales. Les deux premières composantes représentent donc 98.335% de l'ensemble initial, ce qui nous semble très élevé.

⁹ Determinant = 6.360E-16

¹⁰ Extraction Method: Principal Component Analysis.

Nous voyons donc que le premier composant génère une valeur propre élevée. Nous devons en tout cas prendre ce composant en considération. En fait le critère de Kaiser nous dit de prendre toutes les composantes avec une valeur propre supérieure à un. Nous devrions, avec ce critère, laisser le deuxième composant de côté, car ce dernier explique moins que l'une des variables initiales.

Nous avons cependant opté pour le prendre dans l'analyse par le fait qu'il explique presque 1% de la variation (même variance que les variables initiales, donc il apporte autant d'informations). Cela nous donne une deuxième composante principale que le scree-test valorise comme acceptable. Dans le Scree-Plot, on retient les valeurs propres se situant au-dessus des « éboulis ».



C'est la raison pour laquelle nous avons configurés le logiciel afin de prendre les deux premières composantes (partie droite du tableau ci-dessus).

Le tableau suivant nous montre la « Kaiser-Meyer-Olkin Measure of Sampling Adequacy » et le test de Bartlett pour la sphéricité. Les deux sont significatifs.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.781
Bartlett's Test of Sphericity	Approx. Chi-Square	4741.334
	df	28
	Sig.	.000

Nous devons maintenant interpréter ces deux composantes. Le sens caché est révélé par les variables initiales auxquelles elles sont le plus corrélées. SPSS nous calcule ces corrélations (loadings). Toutes les variables initiales, sauf les importations nettes, sont fortement corrélées avec la première composante. Nous pourrions la nommer par exemple « Industrialised & Developed ». La deuxième n'est que fortement corrélée avec cette importation nette. Donc, nous l'appelons « Dependency ». En fait, il est

logique que cette variable ne se trouve pas dans le nuage des autres. Si un pays importe de l'électricité, on doit la générer dans un autre pays. Nous ne pouvons pas montrer ce lien directement.

Component Matrix¹¹ - 2 components extracted

	Component 1	Component 2
Population	.907	-.306
GDP	.980	.147
GDP (PPP)	.999	1.457E-02
Energy Prod.	.976	-.206
Net Imports	.401	.910
TPES	.998	-4.668E-02
Elec. Cons.	.995	4.449E-02
CO2 Emissions	.997	-4.322E-02

Le « Display factor score coefficient matrix » montre les coefficients par lesquels nous devons multiplier les variables pour obtenir les scores des facteurs. Ceci nous permet en même temps d'obtenir la régression multiple. Les chiffres dans la première colonne de la matrice suivante nous donnent le premier axe et donc la régression.

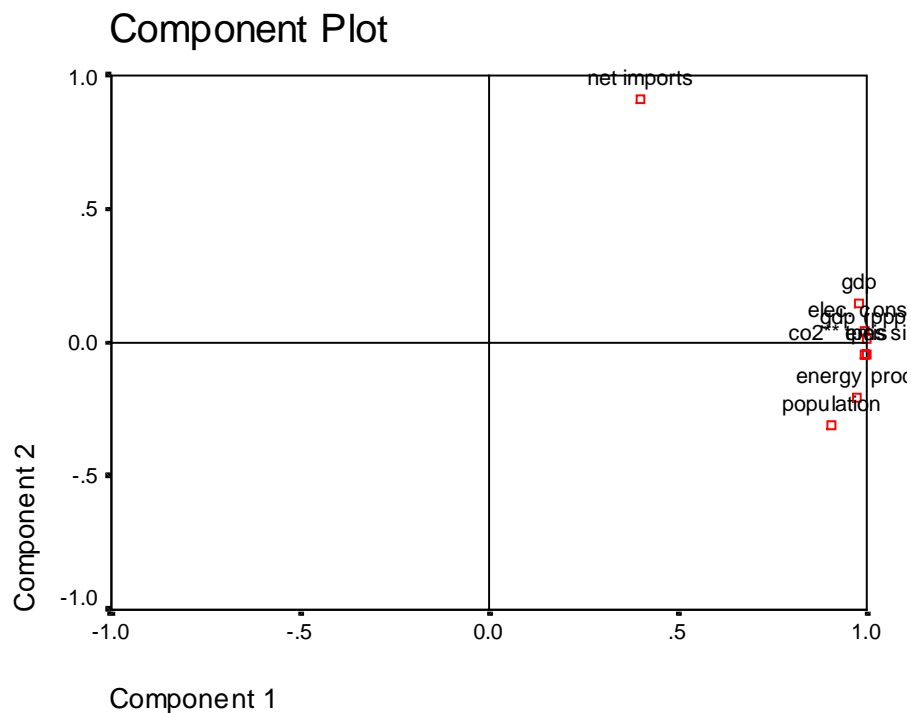
Component Score Coefficient Matrix¹²

	Component 1	Component 2
Population	.132	-.308
GDP	.143	.149
GDP (PPP)	.145	.015
Energy Prod.	.142	-.208
Net Imports	.058	.917
TPES	.145	-.047
Elec. Cons.	.145	.045
CO2 Emissions	.145	-.044

Le graphique nous facilite l'interprétation des composantes. L'axe horizontal, « Industrialised & Developed », est donc fortement corrélé avec les variables dans le nuage périphérique et près du premier. La variable « Net imports » l'est aussi, mais moins fortement avec ce premier axe. Les pays industrialisés sans grande production indigène doivent naturellement importer de l'énergie. Cette importation est par contre périphérique et dans la direction du deuxième axe, donc corrélée.

¹¹ Extraction Method: Principal Component Analysis

¹² Extraction Method: Principal Component Analysis



Pour une vérification complémentaire de la qualité de l'interprétation, nous examinons maintenant les « communalities » des variables. Une communalité est la part de la variance des variables initiales, restituée par les deux composantes retenues. Il n'y a pas de communalités près de 0, donc les variables sont bien prises en compte par les deux composantes.

Communalities¹³

	Initial	Extraction
Population	1.000	.916
GDP	1.000	.983
GDP (PPP)	1.000	.998
Energy Prod.	1.000	.996
Net Imports	1.000	.989
TPES	1.000	.998
Elec. Cons.	1.000	.991
CO2 Emissions	1.000	.996

Avec les nouvelles composantes, nous avons naturellement le tableau suivant. Les composantes sont orthogonales et n'ont par conséquent pas de covariance.

Component Score Covariance Matrix¹⁴

Component	1	2
1	1.000	.000
2	.000	1.000

¹³ Extraction Method: Principal Component Analysis

¹⁴ Extraction Method: Principal Component Analysis

Conclusion de l'analyse en composantes principales

Cette partie nous a fourni une simplification du tableau de base et une compréhension améliorée de celui-ci. Ces variables, sauf une, sont fortement corrélées et nous pouvons donc les synthétiser dans une variable. Celles qui se trouvent dans le nuage de la première composante principale pourraient bien expliquer les émissions de CO₂. Mais ceci ne veut pas dire que notre modèle de base est juste. Nous sommes convaincus que nous pouvons encore améliorer la capacité explicative et prédictive par la prise en considération d'autres variables. Mais ceci conduirait trop loin, et serait un excellent sujet pour une recherche approfondie.

La classification

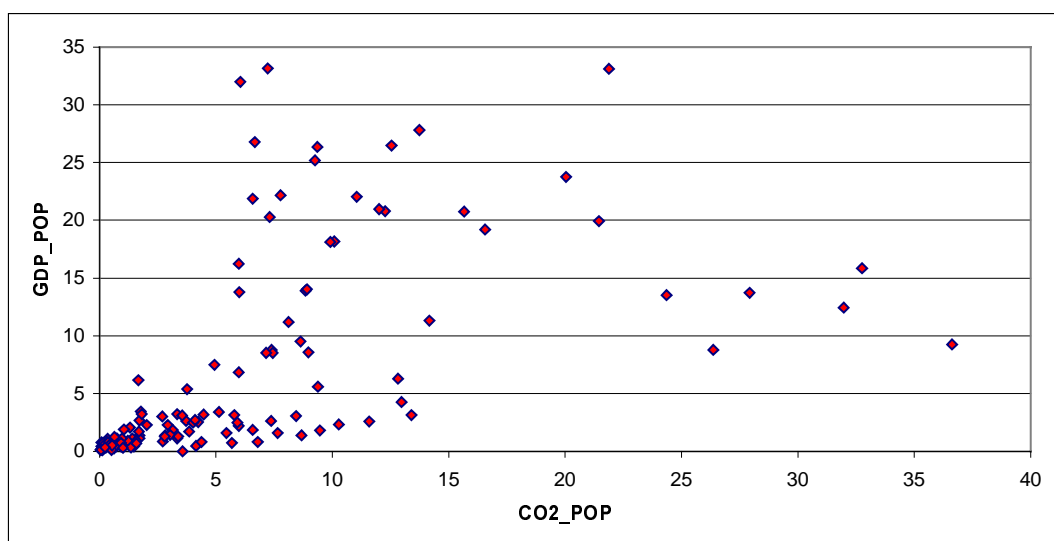
Nous allons, dans cette section, ranger les différents pays en un nombre déterminé de classes afin de pouvoir faire ressortir des tendances liées aux facteurs considérés

Introduction à la classification

Nous nous sommes jusqu'ici intéressés à la pollution au niveau mondial, ceci pour plus de 130 pays pris séparément. Nous avons effectué nos différentes analyses précédentes sur chacune de ces variables. Il peut cependant être fort intéressant de regrouper les différents pays en des classes (clusters en anglais). Ceci nous permettra de faire ressortir les tendances des différents groupements.

Pour effectuer notre classification, nous allons nous intéresser à deux variables qui sont ressorties comme étant essentielles dans nos analyses précédentes. Il s'agit du PIB et de la quantité de CO₂. Afin de pouvoir comparer les différentes valeurs obtenues, nous avons ici rapporté ces valeurs à l'individu en divisant celles-ci par la taille de la population du pays considéré.

Nous obtenons ainsi le graphique suivant, qui représente le PIB et la quantité de CO₂ rapportés aux individus des différents pays.

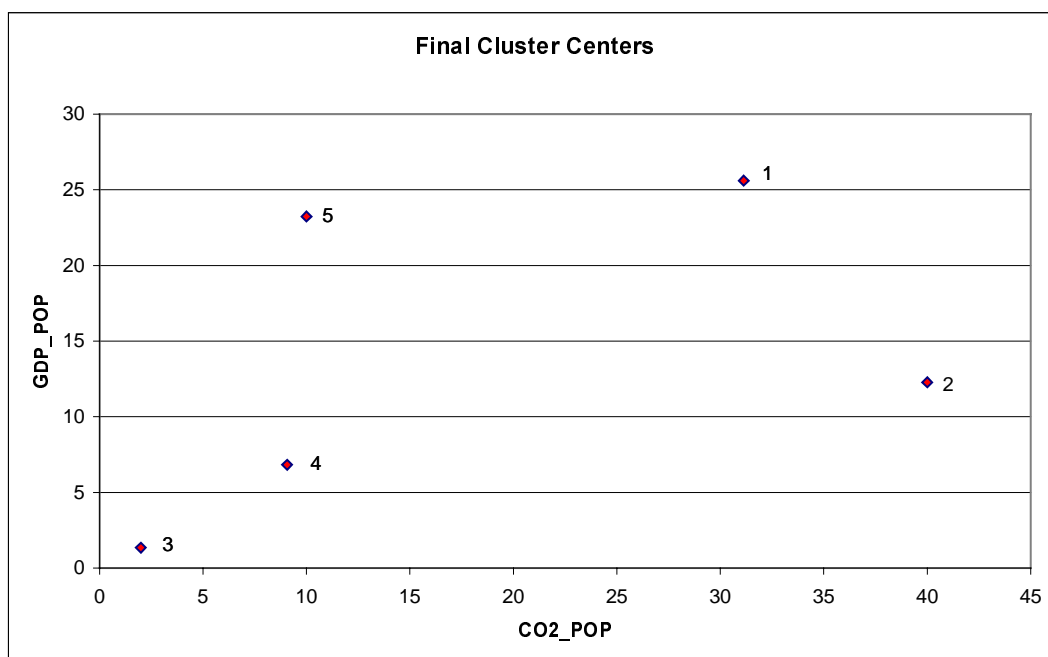


Les classifications effectuées

Afin d'effectuer les classifications, nous avons utilisé le logiciel SPSS et sa fonction « K-mean cluster » qui tente d'identifier des groupes relativement homogènes basés sur des caractéristiques prédéfinies, en utilisant un algorithme nécessitant la définition préalable du nombre de classes. Nous avons ainsi effectué une série de tests avec différents nombres de groupes pour les répartitions. Ci-dessous est représentée celle en 5 clusters. Il est évident que la meilleure classification est celle comprenant, dans notre cas, 131 clusters (un par pays). Une telle approche ne nous permet cependant pas de découvrir une tendance « générale » pour un groupe de pays.

Classification en 5 classes (clusters)

La première idée de répartition consisterait à répartir suivant 5 grands groupements qui sont « intuitivement » l'Europe, l'Amérique du nord, l'Amérique du sud, l'Asie ainsi que l'Afrique. En effectuant une analyse avec 5 clusters, nous remarquons cependant que cette intuition n'est pas fondée. Nous obtenons en effet la situation suivante (seule la position du centre des clusters est représentée dans ce graphique) :



Cluster	CO2_POP	GDP_POP	Nb. pays
1	31.1233	25.6038	3
2	39.9967	12.263	6
3	1.9968	1.3383	81
4	9.0665	6.8264	23
5	10.0028	23.2346	18

Cluster 1 :

Celui-ci regroupe le Luxembourg, Singapour, et les USA. Il peut sembler étrange de trouver ces trois pays dans un même groupe, mais cela s'explique par le fait que le rapport entre la quantité de CO₂ et le PIB est environ égal à 1,2. Il va de

soi que ces quantités ne sont pas du tout dans les mêmes proportions. Cependant, au niveau des individus, celles-ci sont relativement proches.

Cluster 2 :

Nous retrouvons ici des pays tels le Brunei, Koweït, le Qatar et les Émirats Arabes Unis. Dans ce cas, le rapport entre le CO₂ et le PIB est de plus de 3. La logique semble bien dicter le regroupement de ces pays car il s'agit de producteurs de pétrole à faible population.

Cluster 3 :

Ce groupe, qui à un rapport de 1.5 entre le CO₂ et le PIB, regroupe d'une façon générale presque tous les pays d'Afrique et d'Amérique du Sud tels l'Algérie, l'Argentine, le Brésil, le Congo, l'Égypte, le Mexique, Cuba, ainsi que l'Inde, l'Iran, l'Iraq, ... Le PIB par habitant est de plus très faible dans ces pays.

Cluster 4 :

Nous trouvons ici notamment la République Tchèque, la Grèce, Israël, la Corée, la Chine, la Libye, la Nouvelle Zélande, la Pologne, la Russie, l'Afrique du Sud et l'Espagne. Le rapport qui nous intéresse dans ces groupes (CO₂ / PIB) est ici de 1,5. Il est important de noter que dans le cas de cette classe, le PIB par habitant est environ 4,5 fois supérieur à celui de la classe 3

Cluster 5 :

Ce dernier groupe se compose de pays tels l'Australie, l'Autriche, la Belgique, le Canada, les pays de l'Europe du nord, la France, l'Italie, l'Allemagne, le Japon, la Suisse, ... Dans ce cas, nous arrivons à un rapport de environ 0,5 entre le CO₂ et le PIB.

Ci-dessous se trouvent les informations plus « statistiques » concernant la répartition des différents clusters dans le graphique, ainsi que leur éloignement respectif.

Distances entre les centres des groupes

Cluster	1	2	3	4	5
1	0.0000				
2	16.0223	0.0000			
3	30.8972	30.0556	0.0000		
4	22.3150	21.6247	8.9498	0.0000	
5	11.3701	22.8064	23.3140	16.4349	0.0000

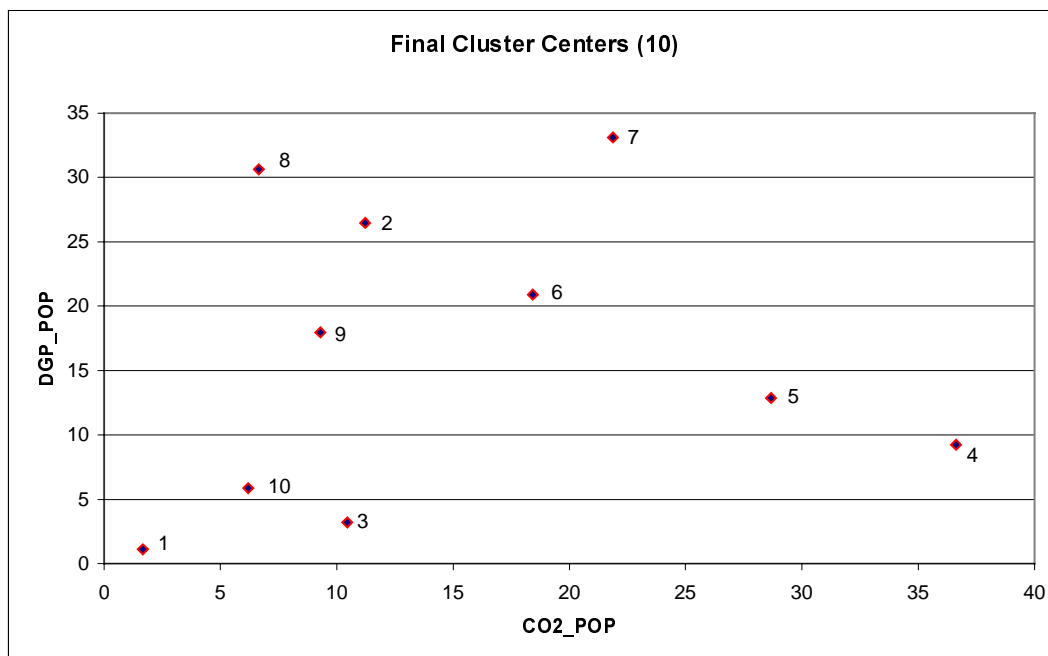
ANOVA

Variable	Cluster MS	DF	Error MS	DF	F	Prob
CO2_POP	1485.8748	4	5.536	126	268.3911	0
GDP_POP	2123.2934	4	8.016	126	264.8714	0

Nous remarquons ici que les groupes sont assez éloignés les uns des autres pour être utiles et utilisables.

Classification en 10 classes (clusters)

Si nous effectuons une classification en 10 groupes par exemple, nous constatons immédiatement que la réparation est nettement plus prononcée. Ce genre de regroupements peut être très intéressant lorsque nous souhaitons affiner notre analyse en fonction des groupes de pays.



Cluster	CO2_POP	GDP_POP	Size
1	1.6774	1.1035	73
2	11.225	26.4481	4
3	10.461	3.1951	10
4	36.63	9.25	1
5	28.67	12.8656	5
6	18.435	20.9077	4
7	21.87	33.0952	1
8	6.6433	30.6436	3
9	9.2992	17.9689	13
10	6.1941	5.8761	17

Nous remarquons cependant très rapidement un problème avec un trop grand nombre de classes. Nous arrivons, déjà avec 10 classes, à la situation où des clusters ne contiennent qu'un unique pays. C'est par exemple le cas du cluster n° 4 qui ne contient que les Antilles et du cluster n° 7 qui ne contient que le Luxembourg.

Commentaires sur la classification

La première remarque qui découle logiquement d'une telle approche est qu'il n'est pas possible de faire, en terme d'émission de CO₂ par les pays, une répartition géographique qui peut paraître intéressante lors d'une première réflexion. Comme nous l'avons très bien remarqué dans la situation à 5 groupes, il faut plutôt effectuer une classification sur la base de critères économiques et politiques.

C'est ainsi que nous avons eu le regroupement, dans le cas à 5 groupes, des pays d'Europe occidentale avec le Japon et l'Australie ou des pays producteurs de pétrole entre eux. Le regroupement entre les USA, le Luxembourg et Singapour est un exemple très clair de répartition basée sur des critères économiques.

Il faut aussi faire très attention avec le « dosage » du nombre de groupes à former. A partir d'une certaine quantité, l'information n'est plus utile comme synthèse puisqu'il apparaît des groupes ne contenant qu'un unique pays. Il faut cependant noter que la classification en un grand nombre de groupes nous permet une bien plus grande précision.

Conclusion

Cette section contient une synthèse des différents points étudiés précédemment.

Conclusion sur le sujet choisi

La problématique des gaz à effet de serre commence à mobiliser le monde entier. Il est vrai que nous ne pouvons continuer sur la voie actuelle. Plusieurs études nous montrent la forte corrélation entre le réchauffement de la terre et les émissions de gaz à effet de serre. A titre d'exemple, un réchauffement de 3 degrés devrait suffire pour faire disparaître les Pays-Bas de la surface du globe.

Les émissions de CO₂ comptent pour 60 % du total des émissions. De plus, cette molécule est l'un des gaz sur lequel nous avons la possibilité d'agir.

Mais pour pouvoir agir, il est indispensable de comprendre la problématique. Intuitivement, nous pouvons déjà indiquer quelques causes, mais les relations qui existent entre les variables, et le degré d'influence provenant de celles-ci sont moins claires. Pour comprendre cette problématique, nous devons effectuer des recherches profondes.

C'est ce que nous avons entrepris ici, à une échelle réduite cependant. Les données de la publication de l'Agence internationale de l'électricité (IEA) ont formé la base de nos traitements statistiques. Nous avons essayé de retrouver quelques résultats intéressants. En commençant avec une simple régression, nous avons ensuite essayé de voir plus claire dans le tableau de base (quelles sont les variables sous-jacentes) en faisant usage de l'analyse factorielle (ceci nous a généré en même temps l'analyse multiple). Nous avons terminé par une classification pour regrouper les pays dans des groupes plus ou moins homogènes.

A partir de nos résultats, il serait intéressant de refaire les premières analyses dans les groupes retrouvés dans la troisième partie. Il est en effet clair que les analyses dans les deux premières parties se situent sur un plan trop agrégé.

Conclusions sur le travail effectué

Des analyses assez détaillées ne sont intéressantes qu'après avoir vu la globalité de la situation. Notre travail pourrait donc être considéré comme un balayage (superficiel) d'un domaine fort complexe.

Nous avons utilisé le programme SPSS pour les analyses. Nous avons constaté que le logiciel nous fournit sans problèmes les résultats des analyses, mais que le vrai travail ne commence qu'ensuite, c'est-à-dire lors de l'interprétation de ces résultats.

Cette introduction au logiciel SPSS nous a également montré ses lacunes et limites tel le manque de compatibilité avec Microsoft Word.

Les méthodes utilisées

S'il fallait résumer les différentes méthodes utilisées, nous aurions ceci :

Régression Linéaire

Cette manière d'aborder le problème à permis de « valider » notre intuition concernant la forte relation entre la quantité de CO₂ émise par un pays et son PIB. Certains cas « extrêmes » tels la Russie ne peuvent cependant pas confirmer le modèle de régression linéaire obtenu. Cela prouve que cette méthode, quoique très utile dans une première approche, est rarement suffisante pour expliquer complètement un phénomène.

Analyse factorielle

Il s'agit d'une manière facile de déceler les composantes sous-jacentes. Nous avons trouvé deux composantes: « industrialised and developed » et « dependency ». Il est intéressant de noter que la première composante représente également une régression multiple.

Classification

Méthode très utile pour regrouper les différentes composantes de la population observée en différents groupes. Ceci permet d'effectuer des analyses et des réflexions à un niveau supérieur. Il faut cependant faire très attention au dosage du nombre de classes pour que ce regroupement reste pertinent et utile.

Annexe

Cette partie contient les données utilisées pour le cas traité ci-dessus et proviennent du le « KEY WORLD ENERGY STATISTICS » de l'Agence Internationale de l'Énergie (IEA), édition 1999

Annexe 1 - Selected Energy Statistics for 1996

Region	N°	Population	GDP	GDP (PPP)	Energy Prod.	Net Imports	TPES	Elec. Cons.*	CO2** Emissions
		(million)	(billion 90 US\$)	(billion 90 US\$)	(Mtoe)	(Mtoe)	(Mtoe)	(TWh)	(Mt of CO2)
World (a)		5624.44	24193.78	30941.41	9434.41	..	9447.86	12539.58	22620.46
OECD		1092.29	19372.64	18441.99	3753.75	1338.55	5019.76	8193.1	12117.05
Middle East		155.2	510.37	579.01	1127.25	-798.84	315.84	315.2	882.27
Fomer USSR		291.95	488.56	1006.11	1198.02	-277.01	932.42	1122.85	2346.07
Non-OECD Europe (b)		60.42	138.71	240.11	66.57	57.02	121.55	186.99	319.6
China		1215.41	696.66	3594.2	1100.39	-7.11	1096.8	999.48	3141.93
Asia (c)		1698.74	1409.5	3925.24	866.88	113.93	960.8	819.97	1867.12
Latin America		389.93	1075.77	2027.45	539.38	-102.76	424.17	547.22	840.36
Africa		720.5	501.58	1127.31	782.16	-328.63	445.9	354.78	691.48
Albania	1	3.29	2.01	0	1.08	0.11	1.19	3.02	1.76
Algeria	2	28.73	65.06	86.42	116.21	-91.65	24.15	16.69	58.72
Angola	3	11.1	3.89	3.52	40.49	-33.82	6.02	0.74	3.07
Argentina	4	35.22	189.38	245.53	74.86	-15.25	58.92	60.36	132.13
Armenia	5	3.77	1.54	10.96	0.74	1.05	1.79	3.87	2.57
Australia	6	18.29	350.84	332.63	189.04	-88.6	100.61	165.45	302.96
Austria	7	8.06	178.73	146.18	7.79	19.55	27.19	52.54	62.67
Azerbaijan	8	7.58	3.6	10.36	14.39	-2.53	11.86	13.75	31.39
Bahrain	9	0.6	5.27	5.79	7.64	-1.13	6.55	4.71	15.79
Bangladesh	10	121.67	28.51	130.39	21.5	2.43	23.93	8.04	20.61
Belarus	11	10.3	22.56	49.16	3.15	19.93	24.57	28.51	61.54
Belgium	12	10.16	211.15	183.9	12.14	49.05	56.4	76.46	124.61
Benin	13	5.63	2.41	7.99	1.95	-0.03	1.92	0.27	0.43
Bolivia	14	7.59	5.64	17.04	5.24	-1.63	3.63	2.85	7.5
Bosnia-Herzegovina	15	3.18	1.31	0	0.63	1.12	1.75	1.9	4.66
Brazil	16	161.37	557.75	877.7	112.32	55.37	163.37	276.88	285.6
Brunei	17	0.29	3.98	2.84	19.06	-15.44	3.35	1.72	8.1
Bulgaria	18	8.36	15.43	28.48	10.35	12.95	22.61	36.69	54.96
Cameroon	19	13.68	10.71	21.63	10.02	-5.02	5	2.32	1.17
Canada	20	29.96	621.15	568.69	357.28	-120.74	236.17	491.81	469.56
Chile	21	14.42	46.58	165.49	7.84	13.18	20.46	28.1	47.95
Chinese Taipei	22	21.47	240.52	355.86	11.5	59.67	67.44	124.95	174.06
Colombia	23	37.45	51.4	189.85	66.74	-34.6	31.39	35.02	63.99
Congo	24	2.71	2.92	5.32	11.49	-9.5	1.2	0.55	0.94

Statistique : Analyse de données – La pollution mondiale

Dem. Rep. of Congo	25	45.23	6.42	30.32	13.69	0.14	13.8	5.92	4.87
Costa Rica	26	3.44	7.12	17.53	0.74	1.63	2.25	4.46	4.49
Croatia	27	4.77	8.91	0	3.91	2.9	6.76	11.81	14.99
Cuba	28	11.02	15.2	34.31	6.71	9.25	15.95	11.5	31.03
Cyprus	29	0.74	7.04	9.55	0.01	2.2	2.12	2.44	6.39
Czech Republic	30	10.32	26.71	93.16	31.94	9.05	40.4	59.1	119.56
Denmark	31	5.26	146.26	103.97	17.59	5.65	22.87	35.47	72.32
Dominican Republic	32	7.96	8.92	25.02	1.46	3.72	5.19	5.11	11.28
Ecuador	33	11.7	12.92	47.21	21.87	-12.67	8.55	7.49	20.09
Egypt	34	59.27	53.37	179.17	59.76	-19.52	37.79	57.56	94.21
El Salvador	35	5.81	6.59	15.46	2.61	1.46	4.06	3.06	4.37
Estonia	36	1.47	4.65	9.14	3.85	1.77	5.62	6.54	19.62
Ethiopia	37	58.23	8.44	25.11	15.54	1.09	16.57	1.3	3.03
Finland	38	5.12	135.46	83.25	13.57	17.32	31.48	70.32	64.24
France	39	58.38	1276.16	1077.21	130.09	125.13	254.2	412.67	384.3
Gabon	40	1.13	6.97	7.6	19.71	-18.01	1.58	0.87	1.87
Georgia	41	5.41	3.54	7	0.7	0.87	1.58	5.67	2.59
Germany	42	81.9	1805.4	1421.76	140.53	209.46	349.55	524.61	904.68
Ghana	43	17.52	7.65	31.77	5.6	1.08	6.66	5.58	4.31
Gibraltar	44	0.03	0.34	0.35	0	0.96	0.13	0.09	0.41
Greece	45	10.48	92.06	112.4	8.81	19.15	24.39	40.65	77.6
Guatemala	46	10.93	9.7	30.68	4	1.12	5.22	3.74	5.9
Haiti	47	7.34	2.29	3.85	1.6	0.37	1.97	0.29	1.11
Honduras	48	6.1	3.74	4.5	1.76	1.19	2.92	2.05	3.54
Hong Kong, China	49	6.31	102.46	139.51	0.05	14.01	12.19	31.64	37.83
Hungary	50	10.19	32.15	65.01	12.84	13.3	25.47	32.61	59.1
Iceland	51	0.27	6.8	5.05	1.4	0.88	2.27	4.67	2.49
India	52	945.12	411.17	1224.52	390.6	58.49	450.29	358.89	863.19
Indonesia	53	197.06	179.25	729.41	219.19	-86.69	132.42	58.78	237.85
Iran	54	62.51	158.14	78.44	220.89	-129.64	89.34	72.47	264.84
Iraq	55	21.37	23.63	18.9	32.58	-7.56	25.03	29	71.4
Ireland	56	3.62	65.75	57.85	3.47	8.65	11.96	17.41	36.48
Israel	57	5.69	79.03	79.79	0.57	15.32	16.18	30.21	50.39
Italy	58	57.47	1164.68	1007.88	29.3	135.47	161.14	264.96	419.96
Ivory Coast	59	14.35	12.56	21.44	4.76	1.4	5.3	2.56	4.35
Jamaica	60	2.55	4.41	9.54	0.55	3.09	3.72	5.39	9.79
Japan	61	125.86	3315.66	2590.91	102.46	415.63	510.36	975.26	1177.71
Jordan	62	4.31	6.07	21.56	0.18	4.31	4.49	5.46	13.07
Kazakhstan	63	16.47	22.35	52.97	62.63	-19.25	43.38	56.54	142.77
Kenya	64	27.36	9.65	35.96	11.25	2.24	13.28	3.55	6.07
Korea	65	45.55	389.75	542.88	22.75	147.53	162.87	215.43	408.87
DPR of Korea	66	22.45	0	0	21.05	2.95	24	5.61	80.21
Kuwait	67	1.59	21.52	31.78	112.6	-98.56	13.86	25.48	38.74
Kyrgyzstan	68	4.58	1.51	3.82	1.44	1.51	2.95	7.13	6.16
Latvia	69	2.49	6.58	9.55	1	3.43	4.17	4.9	9.26
Lebanon	70	4.08	5.25	19.98	0.2	4.55	4.75	6.73	13.74
Libya	71	5.17	28.86	27.33	77.74	-62.74	14.91	18.18	48.43
Lithuania	72	3.71	9.28	14.29	4.17	4.9	8.95	9.85	14.86
Luxembourg	73	0.42	13.9	11.84	0.04	3.42	3.45	6.06	9.1
FYR of Macedonia	74	1.98	1.49	0	1.79	1.6	3.42	5.68	11.25
Malaysia	75	20.57	70.32	178.19	69.56	-28.16	41.21	45.69	105.51
Malta	76	0.37	3.15	4.55	0	0.97	0.89	1.44	2.77
Mexico	77	96.58	298.09	675.72	213.52	-75.56	141.38	138.98	342.19
Moldova	78	4.33	3.82	7.66	0.05	4.4	4.6	6.34	11.78
Morocco	79	27.02	30.5	97.54	0.87	7.7	8.82	12	25.99
Mozambique	80	18.03	2.12	3.5	7.25	0.58	7.81	1.24	2.35

Statistique : Analyse de données – La pollution mondiale

Myanmar	81	45.88	34.3	120.09	11.83	0.68	12.77	2.71	7.13
Nepal	82	22.04	4.9	21.83	6.37	0.59	6.97	0.84	1.85
Netherlands	83	15.52	325.09	284.04	73.38	14.15	75.8	92.31	186.4
Netherlands Antilles	84	0.2	1.85	2.44	0	4.44	2.66	0.89	7.4
New Zealand	85	3.64	51.01	54.93	13.58	2.87	16.3	32.27	32.39
Nicaragua	86	4.5	1.14	4.39	1.49	0.93	2.39	1.38	2.75
Nigeria	87	114.57	33.73	119.76	170.45	-87.7	82.67	10.22	47.99
Norway	88	4.38	145.27	98.6	208.15	-182.59	23.15	105.29	31.62
Oman	89	2.17	14.83	32.16	47.31	-42.44	4.85	6.8	12.99
Pakistan	90	133.51	52.95	265.96	41.49	14.7	55.9	43.78	89.09
Panama	91	2.67	7.13	12.82	0.76	2.49	2.28	3.13	4.54
Paraguay	92	4.96	6.25	15.28	6.68	-2.53	4.29	4.71	3.18
Peru	93	24.29	45.59	91.08	12.35	1.59	13.93	14.61	25.54
Philippines	94	71.9	52.4	160.76	17.23	19.57	37.99	30.45	65.55
Poland	95	38.62	69.59	229.14	103.47	5.62	108.41	122.02	365.47
Portugal	96	9.94	74.54	115.6	2.43	16.96	19.15	32.16	49.07
Qatar	97	0.66	8.2	8.93	29.57	-22.72	8.73	6.17	21.03
Romania	98	22.61	36.4	76.21	31.32	14.4	45.82	54.97	123.22
Russia	99	147.74	342.08	687.89	948.71	-332.65	615.9	755.6	1516.67
Saudi Arabia	100	19.41	121.71	181.36	475	-380.9	92.24	89.62	248.68
Senegal	101	8.53	6.43	13.1	1.57	0.99	2.59	0.99	3.15
Singapore	102	3.04	60.59	84.14	0.06	37.65	23.85	23.02	65.29
Slovakia	103	5.34	14.09	34.44	4.82	13.01	17.45	27.17	39.4
Slovenia	104	1.99	16.92	0	2.76	3.46	6.17	10.38	14.25
South Africa	105	37.64	114.38	178.55	127.86	-27.45	99.08	179.72	317.29
Spain	106	39.27	540.93	521.3	32.67	74.79	101.41	159.91	235.57
Sri Lanka	107	18.3	10.38	49.99	4.2	2.86	6.79	3.76	7.78
Sudan	108	27.27	12.52	10.83	9.49	1.31	10.79	1.41	4.21
Sweden	109	8.9	238.2	156.96	31.8	21.42	52.57	135.69	59.36
Switzerland	110	7.11	227.52	146.27	10.48	15.16	25.62	52.21	42.92
Syria	111	14.5	18.5	44.73	33.71	-18.79	14.54	17	40.42
Tajikistan	112	5.93	1.81	5.44	1.34	2.17	3.51	13.56	5.93
Tanzania	113	30.49	4.26	17.29	13.07	0.75	13.8	1.76	2.21
Thailand	114	60	137.05	406.32	43.95	38.07	79.99	80.58	175.22
Trinidad	115	1.3	5.52	7.11	13.57	-4.94	7.89	4.09	16.82
Tunisia	116	9.13	15.88	44.83	6.29	0.49	6.68	6.77	15.43
Turkey	117	62.7	188.84	417.31	26.78	39.36	65.52	78.94	168.54
Turkmenistan	118	4.6	3.73	9.98	32.62	-20.46	12.16	6.17	31.25
Ukraine	119	50.72	80.51	157.36	79.04	74.89	153.94	162	387.91
United Arab Emirates	120	2.53	40.13	28.62	148.82	-106.14	32.34	19.74	82.91
United Kingdom	121	58.78	1064.57	1021.16	269.12	-33.92	234.72	334.93	582.8
United States	122	265.56	6316.4	6316.4	1687.3	470.38	2134.96	3462.92	5324.52
Uruguay	123	3.2	10.4	25.71	1.02	2.17	2.95	5.31	5.84
Uzbekistan	124	23.23	19.31	51.01	47.41	-4.94	42.41	42.45	101.79
Venezuela	125	22.31	56.04	160.23	193.99	-137.39	54.96	59.88	132.15
Vietnam	126	75.36	10.5	0	38.44	-4.69	33.75	13.74	39.23
Yemen	127	15.78	8.08	26.97	18.18	-15.15	2.94	1.81	8.29
FDR of Yugoslavia	128	10.57	33.64	0	10.99	3.43	14.42	34.41	47.3
Former Yugoslavia	129	22.97	62.27	86.54	20.07	12.17	32.53	64.18	92.45
Zambia	130	9.22	3.18	6.53	5.36	0.43	5.79	5.43	2.07
Zimbabwe	131	11.25	7.56	20.15	8.72	1.68	10.44	9.03	17.77

Annexe 2 - Selected Energy Indicators for 1996

Region	N°	TPES/Pop	TPES/GDP	TPES/GDP (PPP)	Elec. Cons./Pop	CO2/TPES	CO2/Pop	CO2/GDP	CO2/GDP (PPP)
		(toe/capit)	(toe/000 90 US\$)	(toe/000 90 US\$ PPP)	(kWh/capit)	(t CO2/to)	(t CO2/capita)	(kg CO2/90 US\$)	(kg CO2/90 US\$ PPP)
World (a)		1.68	0.39	0.31	2229	2.39	4.02	0.93	0.73
OECD		4.6	0.26	0.27	7501	2.41	11.09	0.63	0.66
Middle East		2.04	0.62	0.55	2031	2.79	5.69	1.73	1.52
Former USSR		3.19	1.91	0.93	3846	2.5	8.03	4.8	2.33
Non-OECD Europe (b)		2.01	0.88	0.51	3095	2.63	5.33	2.34	1.33
China		0.9	1.57	0.31	822	2.87	2.59	4.51	0.87
Asia (c)		0.57	0.68	0.24	483	1.94	1.1	1.33	0.48
Latin America		1.09	0.39	0.21	1403	1.98	2.16	0.78	0.41
Africa		0.62	0.89	0.4	492	1.55	0.96	1.38	0.61
Albania	1	0.36	0.59	0	919	1.48	0.54	0.88	0
Algeria	2	0.84	0.37	0.28	581	2.43	2.04	0.9	0.68
Angola	3	0.54	1.55	1.71	66	0.51	0.28	0.79	0.87
Argentina	4	1.67	0.31	0.24	1714	2.24	3.75	0.7	0.54
Armenia	5	0.47	1.16	0.16	1024	1.43	0.68	1.66	0.23
Australia	6	5.5	0.29	0.3	9047	3.01	16.57	0.86	0.91
Austria	7	3.37	0.15	0.19	6520	2.31	7.78	0.35	0.43
Azerbaijan	8	1.56	3.3	1.14	1813	2.65	4.14	8.72	3.03
Bahrain	9	10.94	1.24	1.13	7858	2.41	26.36	2.99	2.73
Bangladesh	10	0.2	0.84	0.18	66	0.86	0.17	0.72	0.16
Belarus	11	2.39	1.09	0.5	2769	2.5	5.98	2.73	1.25
Belgium	12	5.55	0.27	0.31	7528	2.21	12.27	0.59	0.68
Benin	13	0.34	0.8	0.24	48	0.23	0.08	0.18	0.05
Bolivia	14	0.48	0.64	0.21	376	2.06	0.99	1.33	0.44
Bosnia-Herzegovina	15	0.55	1.33	0	597	2.66	1.47	3.56	0
Brazil	16	1.01	0.29	0.19	1716	1.75	1.77	0.51	0.33
Brunei	17	11.54	0.84	1.18	5938	2.42	27.92	2.04	2.85
Bulgaria	18	2.71	1.47	0.79	4391	2.43	6.58	3.56	1.93
Cameroon	19	0.37	0.47	0.23	169	0.23	0.09	0.11	0.05
Canada	20	7.88	0.38	0.42	16413	1.99	15.67	0.76	0.83
Chile	21	1.42	0.44	0.12	1949	2.34	3.33	1.03	0.29
Chinese Taipei	22	3.14	0.28	0.19	5820	2.58	8.11	0.72	0.49
Colombia	23	0.84	0.61	0.17	935	2.04	1.71	1.24	0.34
Congo	24	0.45	0.41	0.23	204	0.78	0.35	0.32	0.18
Dem. Rep. of Congo	25	0.31	2.15	0.46	131	0.35	0.11	0.76	0.16
Costa Rica	26	0.65	0.32	0.13	1295	2	1.3	0.63	0.26
Croatia	27	1.42	0.76	0	2475	2.22	3.14	1.68	0
Cuba	28	1.45	1.05	0.47	1044	1.95	2.82	2.04	0.9
Cyprus	29	2.87	0.3	0.22	3301	3.01	8.64	0.91	0.67
Czech Republic	30	3.92	1.51	0.43	5729	2.96	11.59	4.48	1.28
Denmark	31	4.35	0.16	0.22	6740	3.16	13.74	0.49	0.7
Dominican Republic	32	0.65	0.58	0.21	641	2.17	1.42	1.26	0.45
Ecuador	33	0.73	0.66	0.18	640	2.35	1.72	1.55	0.43
Egypt	34	0.64	0.71	0.21	971	2.49	1.59	1.77	0.53
El Salvador	35	0.7	0.62	0.26	527	1.08	0.75	0.66	0.28
Estonia	36	3.83	1.21	0.61	4459	3.49	13.39	4.22	2.15
Ethiopia	37	0.28	1.96	0.66	22	0.18	0.05	0.36	0.12
Finland	38	6.14	0.23	0.38	13723	2.04	12.54	0.47	0.77
France	39	4.35	0.2	0.24	7069	1.51	6.58	0.3	0.36
Gabon	40	1.4	0.23	0.21	776	1.18	1.66	0.27	0.25
Georgia	41	0.29	0.44	0.23	1048	1.64	0.48	0.73	0.37

Statistique : Analyse de données – La pollution mondiale

Germany	42	4.27	0.19	0.25	6406	2.59	11.05	0.5	0.64
Ghana	43	0.38	0.87	0.21	318	0.65	0.25	0.56	0.14
Gibraltar	44	4.6	0.39	0.38	3034	3.09	14.16	1.21	1.18
Greece	45	2.33	0.26	0.22	3879	3.18	7.4	0.84	0.69
Guatemala	46	0.48	0.54	0.17	342	1.13	0.54	0.61	0.19
Haiti	47	0.27	0.86	0.51	39	0.57	0.15	0.49	0.29
Honduras	48	0.48	0.78	0.65	336	1.21	0.58	0.95	0.79
Hong Kong, China	49	1.93	0.12	0.09	5013	3.1	5.99	0.37	0.27
Hungary	50	2.5	0.79	0.39	3199	2.32	5.8	1.84	0.91
Iceland	51	8.44	0.33	0.45	17353	1.1	9.26	0.37	0.49
India	52	0.48	1.1	0.37	380	1.92	0.91	2.1	0.7
Indonesia	53	0.67	0.74	0.18	298	1.8	1.21	1.33	0.33
Iran	54	1.43	0.56	1.14	1159	2.96	4.24	1.67	3.38
Iraq	55	1.17	1.06	1.32	1357	2.85	3.34	3.02	3.78
Ireland	56	3.3	0.18	0.21	4808	3.05	10.08	0.55	0.63
Israel	57	2.84	0.2	0.2	5308	3.11	8.85	0.64	0.63
Italy	58	2.8	0.14	0.16	4610	2.61	7.31	0.36	0.42
Ivory Coast	59	0.37	0.42	0.25	179	0.82	0.3	0.35	0.2
Jamaica	60	1.46	0.84	0.39	2114	2.63	3.84	2.22	1.03
Japan	61	4.05	0.15	0.2	7749	2.31	9.36	0.36	0.45
Jordan	62	1.04	0.74	0.21	1267	2.91	3.03	2.15	0.61
Kazakhstan	63	2.63	1.94	0.82	3432	3.29	8.67	6.39	2.7
Kenya	64	0.49	1.38	0.37	130	0.46	0.22	0.63	0.17
Korea	65	3.58	0.42	0.3	4730	2.51	8.98	1.05	0.75
DPR of Korea	66	1.07	0	0	250	3.34	3.57	0	0
Kuwait	67	8.72	0.64	0.44	16022	2.8	24.36	1.8	1.22
Kyrgyzstan	68	0.65	1.96	0.77	1557	2.09	1.35	4.08	1.61
Latvia	69	1.68	0.63	0.44	1966	2.22	3.72	1.41	0.97
Lebanon	70	1.16	0.9	0.24	1650	2.89	3.37	2.62	0.69
Libya	71	2.89	0.52	0.55	3518	3.25	9.37	1.68	1.77
Lithuania	72	2.41	0.97	0.63	2656	1.66	4.01	1.6	1.04
Luxembourg	73	8.28	0.25	0.29	14577	2.64	21.87	0.65	0.77
FYR of Macedonia	74	1.73	2.3	0	2868	3.29	5.68	7.57	0
Malaysia	75	2	0.59	0.23	2222	2.56	5.13	1.5	0.59
Malta	76	2.4	0.28	0.2	3861	3.1	7.43	0.88	0.61
Mexico	77	1.46	0.47	0.21	1439	2.42	3.54	1.15	0.51
Moldova	78	1.06	1.21	0.6	1464	2.56	2.72	3.08	1.54
Morocco	79	0.33	0.29	0.09	444	2.95	0.96	0.85	0.27
Mozambique	80	0.43	3.68	2.23	69	0.3	0.13	1.11	0.67
Myanmar	81	0.28	0.37	0.11	59	0.56	0.16	0.21	0.06
Nepal	82	0.32	1.42	0.32	38	0.27	0.08	0.38	0.08
Netherlands	83	4.88	0.23	0.27	5947	2.46	12.01	0.57	0.66
Netherlands Antilles	84	13.17	1.44	1.09	4411	2.78	36.63	4	3.03
New Zealand	85	4.48	0.32	0.3	8865	1.99	8.9	0.63	0.59
Nicaragua	86	0.53	2.09	0.55	305	1.15	0.61	2.4	0.63
Nigeria	87	0.72	2.45	0.69	89	0.58	0.42	1.42	0.4
Norway	88	5.28	0.16	0.23	24033	1.37	7.22	0.22	0.32
Oman	89	2.23	0.33	0.15	3130	2.68	5.98	0.88	0.4
Pakistan	90	0.42	1.06	0.21	328	1.59	0.67	1.68	0.33
Panama	91	0.85	0.32	0.18	1170	1.99	1.7	0.64	0.35
Paraguay	92	0.86	0.69	0.28	951	0.74	0.64	0.51	0.21
Peru	93	0.57	0.31	0.15	602	1.83	1.05	0.56	0.28
Philippines	94	0.53	0.73	0.24	423	1.73	0.91	1.25	0.41
Poland	95	2.81	1.56	0.47	3160	3.37	9.46	5.25	1.59
Portugal	96	1.93	0.26	0.17	3237	2.56	4.94	0.66	0.42
Qatar	97	13.27	1.07	0.98	9375	2.41	31.96	2.57	2.36

Statistique : Analyse de données – La pollution mondiale

Romania	98	2.03	1.26	0.6	2432	2.69	5.45	3.39	1.62
Russia	99	4.17	1.8	0.9	5114	2.46	10.27	4.43	2.2
Saudi Arabia	100	4.75	0.76	0.51	4617	2.7	12.81	2.04	1.37
Senegal	101	0.3	0.4	0.2	116	1.22	0.37	0.49	0.24
Singapore	102	7.84	0.39	0.28	7561	2.74	21.45	1.08	0.78
Slovakia	103	3.27	1.24	0.51	5086	2.26	7.37	2.8	1.14
Slovenia	104	3.1	0.36	0	5215	2.31	7.16	0.84	0
South Africa	105	2.63	0.87	0.55	4774	3.2	8.43	2.77	1.78
Spain	106	2.58	0.19	0.19	4072	2.32	6	0.44	0.45
Sri Lanka	107	0.37	0.65	0.14	205	1.15	0.43	0.75	0.16
Sudan	108	0.4	0.86	1	52	0.39	0.15	0.34	0.39
Sweden	109	5.91	0.22	0.33	15244	1.13	6.67	0.25	0.38
Switzerland	110	3.6	0.11	0.18	7346	1.68	6.04	0.19	0.29
Syria	111	1	0.79	0.33	1172	2.78	2.79	2.18	0.9
Tajikistan	112	0.59	1.94	0.65	2288	1.69	1	3.27	1.09
Tanzania	113	0.45	3.24	0.8	58	0.16	0.07	0.52	0.13
Thailand	114	1.33	0.58	0.2	1343	2.19	2.92	1.28	0.43
Trinidad	115	6.08	1.43	1.11	3154	2.13	12.97	3.05	2.37
Tunisia	116	0.73	0.42	0.15	741	2.31	1.69	0.97	0.34
Turkey	117	1.05	0.35	0.16	1259	2.57	2.69	0.89	0.4
Turkmenistan	118	2.65	3.27	1.22	1342	2.57	6.8	8.39	3.13
Ukraine	119	3.04	1.91	0.98	3194	2.52	7.65	4.82	2.47
United Arab Emirates	120	12.77	0.81	1.13	7795	2.56	32.75	2.07	2.9
United Kingdom	121	3.99	0.22	0.23	5698	2.48	9.91	0.55	0.57
United States	122	8.04	0.34	0.34	13040	2.49	20.05	0.84	0.84
Uruguay	123	0.92	0.28	0.11	1659	1.97	1.82	0.56	0.23
Uzbekistan	124	1.83	2.2	0.83	1828	2.4	4.38	5.27	2
Venezuela	125	2.46	0.98	0.34	2684	2.4	5.92	2.36	0.82
Vietnam	126	0.45	3.22	0	182	1.16	0.52	3.74	0
Yemen	127	0.19	0.36	0.11	115	2.83	0.53	1.03	0.31
FDR of Yugoslavia	128	1.36	0.43	0	3254	3.28	4.47	1.41	0
Former Yugoslavia	129	1.42	0.52	0.38	2794	2.84	4.11	1.48	1.07
Zambia	130	0.63	1.82	0.89	589	0.36	0.23	0.65	0.32
Zimbabwe	131	0.93	1.38	0.52	803	1.7	1.58	2.35	0.88

*Gross production + imports - exports - transmission/distribution losses.

**CO₂ emissions from fuel combustion only. Emissions are calculated using IEA's energy

(a) World excludes Albania, DPR of Korea and Vietnam.

(b) Non-OECD Europe excludes Albania.

(c) Asia excludes China, DPR of Korea and Vietnam.