

# Projet\_Analyse\_de\_données

moi

2025-04-02

## Projet : Analyse de données financières

### 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.

#### 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 de la capitalisation boursière

- Quels sont les secteurs les plus représentés parmi les entreprises ayant la plus grande capitalisation boursière ?
- Existe-t-il une corrélation entre la capitalisation et d’autres variables financières (PER, price-to-book, etc.) ?

##### 2. Performances ESG et responsabilité sociale

- Quel est l’impact des scores ESG sur la valorisation des entreprises ?
- Les entreprises ayant une forte capitalisation boursière ont-elles tendance à avoir de meilleures performances ESG ?

##### 3. 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 ?

#### 4. 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 ?

## Warning: le package 'ggplot2' a été compilé avec la version R 4.3.3

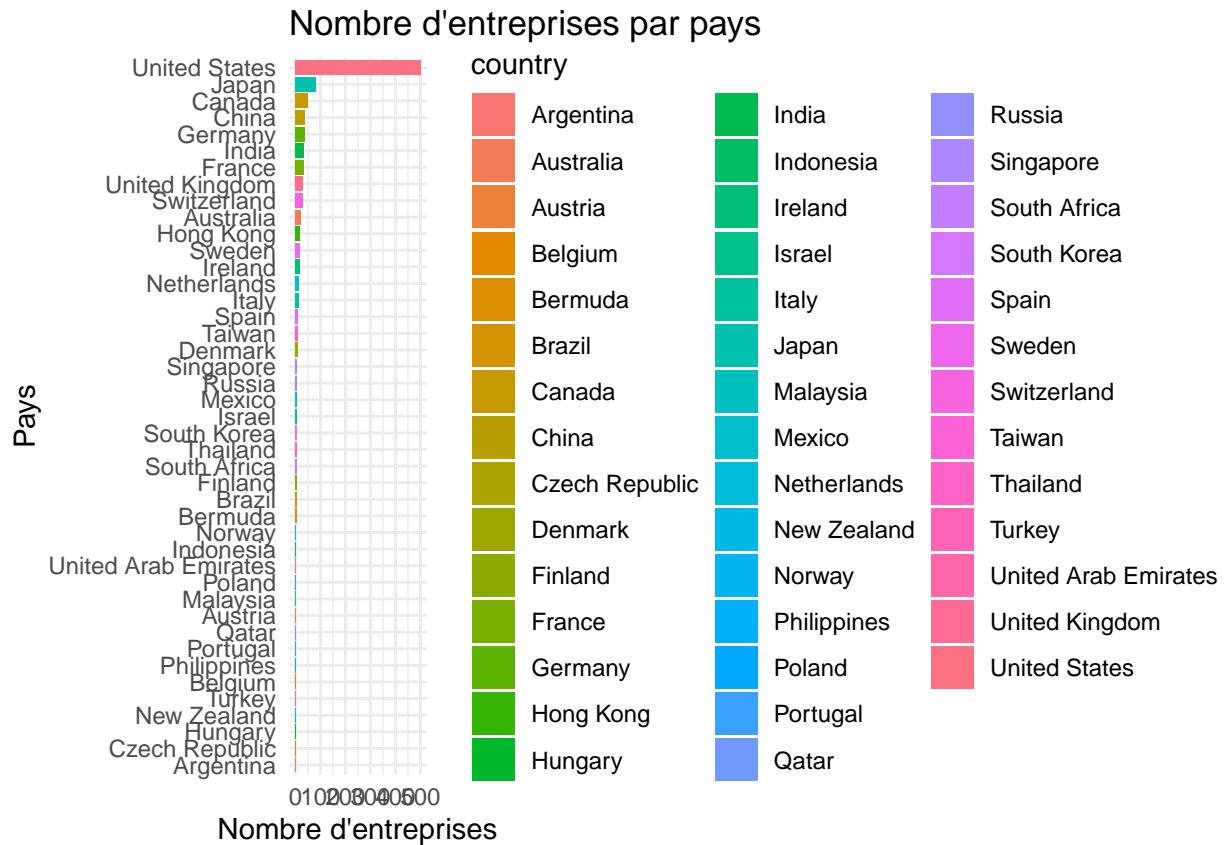
```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Netoyage des données.

Conversion des données en numérique.

### Analyse exploratoire des données

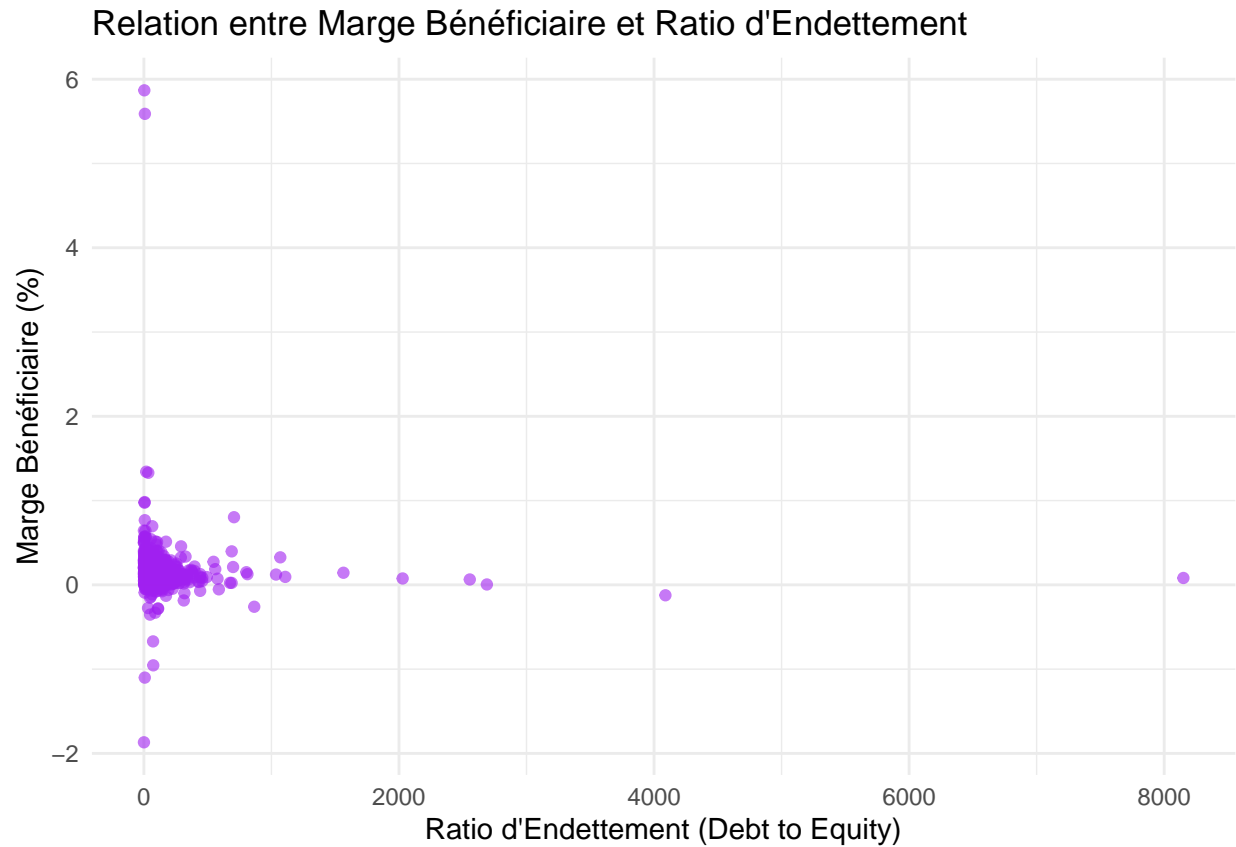
```
## # A tibble: 43 x 2
##   country      Count
##   <chr>         <int>
## 1 United States  501
## 2 Japan          84
## 3 Canada         51
## 4 China          38
## 5 Germany        37
## 6 India          34
## 7 France         33
## 8 United Kingdom  30
## 9 Switzerland    29
## 10 Australia     23
## # i 33 more rows
```



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

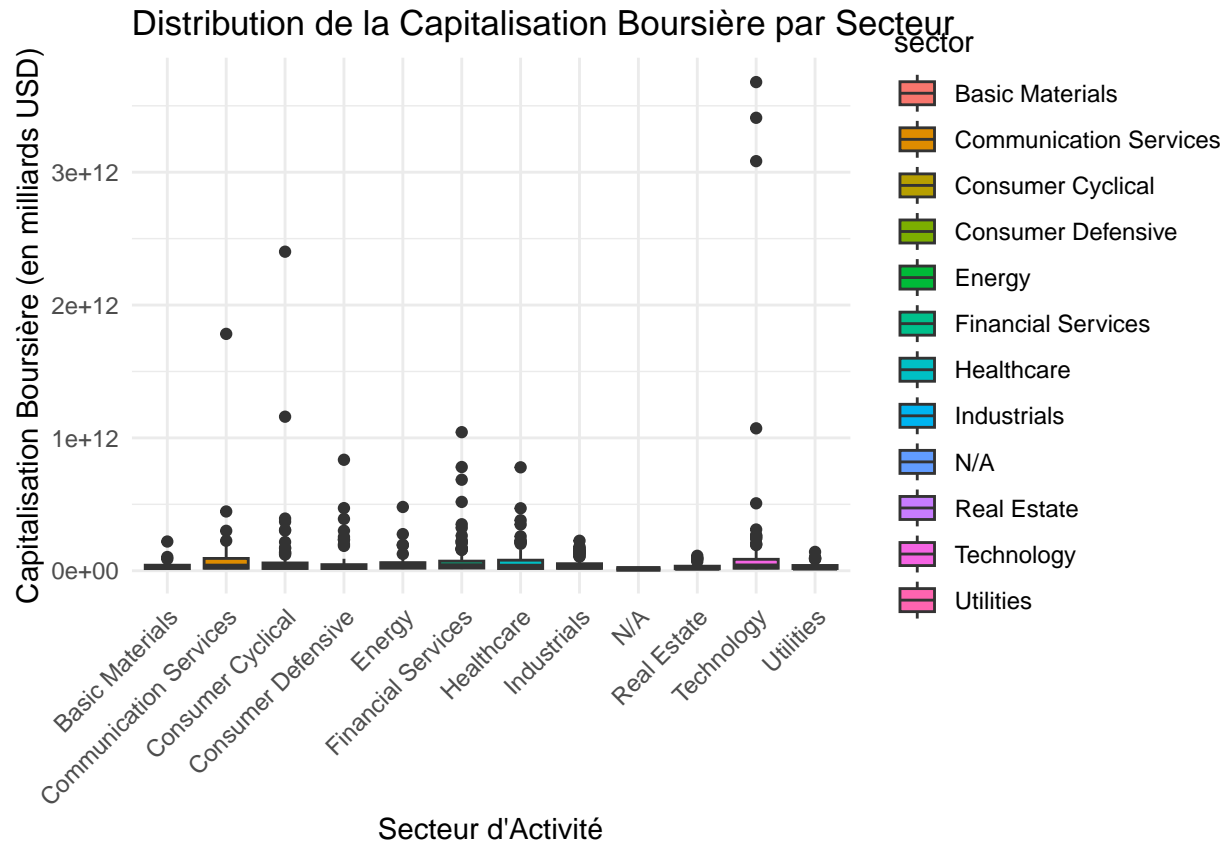
Un scatter plot est pertinent pour analyser la relation entre deux variables financières, comme *profitMargins* et *debtToEquity*, afin de visualiser s'il existe une corrélation entre la rentabilité et l'endettement.

```
## Warning: Removed 144 rows containing missing values or values outside the scale range
## ('geom_point()').
```



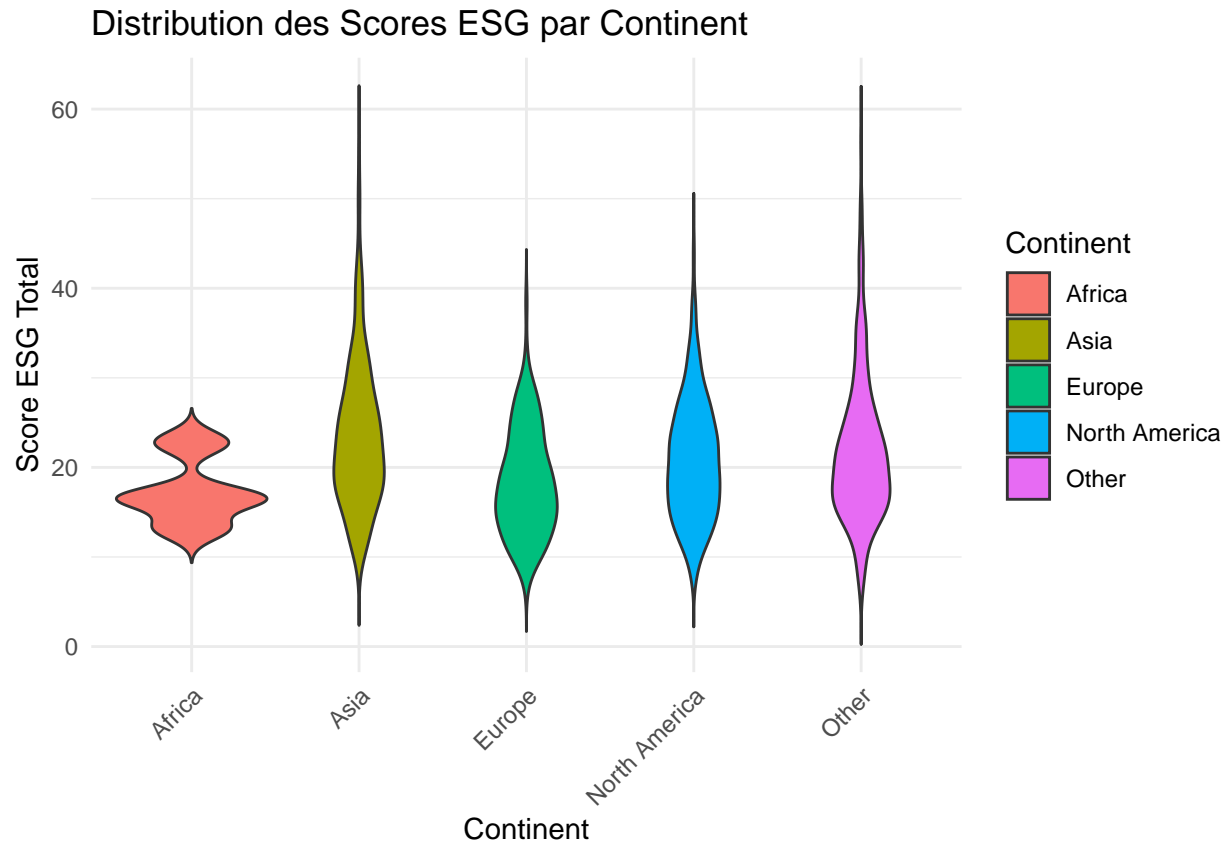
## 2. Distribution de la Capitalisation Boursière par Secteur

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.



### 3. Scores ESG par Continent

Un diagramme en violon pourrait être intéressant pour analyser la distribution des scores ESG (*totalEsg*) en fonction des continents. Cela permettrait de visualiser les écarts dans la performance ESG en fonction de la région géographique.



#### 4. Comparaison de la Rentabilité (ROA vs ROE)

Un graphique en nuage de points est une bonne manière de comparer les deux indicateurs financiers, à savoir le Return on Assets (ROA) et le Return on Equity (ROE), pour chaque entreprise.

```
## Warning: Removed 107 rows containing missing values or values outside the scale range  
## ('geom_point()').
```

## Comparaison de la Rentabilité des Actifs (ROA) et des Fonds Propres (ROE)

