Project-1

Haoran Shi and Jingyao Zhang

2024-10-17

Abstract ## Tina

Introduction ## Tina

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 (Karlan-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:

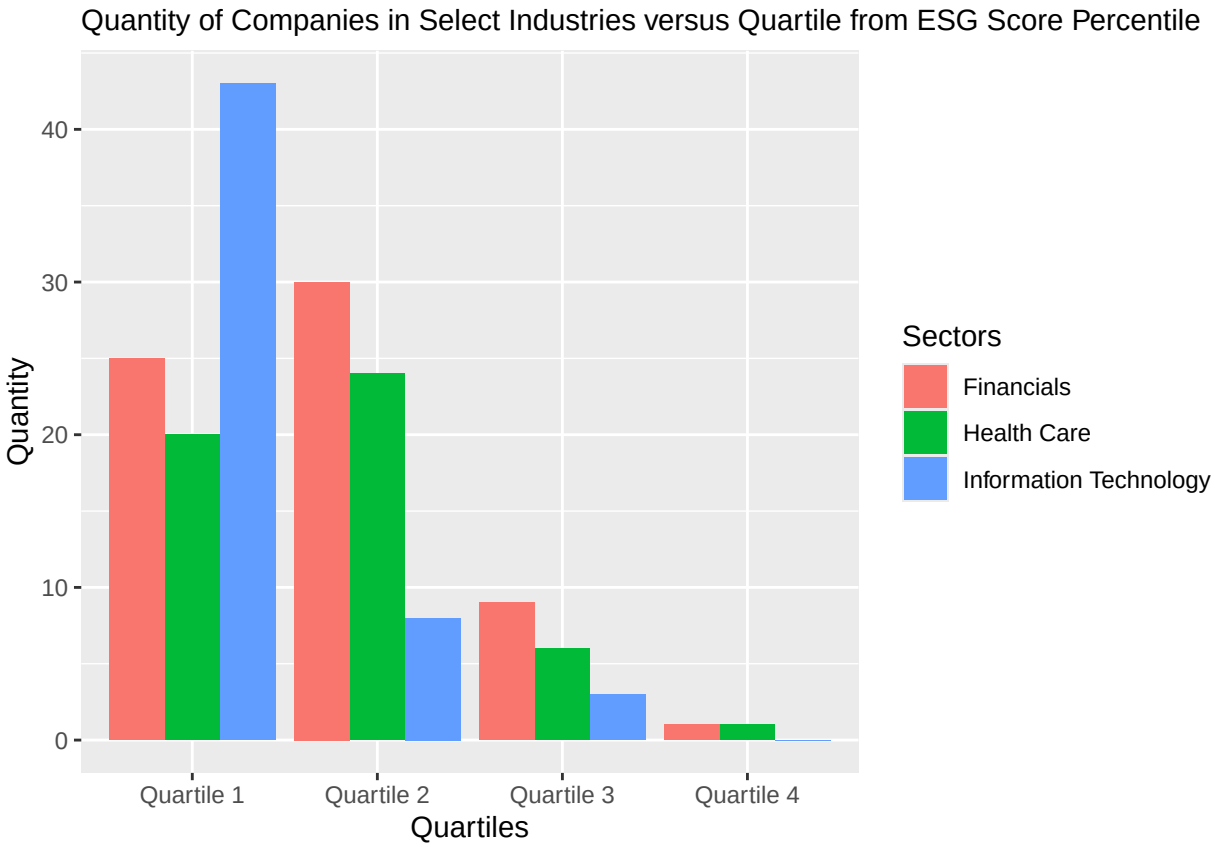
To see top 5% company that with lowest score, and ts. o explore correlation between score and type of industry Within 90th percentile companies, pie chart of percentage each industry in the 90th percentile

```
# Identify top 10% companies with lowest score
```

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 (25.0>=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.S


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]

    }
    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), (

# 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 Industri
```

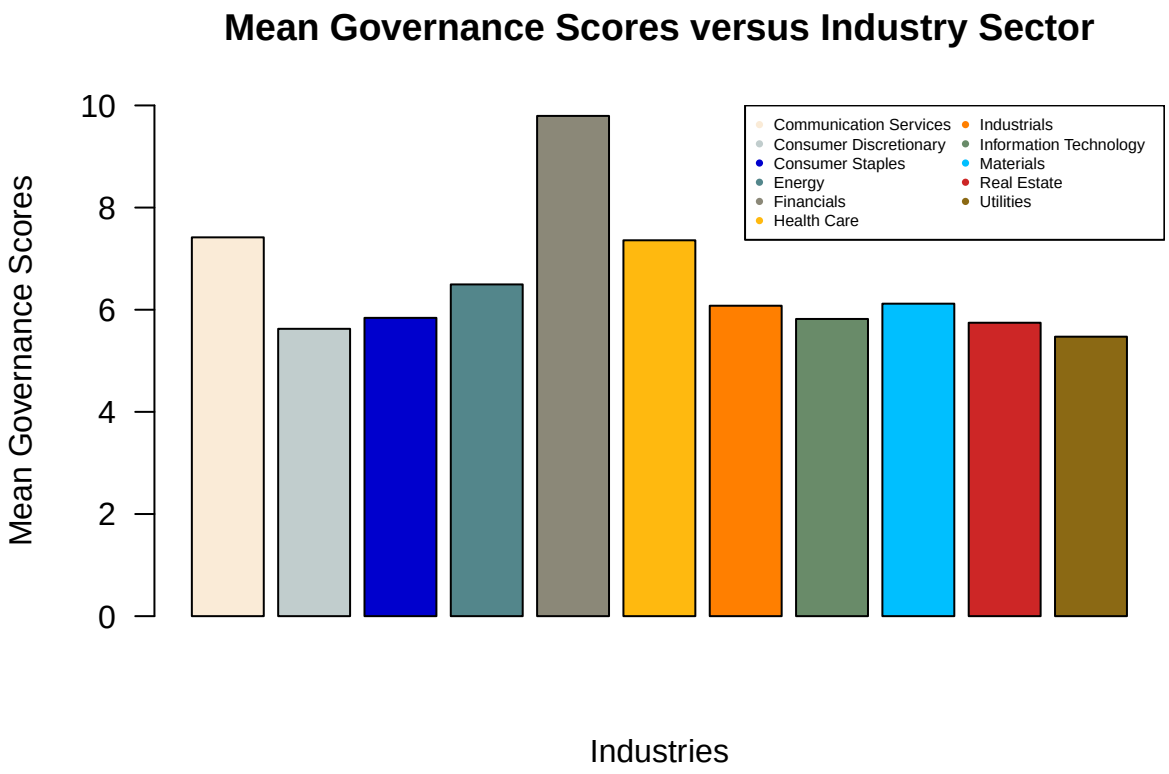


Within the Financials and Health Care sectors, the distribution is such that there is a maximum reached in Quartile 2, meaning that more companies had ESG scores that were higher than 25% of all companies and lower than 50% of all companies. This differs greatly from the Information Technology sector, 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", "Financ

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", ma
legend("topright", col= colours()[c(3,16,29,46,67,76,91,106,121,136,151)], legend = industries, cex = 0
```



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

Contribution Statement

Haoran Shi:

Completed Data, the 2nd and 3rd plots and their analyses, and the references.

Jingyao Zhang:

References

Company ESG Risk Rating - Sustainalytics — sustainalytics.com. https://www.sustainalytics.com/esg-rating/amazon-com-inc/1007896995.

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