

Project-1

Haoran Shi and Jingyao Zhang

2024-10-17

Abstract

The dataset for this project is about the ESG (Environmental, Social, and Governance) scores of S&P 500 companies in 2023-2024. The project attempts to find the relationship between different industries and their ESG scores. The goal is to further explore the industries that focus on ESG index, compare the distribution of information technology, healthcare, and finance in ESG index, and explore the importance that different industries attach to governance scores. As a result, Real Estate, Information Technology, and Consumer Discretionary are the top three industries that focus on ESG index. The Information Technology industry is more conscious of its impact than the Financial and Health Care industries. Surprisingly, the Financial sector has a relatively higher governance score than all other sectors.

Introduction

As society develops, the world is paying more and more attention to the performance and impact of companies in terms of environment, society, and governance. As leaders and models in various industries, the S&P 500 companies impact other companies worldwide. Their performance and management direction play a leading role. Different sectors may pay distinct attention to the ESG index. This project mainly explores the ESG scores of various industries and their emphasis on ESG impact.

Analyzing the industries with the top 10% index can give people an overview of the industry distribution of S&P500 companies. In addition, Information Technology, Healthcare, and Finance are three critical fields that greatly impact people's quality of life and social environment. Analyzing the percentile distribution of companies in these three fields helps us further understand these three industries' current status and trends. Furthermore, since supervision is essential for each sector, looking at the average governance score of each industry can provide a direction for further improvement of society.

Data

The data set selected focuses on the ESG (Environmental, Social, and Governance) scores of companies listed in the S&P 500 index. The environmental score evaluates a company's impact on the environment, considering factors such as carbon emissions, energy consumption, and waste management. The social score assesses the company's relationships with its employees, suppliers, customers, and the communities in which it operates. The governance score measures corporate governance practices, including executive compensation, audits, internal controls, and shareholder rights. Combined, the ESG score represents the overall risk or impact a company has across these three areas, with a lower ESG score indicating a more positive or less harmful impact on society. Companies are ranked relative to each other based on their ESG scores, with higher percentiles assigned to companies that achieve higher ESG scores (*S&P 500 ESG and Stocks Data 2023-24 — Kaggle.com*).

Consider exploring **this link** which gives some insight to how the ESG score translates to a company's ranking and thus their percentile by using Amazon.com,Inc. as an example (*Company ESG Risk Rating - Sustainalytics — Sustainalytics.com*).

A key limitation of this data set lies in the fact that ESG scores are self-reported, meaning their accuracy depends on the transparency and good faith of the reporting companies. Additionally, the factors and weightings used to calculate ESG scores vary by sector, with each sector placing emphasis on different aspects and assigning varying levels of importance to specific factors. This variation complicates the comparability of ESG scores across sectors, raising questions about the fairness and consistency of such comparison (Karlán-Mason).

Analysis and Results

Using R Core Team (2024) and RStudio Team (2020), and the packages Wickham (2016) and Wickham et al. (2023), the following analysis on the data set *S&P 500 ESG and Stocks Data 2023-24 — Kaggle.com* has been conducted:

Industry Distribution of the 10% Companies with the Lowest ESG Scores

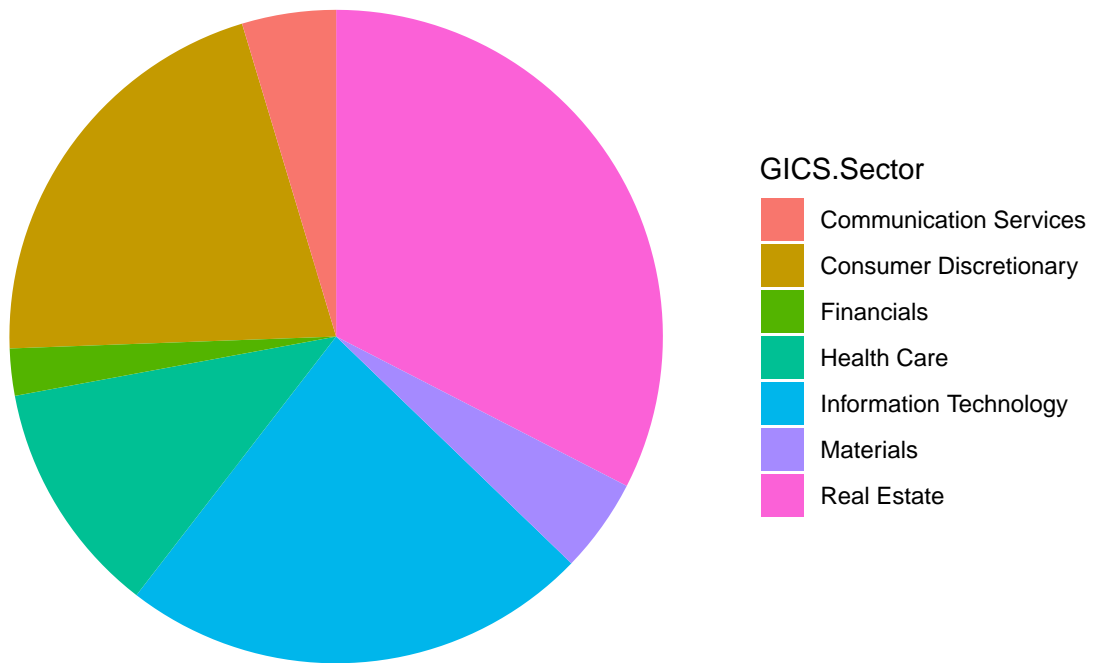
This section aims to investigate the 10% of companies with the lowest ESG total scores and their industry distribution to explore the correlation between score and type of industry.

```
# Identify top 10% companies with lowest score
top_10_percent_value <- quantile(esg.data$totalEsg, probs = 0.1)
companies_with_lowest_score <- esg.data |>
  filter(totalEsg <= top_10_percent_value)

# Create a pie chart to explore the distribution of industries in the 90th
# percentile
sector_counts <- companies_with_lowest_score |>
  group_by(GICS.Sector) |>
  summarise(count = n()) |>
  arrange(desc(count))

ggplot(sector_counts, aes(x = "", y = count, fill = GICS.Sector)) + geom_bar(stat = "identity") +
  coord_polar("y", start = 0) + theme_void() + labs(title = "Industry Distribution in the 90th Percentile")
```

Industry Distribution in the 90th Percentile of ESG Scores



```
print(sector_counts)
```

```
## # A tibble: 7 x 2
##   GICS.Sector      count
##   <chr>          <int>
## 1 Real Estate      14
## 2 Information Technology  10
## 3 Consumer Discretionary    9
## 4 Health Care         5
## 5 Communication Services    2
## 6 Materials          2
## 7 Financials         1
```

The pie chart shows that among the lowest 10% of the S&P 500 ESG index, the real estate industry has the most companies, 14. The information technology and Consumer Discretionary industries are second and third. This reflects that society has attached great importance to the real estate industry's concept of green environmental protection in recent years (*ESG as a Value Driver for Real Estate* — [Www2.deloitte.com](http://www2.deloitte.com)).

Distribution of ESG Score Percentiles within Select Industries

This next section seeks to investigate the percentile distribution the of ESG scores of companies within in 3 sectors: Information Technology, Health Care, and Financials. These 3 sectors were chosen because of their similarity in size, with Information Technology having 54 companies, Health Care having 51 companies, and Financials having a slightly greater number of 65. Additionally, Wickham (2016) was used to construct a grouped bar plot.

```

# creates new column categorizing each entry into quartiles based on their
# percentile
for (i in 1:nrow(esg.data)) {

  if (25 >= esg.data$percentile[i]) {
    esg.data$quartiles[i] = "Quartile 1"
  } else if (50 >= esg.data$percentile[i]) {
    esg.data$quartiles[i] = "Quartile 2"
  } else if (75 >= esg.data$percentile[i]) {
    esg.data$quartiles[i] = "Quartile 3"
  } else {
    esg.data$quartiles[i] = "Quartile 4"
  }
}

# Creates a subset of esg.data with only the 3 selected sectors

esg.data.industries = data.frame(subset(esg.data, GICS.Sector == "Information Technology" |
  esg.data$GICS.Sector == "Health Care" | esg.data$GICS.Sector == "Financials"))

industries = c("Information Technology", "Health Care", "Financials")

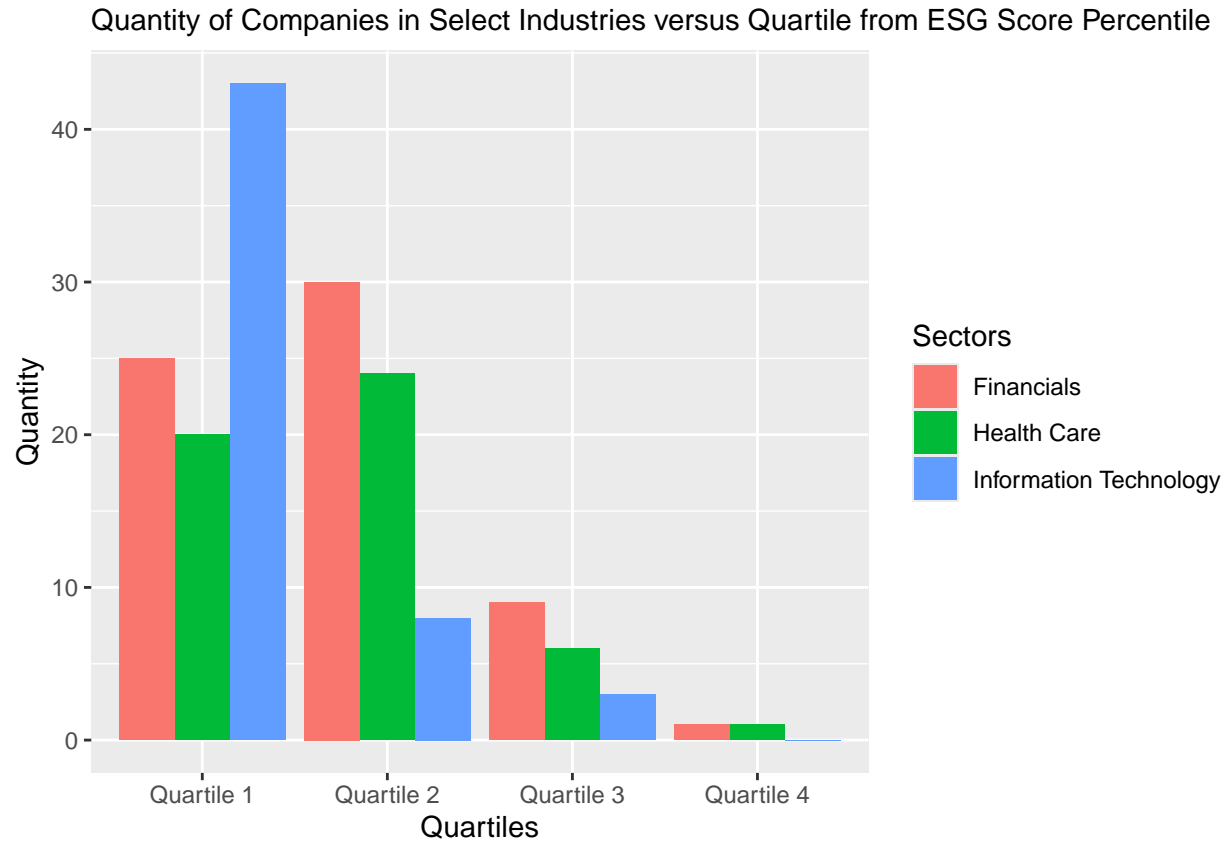
quartiles = c("Quartile 1", "Quartile 2", "Quartile 3", "Quartile 4")

heights = vector("numeric", 12)
mod.index = 0
for (i in 1:4) {
  for (j in 1:3) {
    heights[j + mod.index] = nrow(esg.data.industries[esg.data.industries$GICS.Sector ==
      industries[j] & esg.data.industries$quartiles == quartiles[i], ])
  }
  mod.index = mod.index + 3
}

# creates new data frame with only necessary data
mod.data = data.frame(Quantity = heights, Sectors = c(industries, industries, industries,
  industries), Quartiles = rep(quartiles, each = 3))

# creates grouped bar plot
ggplot(mod.data, aes(fill = Sectors, y = Quantity, x = Quartiles)) + geom_bar(stat = "identity",
  position = "dodge") + labs(title = "Quantity of Companies in Select Industries versus Quartile from
  theme(plot.title = element_text(size = 11))

```



Within the Financials and Health Care sectors, the distribution is such that there is a maximum reached in Quartile 2, meaning that more companies had ESG scores that were higher than 25% of all companies and lower than 50% of all companies. This differs greatly from the Information Technology sector, where the vast majority of companies had a percentile within the first quartile, meaning that these companies had the lowest ESG scores out of all the companies compared. This could suggest that companies within the Information Technology sector are more conscious of their impact on society. This could also be a reflection of the fact that these companies, such as Amazon and Apple, get more press coverage, and are thus under greater public scrutiny than companies from other sectors.

Mean Governance Score Disparity by Sector

In this section, the relationship between governance scores and sector is discussed.

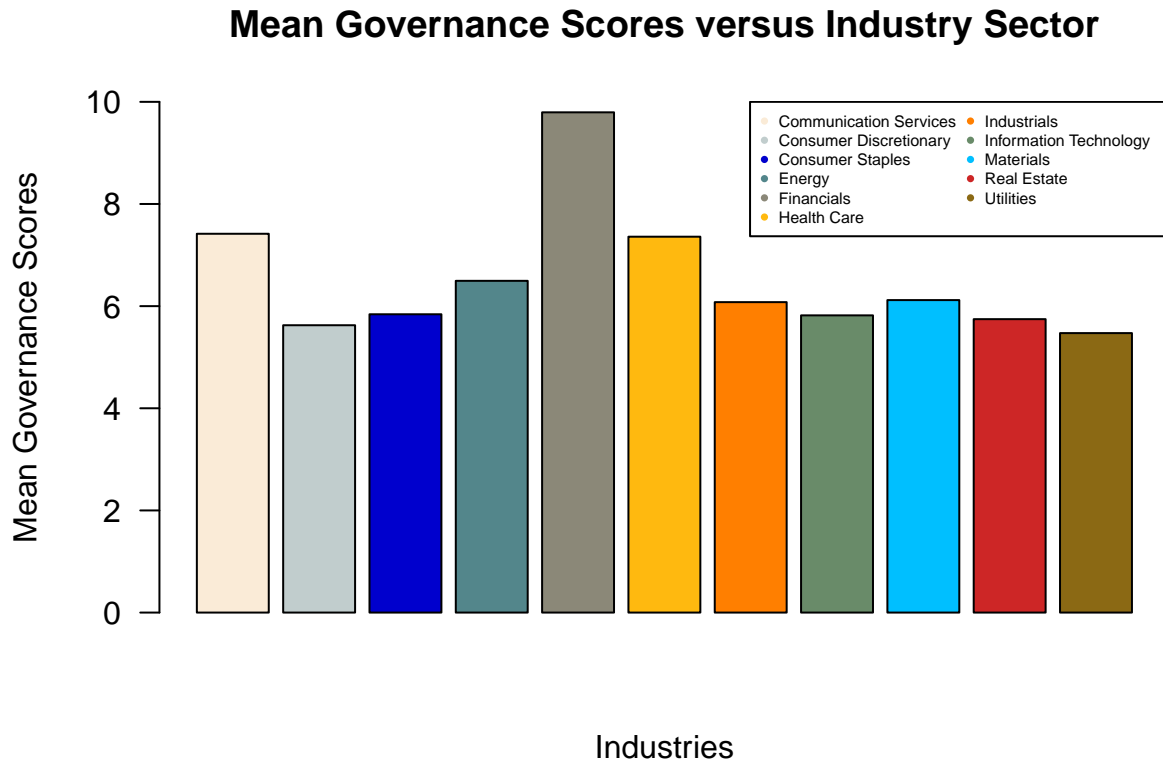
```
industries = c("Communication Services", "Consumer Discretionary", "Consumer Staples",
               "Energy", "Financials", "Health Care", "Industrials", "Information Technology",
               "Materials", "Real Estate", "Utilities")

gov.averages = vector("numeric", length(industries))

for (i in 1:length(gov.averages)) {
  gov.averages[i] = mean(esg.data$governanceScore[which(esg.data$GICS.Sector ==
                                                         industries[i])])
}

barplot(height = gov.averages, ylim = c(0, 10), xlab = "Industries", ylab = "Mean Governance Scores",
```

```
main = "Mean Governance Scores versus Industry Sector", las = 2, col = colours()[c(3,
  16, 29, 46, 67, 76, 91, 106, 121, 136, 151)])
legend("topright", col = colours()[c(3, 16, 29, 46, 67, 76, 91, 106, 121, 136, 151)]),
  legend = industries, cex = 0.5, pch = 16, ncol = 2)
```



This graph compares the mean governance scores for all sectors included in the *S&P 500 ESG and Stocks Data 2023-24* — *Kaggle.com* data set. The average mean governance score across all sectors is 6.5242049. It is noted that while most industries had similar governance scores, the Financials sector had a significantly greater mean governance score of 9.7946154. This suggests that companies within the Financials sector had more concerning corporate governance practices. Additionally, it is worth mentioning that the mean governance score of the Information Technology sector, which is 5.8192593, is one of the lower scores, which aligns with how most companies in this sector had lower overall ESG scores as discussed in prior section.

Conclusion

This project sought to explore the relationships between the different sectors of companies within the S&P 500 and their ESG Scores. Real Estate is the industry that most focuses on the ESG score as shown in the first analysis section. Since Real Estate is usually an important part of business, society is trying to make it more sustainable currently. The Information Technology sector is more concerned about the ESG impact than Finance and Health Care sectors. Financial sector has the highest average governance score within all sectors, which indicates that there is an improvement room for the Finance field.

Contribution Statement

Haoran Shi:

Completed Data, the 2nd and 3rd plots and their analyses, all references and bibliography, and collaborated on Conclusion.

Jingyao Zhang:

Completed Abstract, Introduction, 1st plot and analysis, and collaborated on Conclusion.

References

- Company ESG Risk Rating - Sustainalytics* — *sustainalytics.com*. <https://www.sustainalytics.com/esg-rating/amazon-com-inc/1007896995>.
- ESG as a value driver for real estate* — *www2.deloitte.com*. <https://www2.deloitte.com/cz/en/pages/real-estate/articles/the-impact-of-social-good-on-real-estate.html>.
- Karlan-Mason, G. *What's an ESG Score? How ESG Ratings Work | GreenChoice* — *greenchoicenow.com*. <https://www.greenchoicenow.com/v/esg>.
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- RStudio Team. (2020). *RStudio: Integrated development environment for r*. RStudio, PBC. <http://www.rstudio.com/>
- S&P 500 ESG and Stocks Data 2023-24* — *kaggle.com*. <https://www.kaggle.com/datasets/rikinzala/s-and-p-500-esg-and-stocks-data-2023-24>.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation*. <https://CRAN.R-project.org/package=dplyr>