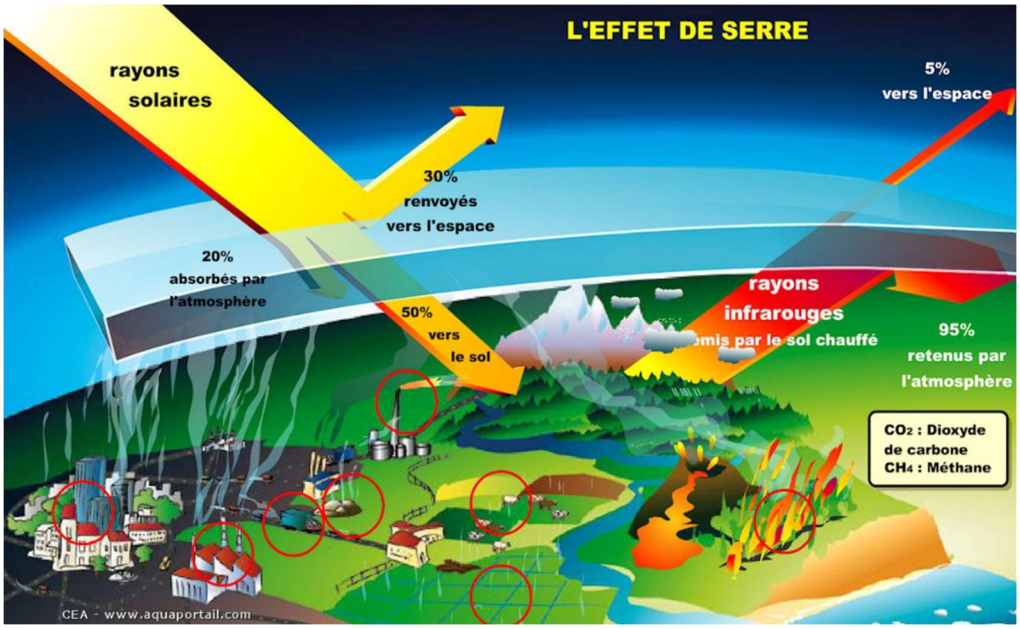
CLASSIFICATION DES PAYS DE L'EUROPE PAR ÉMISSIONS GAZ À EFFET SERRE



Analyse statistique Cluster hiérarchique

Beltrán Vázquez Liniers
bvazquezlin@gmail.com github.com/bvzq

Définition du problème

L'analyse cluster consiste à diviser la population en au moins deux groupes aussi différents que possible mais dont les éléments sont, eux, aussi semblables que possible. Il s'agit donc de maximiser la distance entre groupes et de minimiser la distance à l'intérieur de chaque groupe.

Les données ont été télécharges sur EUROSTAT. On analyse les émissions de gaz à effet serre par pays et par année depuis l'année 1990. L'information est mesurée en tonnes par personne. Le but est de positionner les pays l'un par rapport à l'autre en fonction de leurs émissions.

Deux types de classification seront réalisées. D'une part, on étudiera les émissions pour savoir quels pays sont les plus ou moins polluants. D'autre part, on examinera la réduction de ces émissions pendant les dernières années.

On a donc calculé un ensemble de variables qui mesurent la diférence relative des émissions selon différents intervalles de temps : 1990, 2000, 2010 et 2018. De même on a les valeurs brutes.

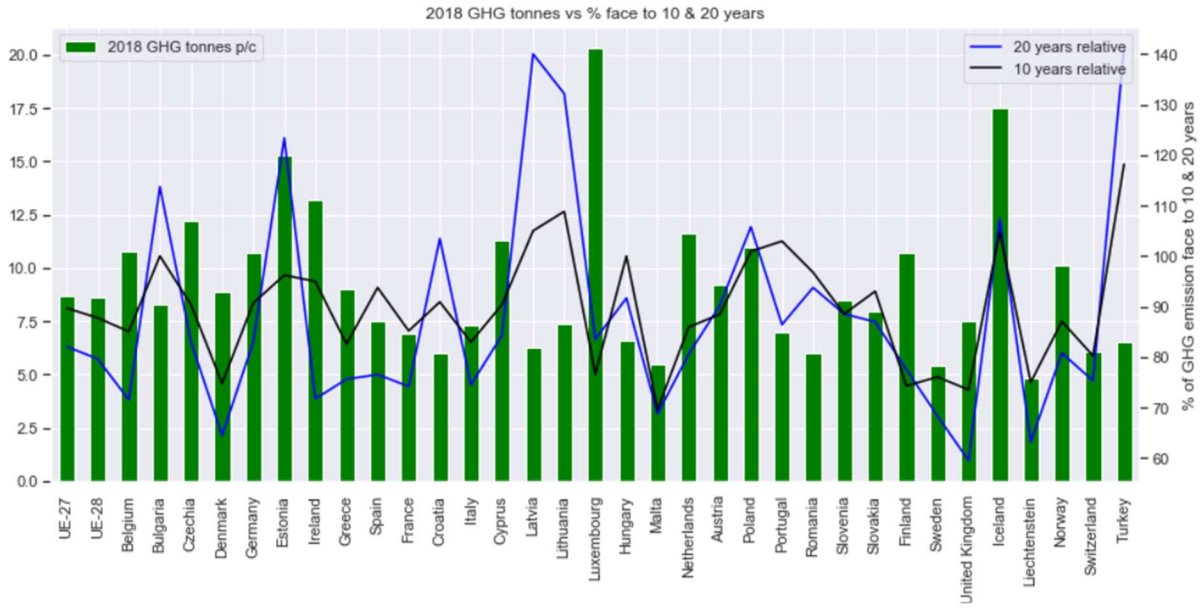
Analyse descriptive

Ensuite, on va réaliser une analyse descriptive des données pour extraire les premières conclusions. Le graphique comporte des barres verticales représentant les émissions cumulées pendant l'année 2018 et deux lignes qui représentent la différence relative par rapport à2010 et 2000.

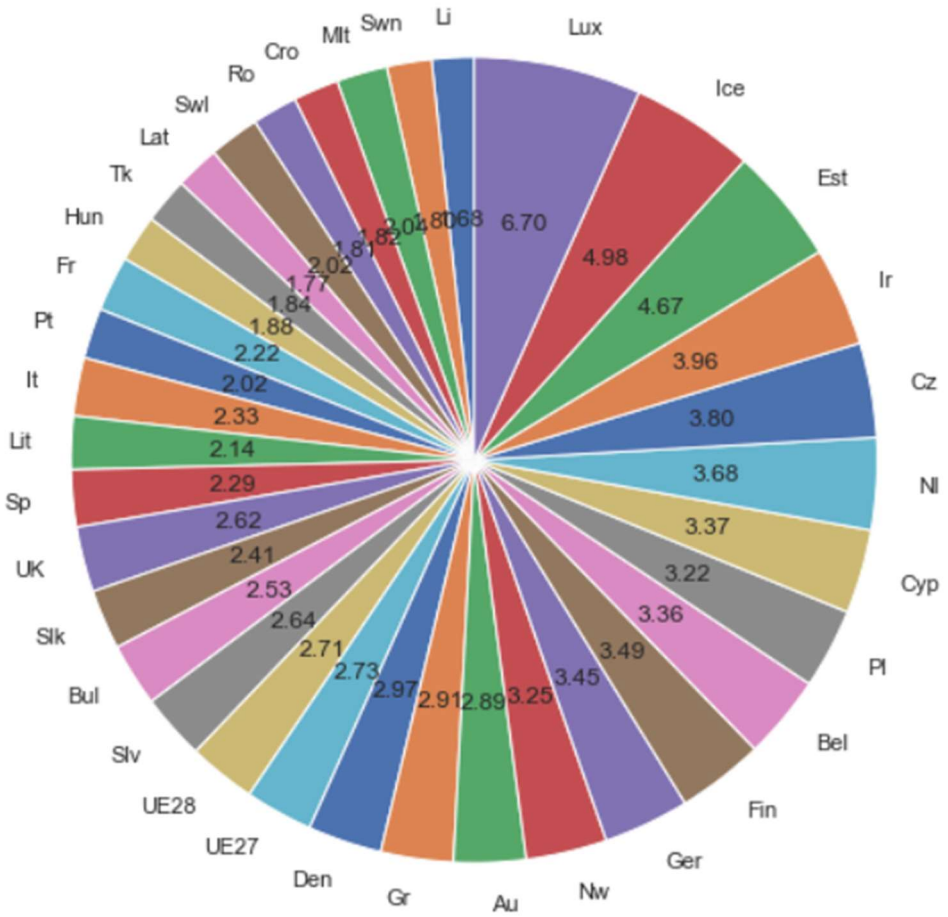
Les deux pays qui émettent le plus en 2018 son le Luxembourg et l'Islande. Viennent ensuite l'Estonie, l'Irlande et la République Tchèque, tous avec plus de 12 tonnes par personne. Le Luxembourg a réduit considérablement ses émissions depuis l'année 2010. En 2018, il a produit 75% de ses émissions de 2010. Par contre, l'Islande les a augmentées de 5%.

A l'opposé se situe la Turquie, qui n'est pas l'un des pays les plus polluants mais qui a néanmoins augmenté ses émissions le plus fortement. Ce pays a produit en 2018 presque 120% de ce qu'il a produit en 2010.

Malte, le Liechtenstein, et la Suède sont les pays les plus respectueux sur le plan environnemental de l'Union Européenne. Ils ont émis environ 5 tonnes par personne pendant l'année 2018, ce qui représente moins de 80% de leurs émissions en 2010.



Enfin, il est intéressant de savoir de qu'émettent les pays les plus polluants par rapport aux autres. Le graphique à secteurs ci-dessous montre que le Luxembourg représente 6.7% des émissions. Ceci trois fois supérieur à la France et 2.5 fois supérieure à la moyenne de l'UE.



Nombre de Clusters

À ce stade de l'analyse, la question la plus fréquente est de savoir combien de groupes on doit créer pour optimiser la classification. En théorie, plus le nombre de groupes est élevé, moins il y aura de dispersion au sein des groupes. Ceci est positif mais s'il y a beaucoup de divisions, alors l'interprétation des clusters est trop compliquée et les différences deviennent trop peu significatives.

Donc, pour atteindre le nombre approprié de groups, on se base sur l'indicateur SSE (Standard Squared Error). On analyse la valeur de cette statistique pour les différents modèles, qu'on a créé en ajoutent un nouveau cluster pour chacun. Quand la réduction du SSE n'est pas significative, par rapport au modèle précèdent avec un cluster moins, on arrête la recherche

On a testé différents modèles avec différent variables et on recommande d'utiliser 2, 3 et 4 clusters, selon l'information que l'on veut introduire.

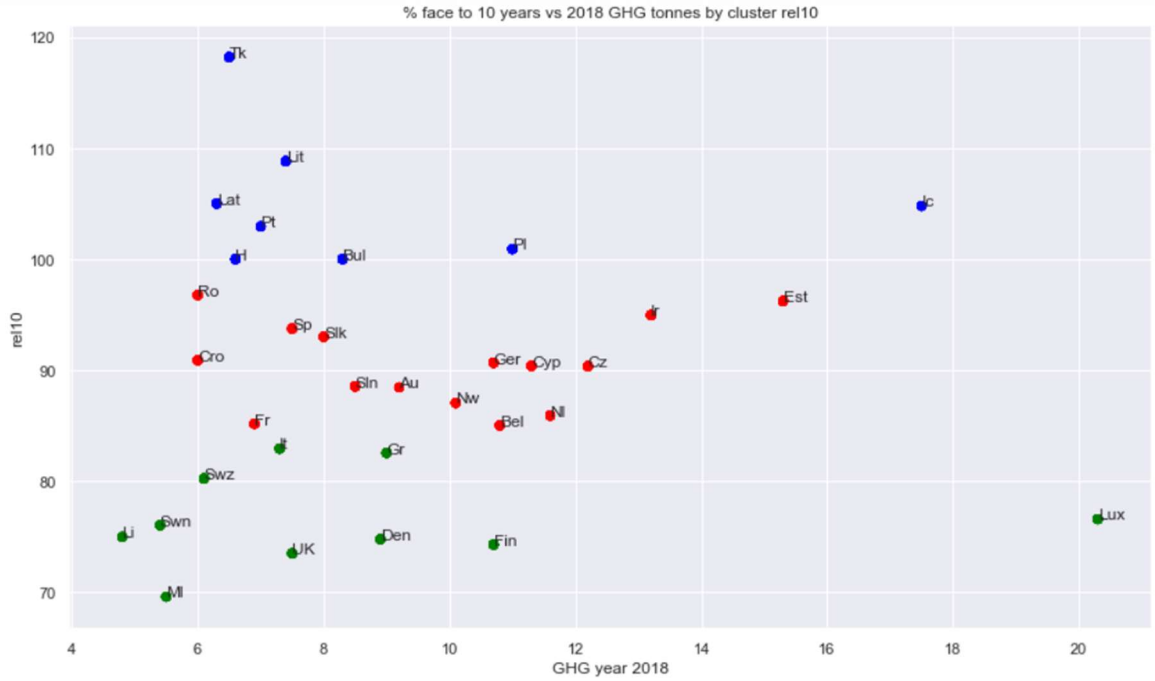
Avant de commencer la clusterisation, il faut expliquer comment les groupes sont créés, il y a deux façons de les construire : hiérarchique et non hiérarchique.

Si l'on applique la première technique, chaque observation est considérée comme un cluster. Les deux clusters les plus proches s'agglomèrent en un seul. Le processus continue jusqu'à ce que l'on obtienne un seul grand groupe. L'analyste doit décider quand arrêter, avec combien de cluster.

Par contre, si l'on applique la méthode non hiérarchique, l'analyste assigne autant de centroïdes qu'il veut faire de groupes et où il veut les situer. À partir d'ici, chaque observation est ajoutée au centroïde le plus proche. Chaque fois que s'ajoute une nouvelle observation à un cluster, les centroïde est recalculé.

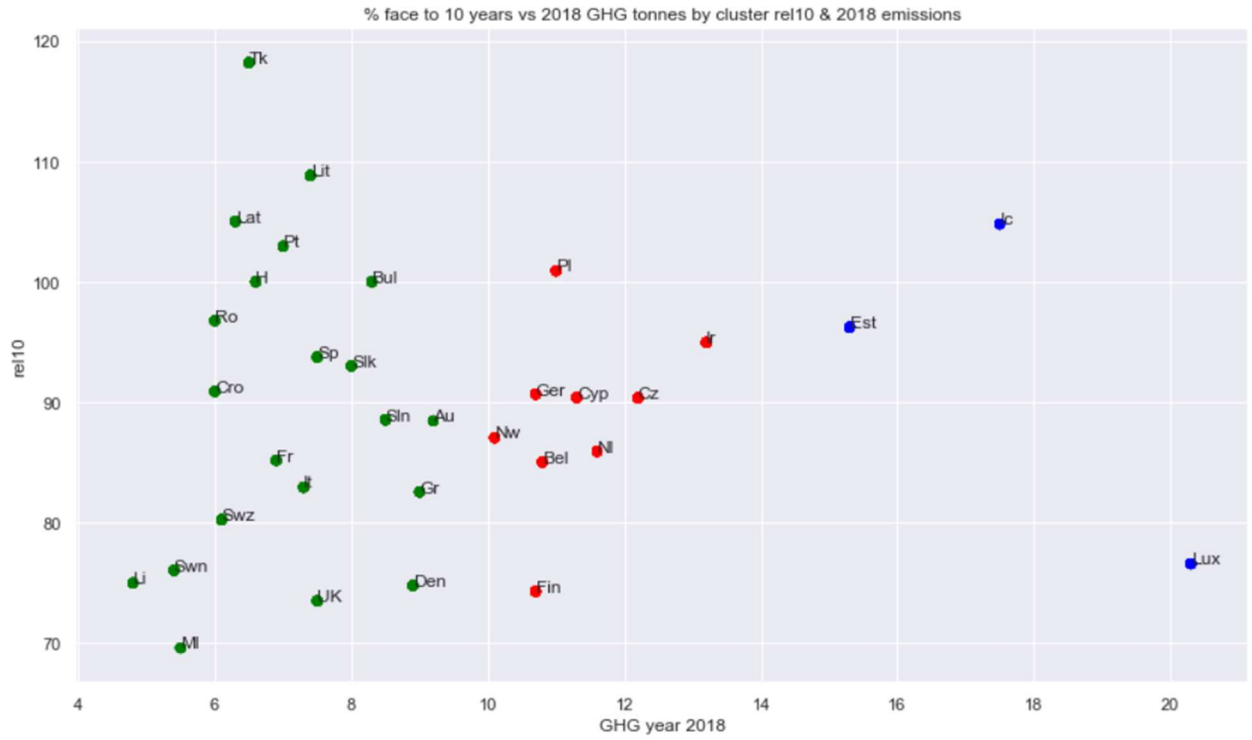
Aucune des deux méthodes n'est préférable à l'autre. Cela dépend des types de données et de la manière dont on veut les travailler. Nous établissons une classification des pays selon la méthode hiérarchique.

Dans le graphique ci-dessous, on a la première classification. Elle est basée sur le pourcentage des émissions en 2018 par rapport à 2010 (variable rel10). Cette information est représentée sur l'axe vertical. Sur l'axe horizontal figurent les émissions brutes en 2018



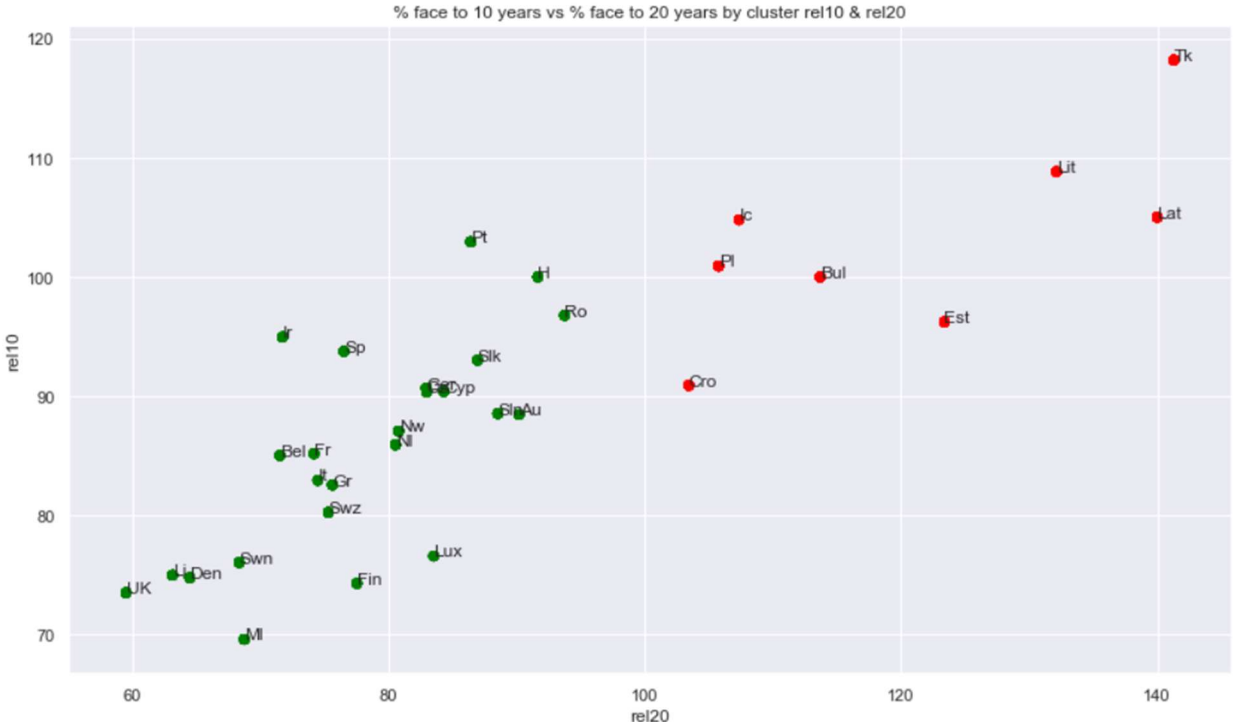
Les pays en vert sont ceux qui ont réduit le plus leurs émissions. Ils ont émis en moyenne 76% par rapport au chiffre de 2010. De l'autre côté figurent, en bleu, les pays qui n'ont pas du tout réduit leurs émissions et les ont, par contre, augmentées. En rouge, apparaissent les pays qui réduisent lentement leurs émissions, dont la France, l'Espagne et l'Allemagne, avec une réduction proche de 10%.

Le graphique suivant montre une clusterisation basée sur les émissions brutes dans l'année 2018.



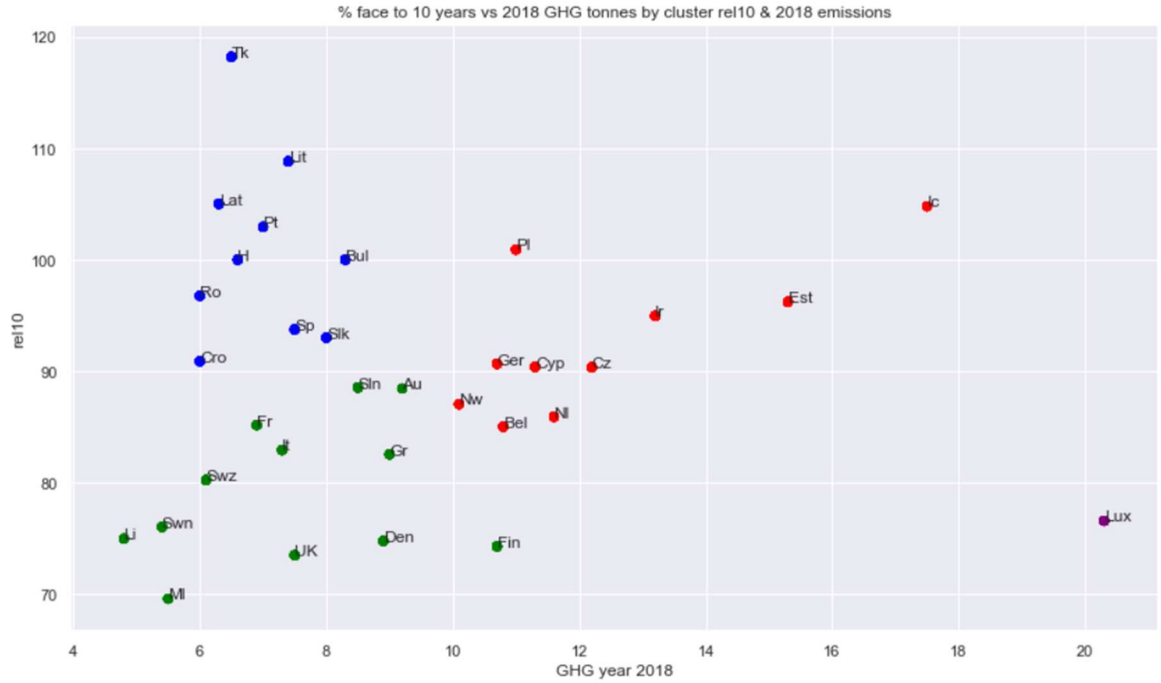
Analyse Cluster avec deux variables

Pour la prochaine étape, on ajoute une deuxième variable, rel20, qui explique le pourcentage des émissions de 2018 par rapport à 2000. Ce que l'on voit ci-dessous, ce sont, à gauche, les pays qui ont réduit leurs émissions pendent les deux dernières décades et, à droite, ceux qui les augmentent. L'axe horizontal représente ici la variable rel20



Quatre pays sont vraiment à la limite entre les deux clusters : le Portugal, la Hongrie, la Romanie et la Croatie. Les trois premiers ont réduit leurs émissions par rapport à l'année 2000 mais ils sont stables par rapport a 2010. De l'autre côté, la Croatie, elle, n'a pas réduit ses émissions dans les 20 dernières années mais bien dans les 10 dernières, peut-être, parce qu'elle les a augmentées entre 2000 et 2010, et réduites entre 2010 et 2018.

La dernière classification que nous allons établir introduit deux variables avec différent échelles. Cela veut dire que les valeurs qu'elles prennent sont très différentes, donc on doit les standardiser. Pour ce faire, on soustrait la moyenne et on divise par la déviation standard. Ces deux variables sont la différence relative au cours des 10 dernières années (rel10) et les émissions brutes en 2018.



Dans ces cas, on obtient 4 clusters. L'un d'eux n'a qu'une observation. C'est le Luxembourg. Cela s'explique par le fait qui il est trop loin des autres donc qu'il ne réussit pas à les rejoindre. Les autres 3 groupes sont bien différenciés. Ceux qui émettent peu et réduisent ses émissions figurent en vert. Ceux qui ne les réduisent pas mais ne polluaient beaucoup figurent en bleu. Enfin, ceux qui produisent des gaz à effet de serre et ne les réduisent pas, figurent en rouge.

Conclusions

D'abord, on note de grandes différences entre les pays de l'UE, en ce qui concerne les émissions de gaz à effet de serre. Bien, à cause des tonnes de ces gaz émis, bien par la réduction de ces émissions dans les dernières décades.

Les nations le plus avancées sur ce point sont : Malte, la Suède, le Royaume Unit et la Suisse. On peut aussi discerner des pays en nette progression, comme la Norvège, l'Allemagne et la Belgique. Enfin, des pays qui n'émettent pas trop mais qui n'ont pas une grande marge de réduction : l'Espagne, le Portugal et la Slovaquie.

Les deux pays les plus remarquables sont le Luxembourg et l'Islande. Ils sont assez riches et à l'avant-garde sur les plans sociaux et économiques, cependant ils émettent beaucoup et, en plus, l'Islande ne réduit pas ses émissions.

Une possible explication à ce phénomène en Luxembourg est qu'il y a beaucoup de travailleurs transfrontaliers. Ils polluent et, à la fin de la journée, ils rentrent dans leurs pays d'origine. Ils ne comptent donc pas comme habitants mais ils contribuent aux émissions de gaz.

En ce qui concerne à l'Islande, il est peut-être difficile de réduire ses émissions en raison de son système de production géothermique et sa baisse densité de population.