

Projet Analyse de données

Malleret Antoine, Morille Adrien

2025-04-02

1. Introduction

Présentation du jeu de données : “Market Cap”

Ce jeu de données contient des informations sur les plus grandes entreprises mondiales en termes de capitalisation boursière. Il inclut diverses variables financières et ESG (Environnement, Social et Gouvernance) qui permettent d’analyser leur performance économique et leur engagement en matière de durabilité. Nous avons réaliser une analyse exploratoire des données pour mieux comprendre les relations entre ces variables que vous pouvez retrouver en annexe.

Structure du jeu de données

- **Nombre de variables** : 45
-
- **Types de variables** :
 - **Quantitatives** : marketCap (capitalisation boursière), trailingPE (PER historique), forwardPE (PER futur), priceToBook (prix/valeur comptable), totalEsg (score ESG), etc.
 - **Qualitatives** : sector (secteur d’activité), country (pays), peerGroup (groupe de pairs), esgPerformance (performance ESG), etc. Vous pouvez retrouver la description de chacune des variables dans l’annexe.

Problématiques soulevées

1. Analyse du jeu de données et prédiction

- Quels sont les secteurs les plus représentés parmi les entreprises ayant la plus grande capitalisation boursière ?
- Existe-t-il des modèles adaptés pour prédire les bénéfices future (forwardPE) et si une entreprise est sur/sous évalué (PriceToBook) ?
- Pouvons nous construire des groupes représentant bien l’ensemble de nos données ?

2. Performances ESG et responsabilité sociale

- Pouvons nous trouver une corrélation entre les secteurs et les scores ESG ?
- Les entreprises ayant une forte capitalisation boursière (nos individus) ont-elles tendance à avoir de meilleures performances ESG ?

2. Classification non supervisée

ACP : Analyse en Composantes Principales

En appliquant une **Analyse en Composantes Principales** (ACP) sur l'ensemble de nos données quantitatives, nous obtenons **31 axes** principaux. Parmi eux, les 15 premiers axes permettent d'expliquer 80 % de la variance totale.

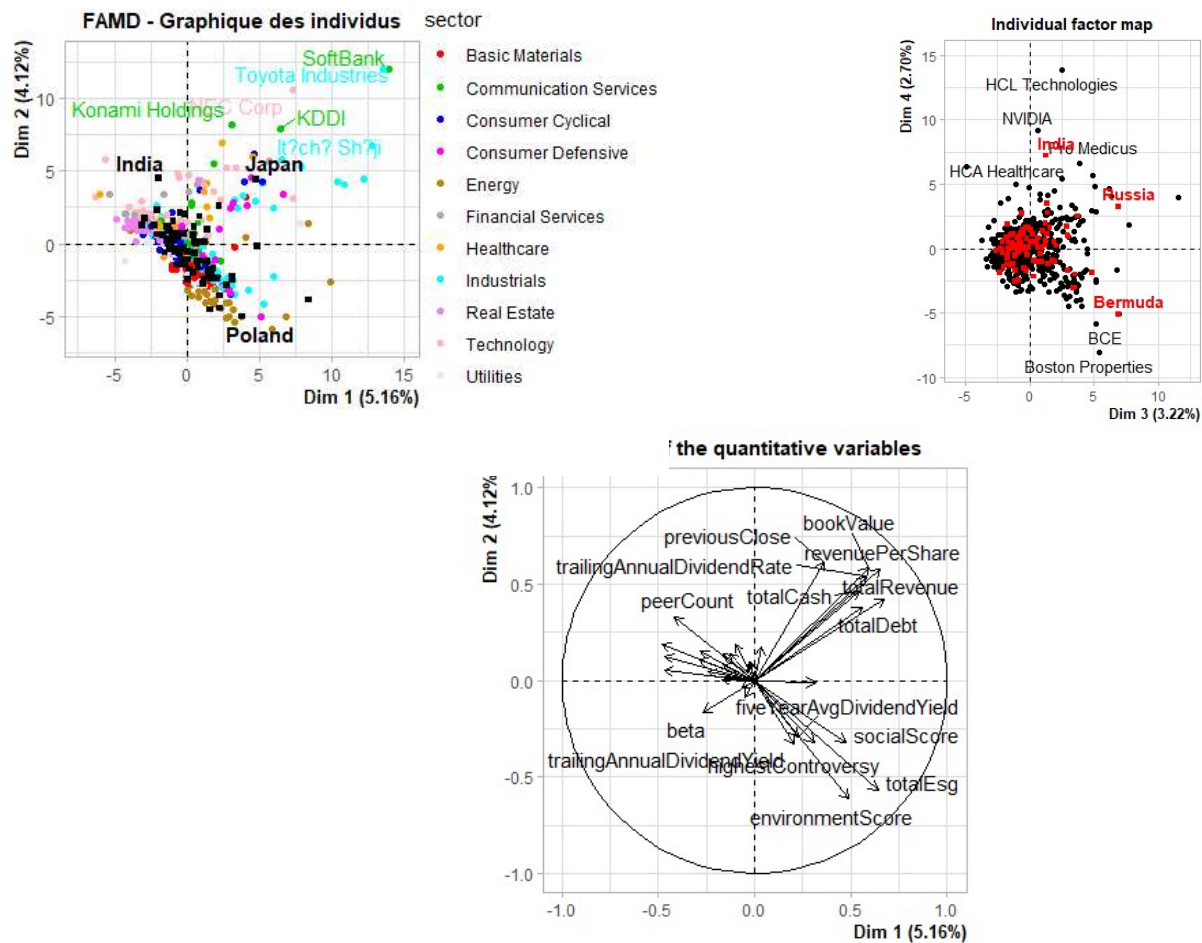
Observations Si nous nous concentrons sur les 5 premiers axes, ils expliquent à eux seuls 40 % de la variance, répartie de la manière suivante :

- Axe 1 : *priceToBook, enterpriseToRevenue, enterpriseToEbitda*
- Axe 2 : *totalRevenue, totalCash, grossMargins, ebitdaMargins*
- Axe 3 : *totalEsg, socialScore, ebitdaMargins*
- Axe 4 : *currentRatio, quickRatio*
- Axe 5 : *trailingAnnualDividendYield, trailingAnnualDividendRate*

Avec une telle quantité de données et un grand nombre d'axes, la visualisation des résultats devient complexe. Pour mieux comprendre les structures sous-jacentes, nous allons donc réaliser une ACP ciblée sur les données **ESG**. Cette approche nous permettra d'identifier les principaux facteurs qui influencent spécifiquement les critères environnementaux, sociaux et de gouvernance.

De plus, l'exclusion des données qualitatives restreint l'analyse alors que nous avons des données qualitatives qui peuvent être très intéressantes à analyser. Nous allons donc réaliser une **Analyse Factorielle des Données Mixtes** (FAMD) pour inclure ces variables qualitatives dans notre analyse.

FAMD : Analyse Factorielle des Données Mixtes

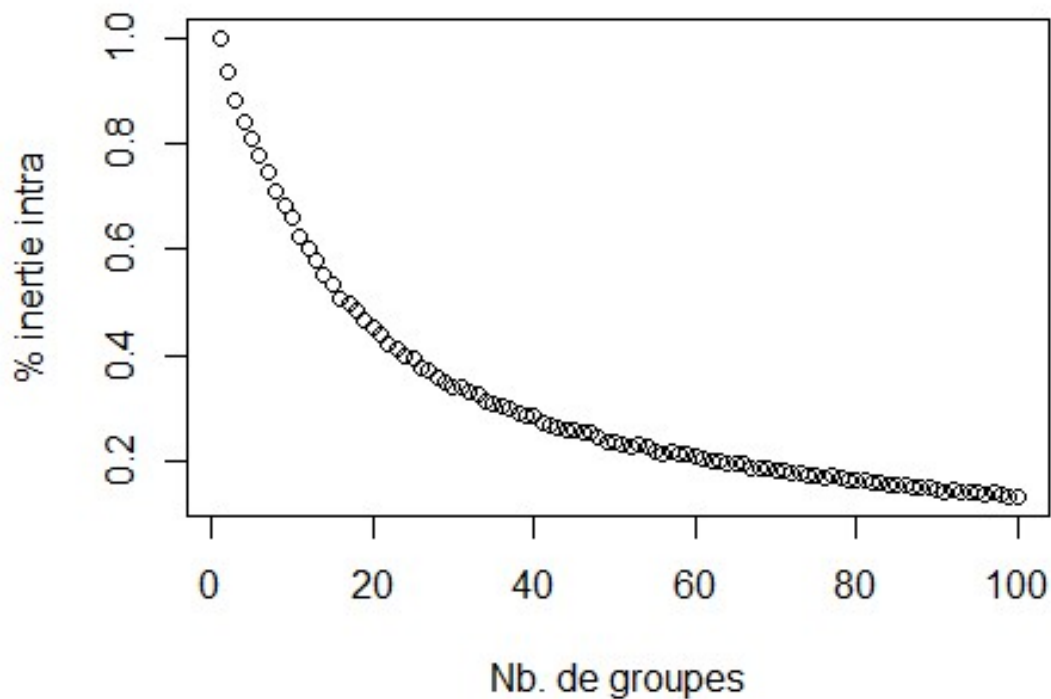


Nous observons que les axes 1 et 2 expliquent environ 10 % de la variance totale, tandis que les axes 3 et 4 expliquent environ 6 % supplémentaires. Cela signifie que les quatre premiers axes capturent environ 16% de la variance totale des données ce qui est très peu. Il nous faudrait 59 axes pour expliquer 80% de la variance.

Clustering

Clustering avec k-means sur les premières composantes On applique k-means pour regrouper les entreprises.

Pour notre clustering K-means, nous choisissons d'utiliser les 30 premières composantes principales issues de la FAMD. Celles-ci permettent de capturer environ 55 % de la variance, tout en réduisant suffisamment la dimension pour obtenir un clustering plus stable et pertinent. Cela permet de conserver l'information essentielle tout en limitant les effets de bruit liés à la haute dimension.

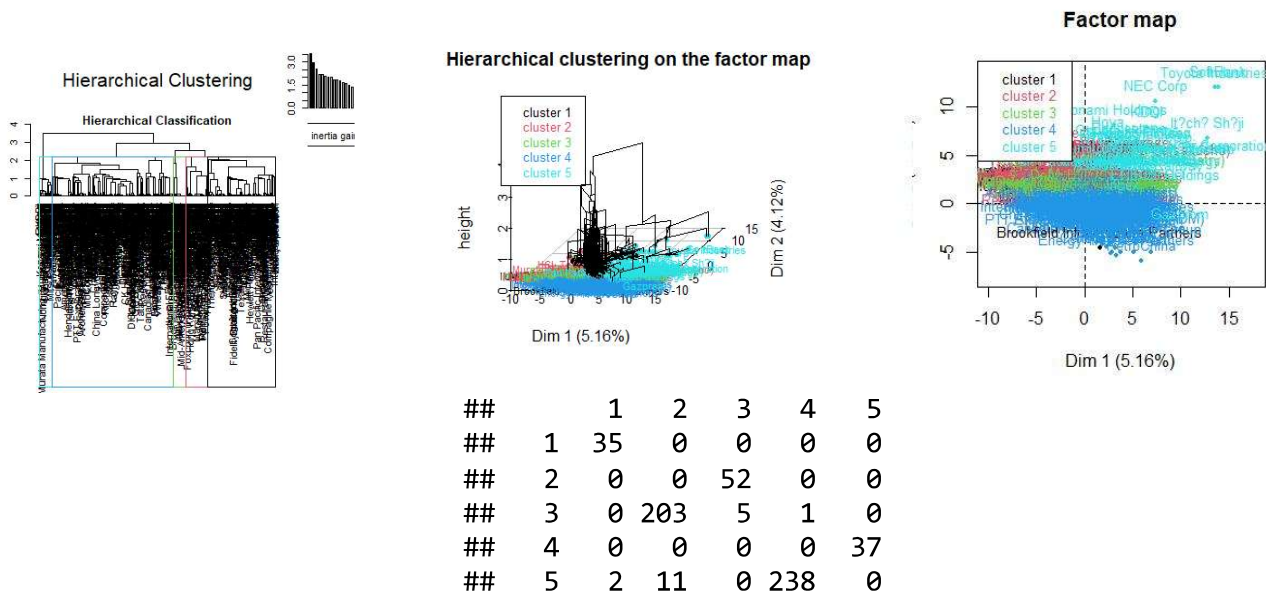


D'après les résultat du graphe de l'inertie intra, nous devrions choisir un nombre de groupes entre 20 et 30 mais pour des raisons de visualisation nous allons en choisir 5. Ce qui nous donne la répartition suivante :

```
## 1  2  3  4  5
## 35 52 37 209 251
```

CAH : Classification Hiérarchique

Avec HCPC On applique la Classification Ascendante Hiérarchique (CAH) pour vérifier si les groupes sont cohérents.



On peut observer que les groupes obtenus par la CAH sont très similaires à ceux obtenus par K-means. A cause des approximations que nous avons dû faire pour en arriver là, il est difficile de faire une analyse pertinentes des différent groupes. Nous pouvons toutefois observer les moyennes des variables par cluster pour essayer de les interpréter:

Cluster 1 – “Responsables et performants” Rentabilité très élevée : marges brutes et EBITDA très importantes.

- Entreprise bien gérée : ratios financiers sains, peu de dette.
- Bon score ESG (~14) : faible impact environnemental, social et de gouvernance.

Des entreprises efficaces sur le plan financier.

Cluster 2 – “Solides mais perfectibles socialement” - Bonne solidité financière : forte trésorerie, bon rendement.

- Rentabilité modérée mais stable.
- ESG correct (~18)

Des entreprises solides avec un profil ESG moyen.

Cluster 3 – “Entreprise à fort impact” - Très grosses capitalisations et CA.

- Ratios de valorisation parfois extrêmes (PE, Price-to-Book).
- Scores ESG très mauvais (~25-27) sur toutes les dimensions : environnement, social, gouvernance.

Des géants mondiaux performants mais très mal notés sur l'impact durable.

Cluster 4 – “Surévalués, peu efficaces” - Ratios de valorisation élevés (PE, Price-to-Book) → potentiellement surévalués.

- Rentabilité opérationnelle plus faible, marges réduites.
- Scores ESG modérés (~16.5).

Profil d'entreprises moins performantes, mais avec un profil ESG moyen.

Cluster 5 – “Rentables mais à fort risque social et environnemental” Rentabilité correcte à bonne, marges bien présentes.

- Ratios financiers stables, bon niveau de cash.
- Score ESG élevé (~26) → fort impact environnemental et social, surtout environnement (>10).

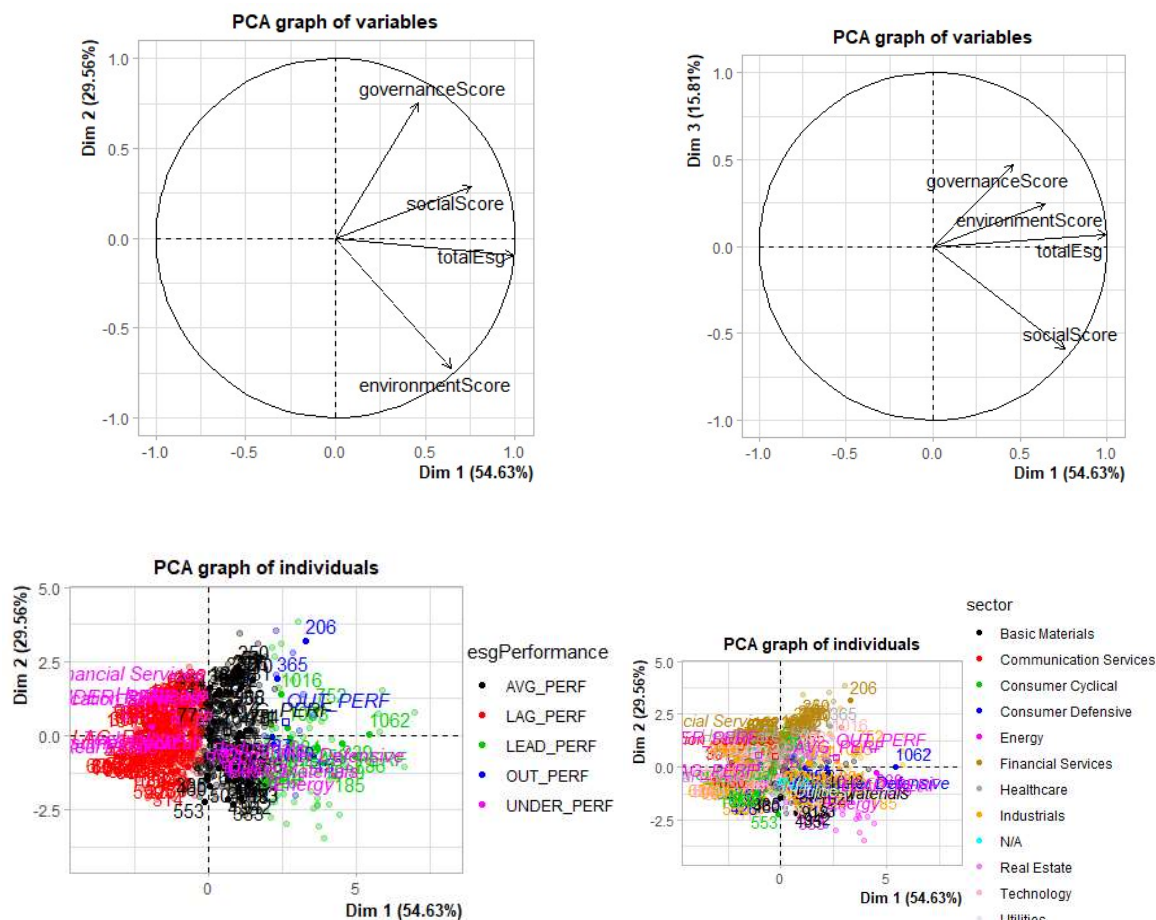
Des entreprises qui semblent performantes, mais très mal notés RSE, en particulier sur l'écologie.

Conclusion

Nous avons réalisé une analyse en composantes principales (ACP) et une analyse factorielle des données mixtes (FAMD) sur les données financières de notre jeu de données. Ces analyses nous ont permis d'identifier des groupes d'entreprises ayant des caractéristiques similaires en termes de performance financière.

3. Analyse des scores ESG

ACP sur les données ESG



Dans cette analyse, plus un score ESG est élevé, plus l'entreprise a un impact négatif sur l'environnement, la société ou sa gouvernance. L'ACP nous permet d'identifier les principaux axes qui structurent ces données et de mieux comprendre comment les entreprises se différencient en fonction de leurs scores ESG.

Analyse des axes principaux L'ACP révèle que trois axes principaux expliquent presque 100% de la variance des données, ce qui signifie qu'ils capturent l'essentiel des différences entre les entreprises en matière d'impact ESG.

Premier axe : un indicateur global de l'impact ESG Le premier axe est principalement influencé par le score ESG total et le score social. Plus une entreprise a une valeur élevée sur cet axe, plus elle accumule des impacts négatifs sur l'ensemble des critères ESG, en particulier sur le plan social. Ce premier axe permet donc de classer globalement les entreprises selon leur empreinte ESG globale.

Deuxième axe : opposition Environnement / Gouvernance Le second axe met en évidence une opposition entre le score environnemental et le score de gouvernance. Cela signifie que certaines entreprises ont un impact plus marqué sur l'environnement (pollution, émissions carbone, mauvaise gestion des ressources), tandis que d'autres se distinguent par des faiblesses en gouvernance (manque de transparence, risques liés à la gestion interne). Cet axe souligne donc deux profils distincts : les entreprises aux impacts négatifs environnementaux forts et celles aux impacts négatifs liés à leur gouvernance.

Troisième axe : différenciation entre Social et Gouvernance Le troisième axe met en évidence une différence entre les entreprises ayant un impact négatif principalement sur le plan social (mauvaise gestion des employés, manque d'inclusion, impact sociétal négatif) et celles qui présentent des lacunes en gouvernance. Il permet d'affiner l'analyse en montrant quelles entreprises ont des problèmes davantage liés à la gestion humaine et sociale, et lesquelles ont des défauts structurels de gouvernance.

Conclusion et interprétation des résultats Cette ACP met en lumière trois grandes tendances :

Le premier axe classe les entreprises selon leur impact ESG global.

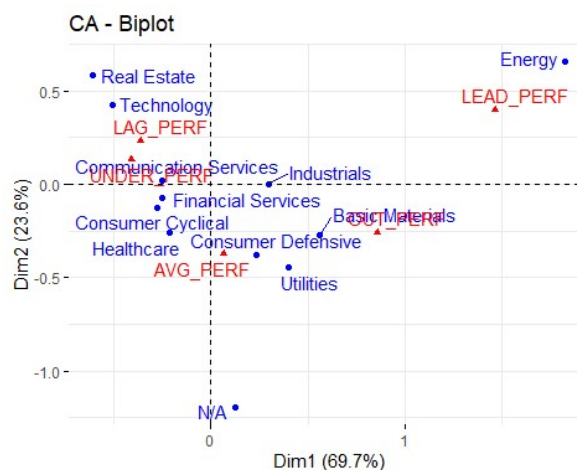
Le second axe différencie les entreprises ayant un impact négatif sur l'environnement ou sur la gouvernance.

Le troisième axe affine l'analyse en opposant l'impact social et l'impact de gouvernance.

Ces résultats montrent que l'ESG ne se limite pas à un seul critère et que les entreprises peuvent avoir des impacts négatifs dans des domaines différents. Cette segmentation permet d'identifier les entreprises ayant les pires pratiques globales, mais aussi celles qui ont des faiblesses spécifiques, que ce soit en matière d'environnement, de gouvernance ou de responsabilité sociale.

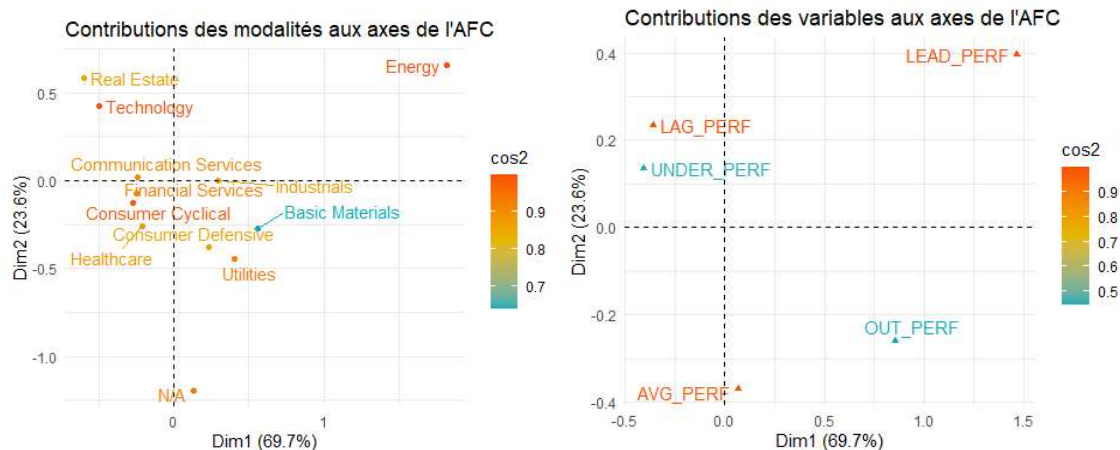
AFC : Analyse Factorielle des Correspondances

Ici nous allons effectuer une **Analyse Factorielle des Correspondances (AFC)** sur les variables sector et esgPerformance de notre jeu de données.



Résumé des variables $OUT_PERF > LEAD_PERF > AVG_PERF > LAG_PERF > UNDER_PERF$ Et comme pour l'ACP, plus la valeur est élevée, plus la l'entreprise impacte négativement. Donc les "meilleurs" secteurs sont ceux qui ont la valeur la plus basse (LAG/UNDER). On remarque également que les deux premiers axes expliquent plus de 90 % de la variance totale. Ce qui est plutôt logique étant donné la relation entre nos variables est assez linéaire.

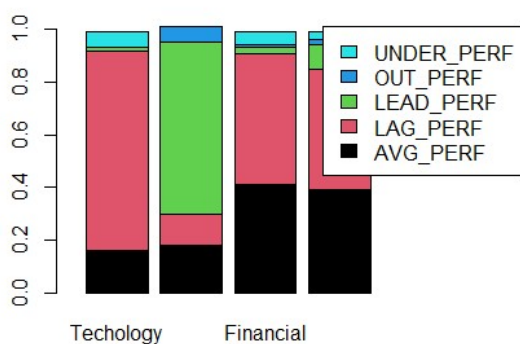
Explications des résultats de l'AFC:



Tout d'abord on voit que toutes les modalités sont bien représentées sur le graphique, ce qui est un bon signe. (le pire étant Basic Materials avec un cos2 légèrement inférieur à 0.7.)

Ensuite, on remarque que les secteurs Technology et Real Estate sont très proches l'un de l'autre avec un score ESG très faible. Les secteurs Energy et Industrials ont quant à eux un score ESG très élevé.

Profils des secteurs

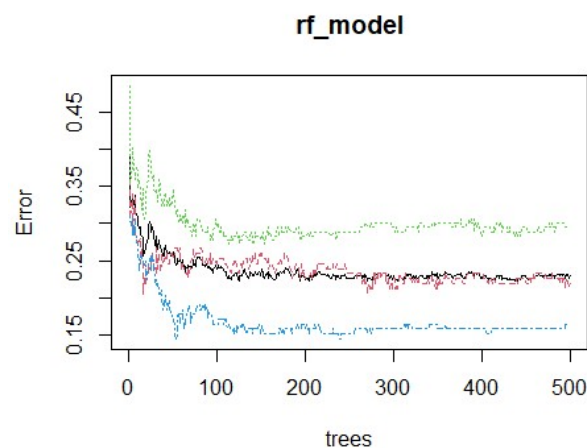


On observe que le secteur Technology est celui qui a le plus de chance d'avoir un score ESG faible. Le secteur Energy est celui qui a le plus de chance d'avoir un score ESG élevé. Le secteur Financial Services est celui qui a le plus de chance d'avoir un score ESG moyen.

4. Classification supervisée

Prédiction de la variable forwardPE avec Random Forest

La variable forwardPE (Forward Price-to-Earnings) est un ratio financier qui mesure le rapport entre le prix actuel de l'action et les bénéfices futurs estimés d'une entreprise. C'est donc une variable intéressante à prédire. Pour se faire, nous avons converti forwardPE en forwardPE_class prenant des valeurs qualitatives ("élevé"(33%), "moyen"(33%), "bas"(33%)) et nous avons découpé les données aléatoirement en données d'entraînement (80%) et données de test (20%) et nous allons effectuer une Random Forest qui s'applique bien dans notre cas car pour prédire forwardPE_class avec nos données mixtes (qualitatives et quantitatives). De part la construction de forwardPE_class nous avons aussi la certitude que nos données sont bien réparties dans les classes.



On remarque que les erreurs sont constantes pour toutes les classes à partir d'environ 100 arbres dans notre random forest. On choisit de prendre le nombre d'arbres égal à 100.

Analyse du graphe et de l'importance

```
##
## pred    Bas Moyen Élevé
## Bas     29     4     0
## Moyen   6     35    10
## Élevé    2      6    25

## Area under the curve: 0.8796

## [1] "accuracy = 0.760683760683761"
```

Nous remarquons pour le choix de 100 arbres, une matrice de confusion avec des erreurs réparties de manière homogène entre les classes. Nous avons aussi une précision de 76% cela veut dire que 76% de nos individus sont classés dans la bonne classe. Nous avons une aire sous la courbe ROC de 0,87 ce qui signifie que le taux de "vrais positifs" est de 87% et le taux de "faux positifs" est de 20%. Cela signifie que notre modèle est assez bon pour prédire la classe de forwardPE_class.

Prédiction de la variable forwardPE avec CART (Classification and Regression Trees)

Nous avons essayé d'utiliser une autre méthode pour la prédiction de forwardPE_class, la méthode CART. Cependant, l'exécution du code n'arrivait jamais à bout, nous avons donc abandonné cette méthode mais nous avons mis le code en annexe.

Prédiction de la variable priceToBook

Nous allons maintenant prédire la variable priceToBook qui est un ratio financier qui mesure le rapport de son prix de marché (valeur de marché) sur la valeur comptable d'une entreprise (valeur réelle). C'est un indicateur important permettant de déceler si une entreprise est sous ou sur évaluée, en particulier, un priceToBook inférieur à 1 indique que l'entreprise est sous-évaluée. Nous allons donc essayer de prédire si une entreprise est sous-évaluée (priceToBook=0) ou surévaluée (priceToBook=1). De ce fait, nous modifions la variable priceToBook de façon à ce que si le ratio est strictement inférieur à 1 alors nous mettons la valeur à zéro et si elle est supérieure à 1 nous mettons 1. Ensuite, nous convertissons cette ligne en facteurs. Nous allons effectuer un maximum de méthodes de prédictions et nous finirons par leur comparaisons afin de déterminer laquelle est la mieux adaptée.

```
## [1] "table de priceToBook"

##
##    0    1
## 132 926

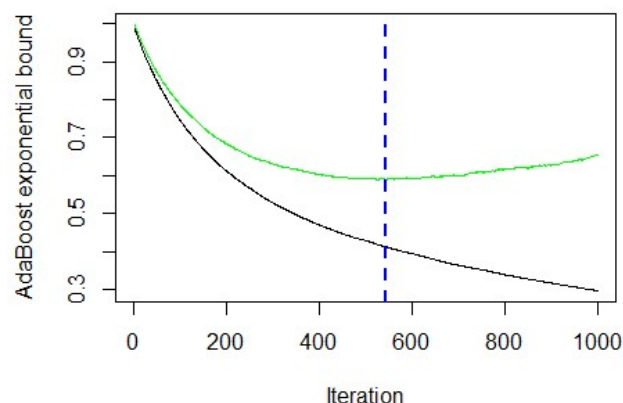
## [1] "table de nos données d'entrainement après rééchantillonnages"

##
##    0    1
## 108 144
```

Nous remarquons que nos données initiales ne sont pas bien réparties, beaucoup d'individus sont dans la classe 1 et peu dans la classe 0, nous avons donc utilisée une méthode de rééchantillonnage SMOTE afin de rééquilibrer le jeu de données d'entraineemnt.

Prédiction de la variable priceToBook avec Adaboost

Nous allons maintenant prédire la variable priceToBook avec la méthode Adaboost. Nous avons conservé uniquement les données quantitatives.



En affichant l'erreur de validation croisée, nous obtenons un nombre d'arbres optimal que nous pouvons maintenant utiliser pour la prédiction

```
##
## class_adaboost    0    1
```

```
##           0 12 11
##           1  1 91

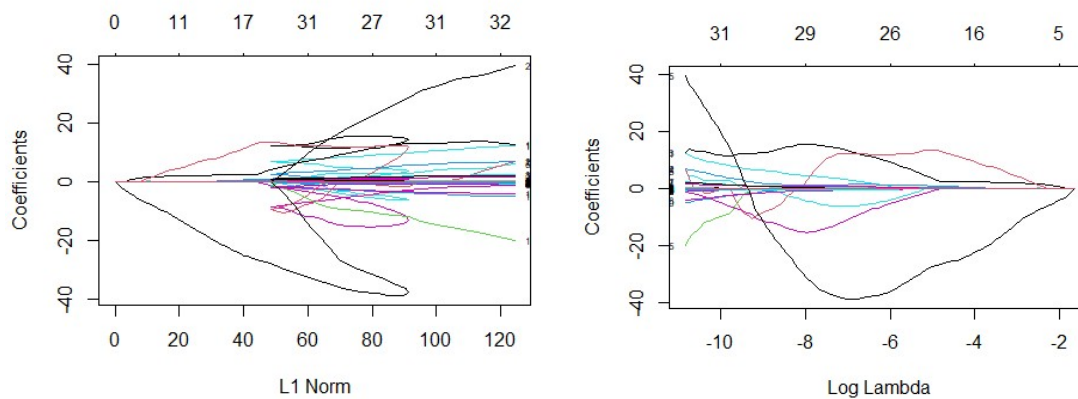
## [1] "accuracy = 0.895652173913044"

## Area under the curve: 0.9306
```

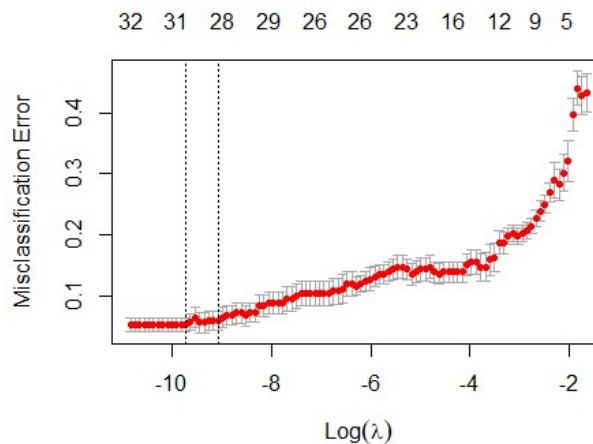
Nous remarquons que la matrice de confusion est bien répartie entre les classes et que nous avons une précision de 88% ainsi qu'une très bonne aire sous la courbe ROC de 0,93 ce qui semble être de bons résultats.

Prédiction de la variable priceToBook avec Lasso

Nous allons maintenant effectuer une régression logistique avec Lasso.



Nous allons choisir lambda par cross validation



Nous allons maintenant choisir le lambda optimal et faire la prédiction sur le jeu de test.

```
##           Y_test
## class_logit_lasso 0  1
##                   0  8  8
##                   1  5 94

## [1] accuracy = 0.8869565 ## Area under the curve: 0.7821
```

Nous avons une matrice de confusion bien répartie entre les classes et une précision de 88% ainsi qu'une aire sous la courbe ROC de 0,83 ce qui est relativement bon

Prédiction de la variable priceToBook avec LDA et QDA

Nous allons maintenant faire une LDA et une QDA. A priori, nous n'avons pas d'information sur l'homoscédasticité de nos données, nous allons donc faire les deux méthodes et les comparer pour voir à posteriori si cela est pertinent.

```
## pred_lda_class  0  1
##                0  6 22
##                1  7 80

## [1] accuracy_lda = 0.7478261

## Area under the curve: 0.7821
```

```
## pred_qda_class  0  1
##                0  2  0
##                1 11 102

## [1] accuracy_qda = 0.9043478

## Area under the curve: 0.5769
```

Nous remarquons des résultats étonnant, malgré des précisions proches, l'aire sous la courbe ROC de la QDA n'est que de 0,52 contre 0,85 pour celle de la lda.

Prédiction de la variable priceToBook avec Random Forest

Nous allons maintenant effectuer une Random Forest

```
## pred_rf_priceToBook  0  1
##                    0 11  2
##                    1  2 100

## [1] 0.9652174

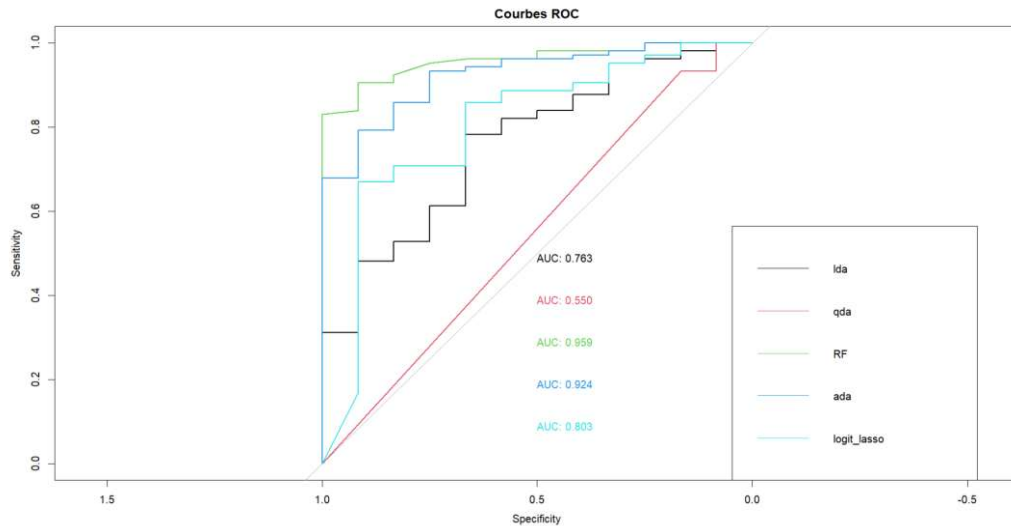
## Area under the curve: 0.9265
```

Nous avons une matrice de confusion bien répartie entre les classes et une précision de 95% ainsi qu'une aire sous la courbe ROC de 0,95. Ce sont des résultats excellents .

Comparaison des modèles et conclusion

Enfin, nous allons comparer tout nos modèles et conclure.

```
##          lda          qda          RF  adaboost  logit_lasso
## accuracy 0.7478261 0.9043478 0.9652174 0.8956522  0.8869565
## AUC      0.7820513 0.5769231 0.9264706 0.9306184  0.7820513
```



Nous observons ainsi que la méthode de Random forest est la meilleure que cela soit en terme d'aire sous la courbe ROC ou de précisions. Cependant, il faudrait faire plus de test pour chacune des méthodes et faire une moyenne pour obtenir véritablement une conclusion définitive.

5. Conclusion

Ce projet a permis d'explorer les apports des méthodes d'apprentissage statistique, tant supervisées que non supervisées, dans le contexte de l'analyse financière enrichie par des données ESG.

Nous avons d'abord mené une phase exploratoire des données (en annexe) puis une phase d'analyse non supervisée à travers une **ACP** (Analyse en Composantes Principales) pour les variables quantitatives, et une **FAMD** (Analyse Factorielle des Données Mixtes) pour prendre en compte à la fois les variables qualitatives et quantitatives présentes dans notre base. La FAMD s'est révélée plus adaptée à notre jeu de données hétérogène qui doit tenir compte de variables qualitatives clef dans une analyse financière (secteurs, région...), et a permis une meilleure visualisation de la structure des individus. Cependant, ces méthodes n'ont pas permis de réduire significativement la dimension des données sans perte d'information, ce qui limite leur usage pour la simplification du modèle.

Par ailleurs, les variables **ESG** ont été intégrées aux analyses, afin d'analyser leurs liens avec les secteurs d'activités et les liens qui les relient.

L'analyse supervisée a porté sur la **prédiction du ratio forwardPE** (bénéfices futurs estimés) via une méthode de Random Forest prenant en compte l'ensemble des données. Plusieurs méthodes ont été testées pour prédire la variable **priceToBook** (estimation de l'évaluation d'une entreprise), telles que la **régression logistique Lasso**, **LDA**, **QDA**, **Adaboost** et enfin une **Random Forest**. Pour faire face au déséquilibre des classes, un rééchantillonnage par **SMOTE** a été utilisé, ce qui a permis d'améliorer les performances de certains modèles. La Random Forest est la méthode la plus pertinente pour la prédiction de la variable priceToBook avec une précision de 95% et une aire sous la courbe ROC de 0,95.

En somme, ce projet montre l'intérêt des outils de classifications appliqués à la finance d'entreprise et à l'analyse ESG. Malgré les limitations rencontrées sur la réduction de dimension, les modèles développés ouvrent la voie à des prédictions fiables, tout en soulignant l'importance d'un traitement adapté des données mixtes pour une analyse pertinente.

Annexe : Explication des variables du jeu de données

Ce jeu de données contient des informations sur les principales entreprises mondiales en termes de capitalisation boursière, incluant des indicateurs financiers, des ratios de performance et des scores ESG.

Variables du jeu de données

Informations générales

- **Rank** : Classement de l'entreprise selon sa capitalisation boursière.
- **Name** : Nom de l'entreprise.
- **Symbol** : Symbole boursier de l'entreprise.
- **country** : Pays d'origine de l'entreprise.
- **sector** : Secteur d'activité (ex : Technologie, Finance, Santé, etc.).
- **Index** : Indice boursier auquel appartient l'entreprise (ex : S&P 500, Nasdaq).
- **Continent** : Continent d'implantation de l'entreprise.

Données de Marché

- **marketCap** : Capitalisation boursière totale de l'entreprise.
- **previousClose** : Prix de clôture de l'action lors de la dernière session.
- **trailingPE** : Ratio prix/bénéfice basé sur les bénéfices passés.
- **forwardPE** : Ratio prix/bénéfice basé sur les bénéfices estimés.
- **priceToBook** : Ratio Prix/Valeur comptable.
- **beta** : Mesure de la volatilité du titre par rapport au marché.

Performance Financière

- **profitMargins** : Marge bénéficiaire nette.
- **enterpriseToRevenue** : Ratio Valeur d'Entreprise/Chiffre d'Affaires.
- **enterpriseToEbitda** : Ratio Valeur d'Entreprise/EBITDA.
- **totalRevenue** : Chiffre d'affaires total.
- **revenuePerShare** : Chiffre d'affaires par action.
- **revenueGrowth** : Croissance du chiffre d'affaires.
- **grossMargins** : Marge brute.
- **ebitdaMargins** : Marge EBITDA.
- **operatingMargins** : Marge opérationnelle.

Rentabilité et Gestion Financière

- **returnOnAssets (ROA)** : Rentabilité des actifs.
- **returnOnEquity (ROE)** : Rentabilité des fonds propres.
- **debtToEquity** : Ratio d'endettement.
- **currentRatio** : Ratio de liquidité courante.
- **quickRatio** : Ratio de liquidité immédiate.
- **totalCash** : Montant total des liquidités.
- **totalDebt** : Montant total des dettes.

- **bookValue** : Valeur comptable de l'entreprise.

Dividendes et Rendement Actionnarial

- **fiveYearAvgDividendYield** : Rendement moyen des dividendes sur 5 ans.
- **payoutRatio** : Pourcentage des bénéfices distribués en dividendes.
- **trailingAnnualDividendYield** : Rendement annuel des dividendes.
- **trailingAnnualDividendRate** : Montant du dividende annuel par action.

Données ESG (Environnement, Social, Gouvernance)

- **totalEsg** : Score ESG total.
- **environmentScore** : Score environnemental.
- **socialScore** : Score social.
- **governanceScore** : Score de gouvernance.
- **ratingYear** : Année de notation ESG.
- **ratingMonth** : Mois de notation ESG.
- **highestControversy** : Niveau de controverse maximal associé à l'entreprise.
- **peerCount** : Nombre d'entreprises comparables pour l'évaluation ESG.
- **esgPerformance** : Catégorie de performance ESG (ex : LEAD_PERF, AVG_PERF, LAG_PERF).
- **peerGroup** : Groupe de comparaison sectoriel pour l'ESG.

Analyse exploratoire des données

Comparaisons sectorielles et géographiques

- Quels sont les pays qui dominent le classement des entreprises à forte capitalisation ?
- Comment les secteurs d'activité diffèrent-ils en termes de ratios financiers et d'engagement ESG ?

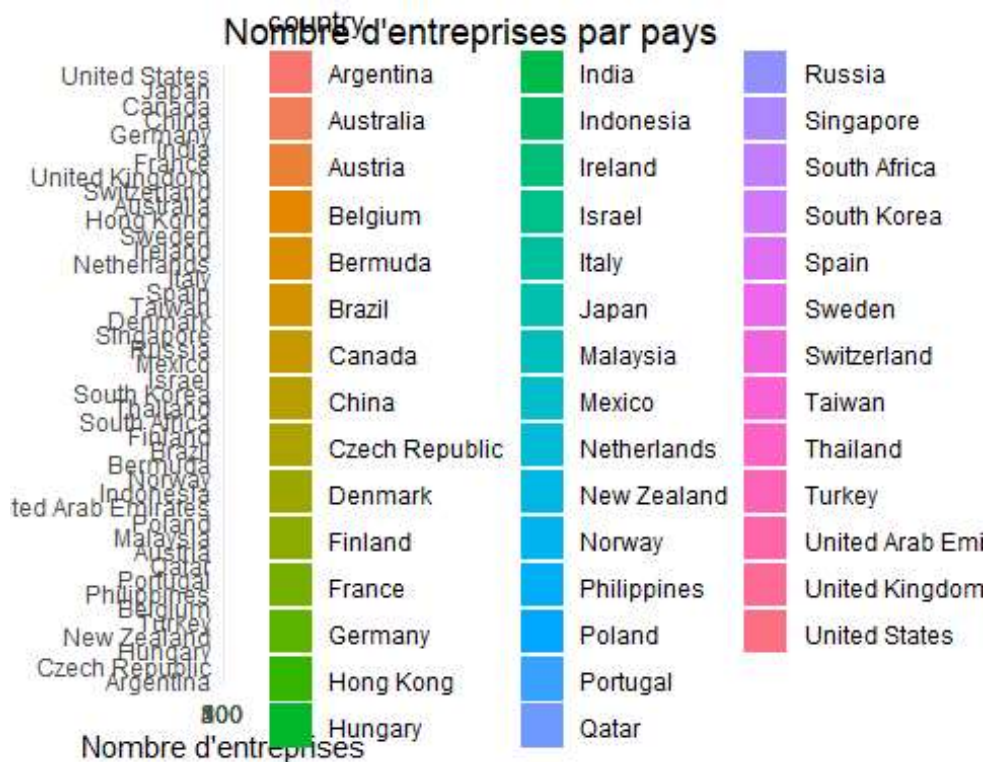
Analyse prédictive et tendances

- Peut-on prédire la performance future d'une entreprise en fonction de ses indicateurs financiers et ESG ?
- Existe-t-il des anomalies statistiques qui mériteraient une étude plus approfondie ? ### 0.
Répartition des entreprises par pays

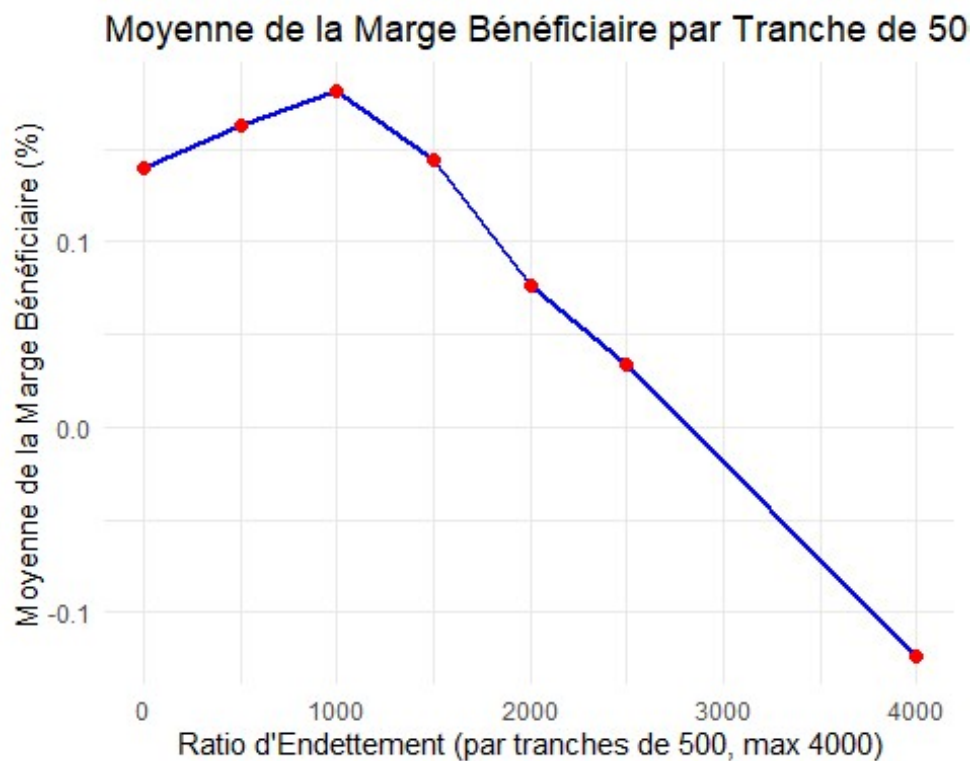
Un barplot est un bon moyen de visualiser la répartition des entreprises par pays. Cela permet de voir rapidement quels pays ont le plus d'entreprises dans le jeu de données.

```
## # A tibble: 43 × 2
##   country      Count
##   <chr>         <int>
## 1 United States    501
## 2 Japan             84
## 3 Canada           51
## 4 China            38
## 5 Germany          37
```

```
## 6 India 34
## 7 France 33
## 8 United Kingdom 30
## 9 Switzerland 29
## 10 Australia 23
## # i 33 more rows
```

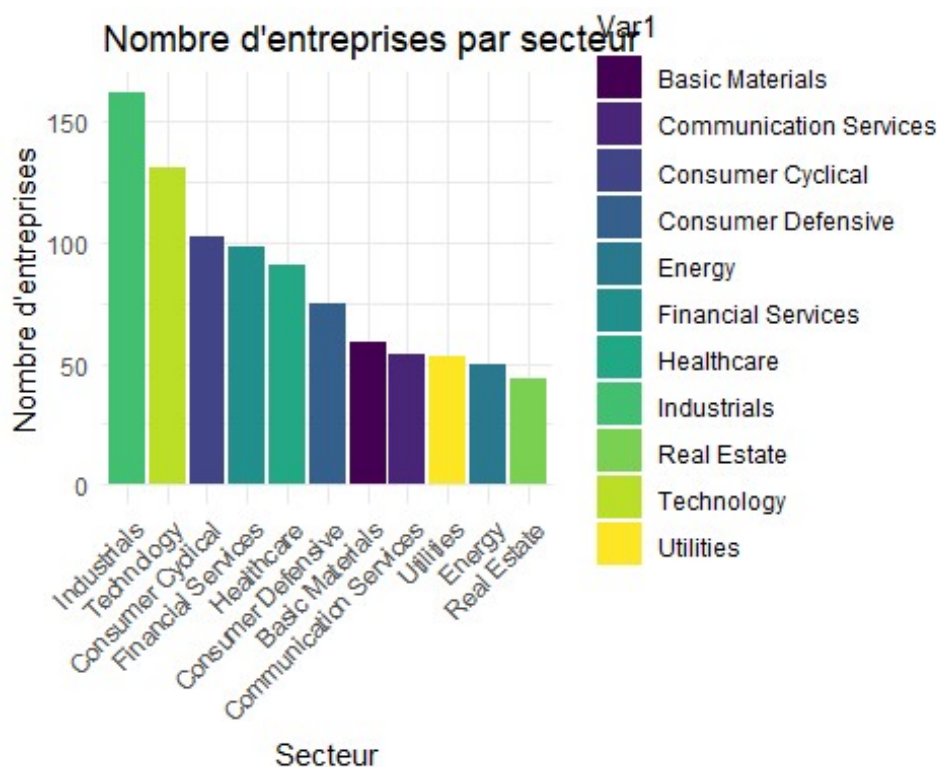


1. Corrélation entre la Marge Bénéficiaire et le Ratio d'Endettement



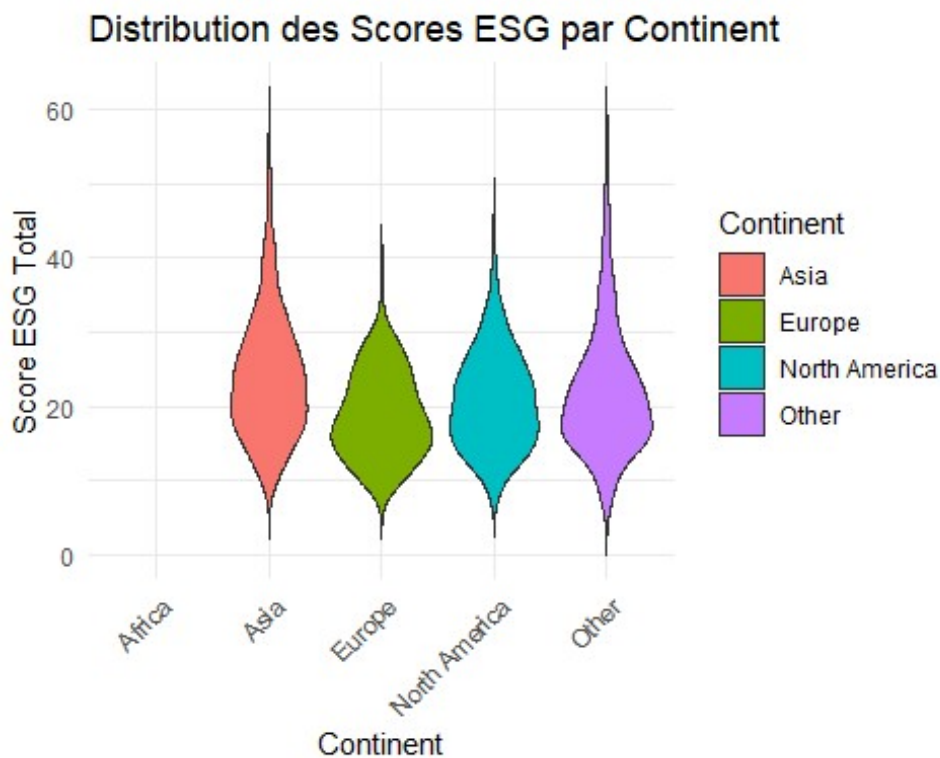
- **Faible endettement (0 - 1000) :** L'endettement modéré semble bénéfique pour la rentabilité, car les entreprises peuvent utiliser les fonds empruntés pour financer leur croissance.
- **Endettement excessif (> 1000) :** Un endettement trop élevé réduit la rentabilité, probablement à cause des coûts d'intérêts accrus et d'un risque financier plus important.
- **Ratio > 3000 - 4000 :** La rentabilité devient négative, ce qui suggère que ces entreprises ne parviennent plus à couvrir leurs charges financières. ### 2. Distribution de la Capitalisation Boursière par Secteur Un boxplot est un excellent moyen de visualiser la distribution de la capitalisation boursière par secteur, ce qui permet de comparer la taille des entreprises dans différents domaines d'activité. Commençons tout d'abord par compter le nombre d'entreprises par secteur pour voir la répartition.

2.Répartition du Nombre d'Entreprises par Secteur et de leur Capitalisation Boursière



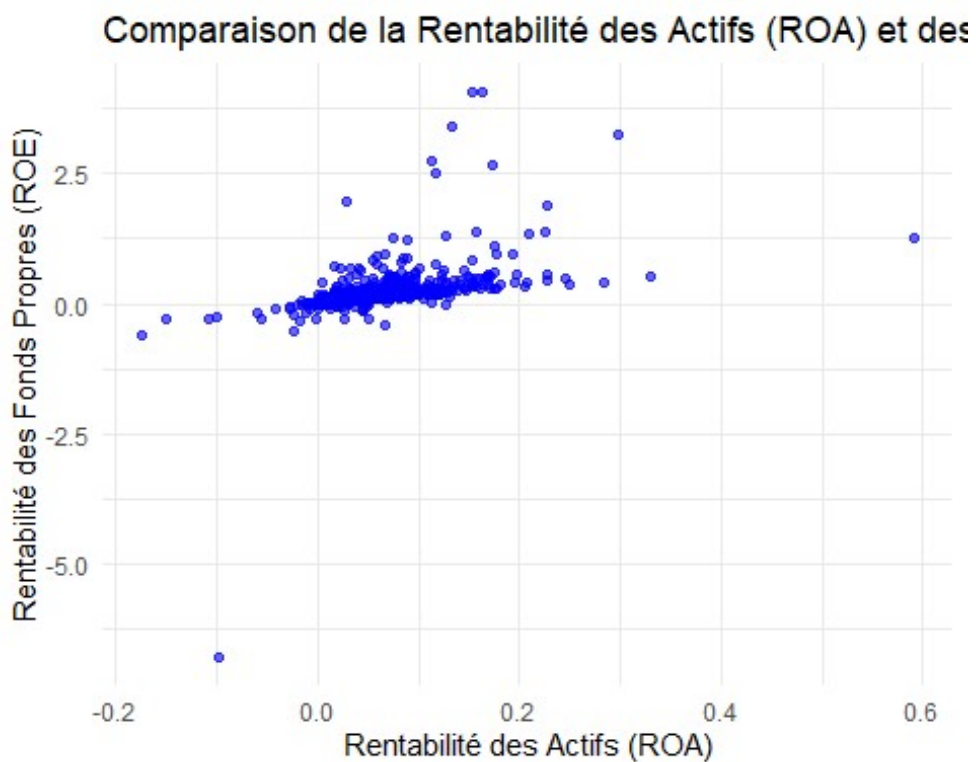
- Industrials et Technology dominant avec le plus grand nombre d'entreprises.
 - Consumer Cyclical, Financial Services et Healthcare occupent une position intermédiaire.
 - Real Estate, Energy et Utilities sont les moins représentés.
- ◆ Cela reflète l'importance des secteurs industriels et technologiques dans l'économie actuelle.

3. Scores ESG par Continent



Attention : Plus le score est élevé, plus l'entreprise impacte négativement l'environnement. - **Asie :** Score ESG le plus élevé, indiquant un impact environnemental plus important. - **Europe et Amérique du Nord :** Score ESG le plus bas, indiquant un impact environnemental relativement faible.

4. Comparaison de la Rentabilité (ROA vs ROE)



- **Corrélation positive** : Les entreprises avec un ROA élevé ont également un ROE élevé, indiquant une gestion efficace des actifs et des fonds propres.