Where I found the data set I am analyzing:

<https://www.kaggle.com/datasets/pritish509/s-and-p-500-esg-risk-ratings>

**What factors influence a corporation’s total ESG (Environmental, Social, and Governance) score?**

**Initial Analysis**

To begin, I hypothesized that there would be minimal or no relationship between the number of total workers in a corporation and its total ESG score. My reasoning was that the ESG score should be driven by other factors unrelated to workforce size.

To test this hypothesis, I conducted a linear regression analysis in R, examining the relationship between the total number of workers and the total ESG score. The regression results, including a scatter plot, indicated a very low correlation between these variables, with an R2R^2R2 value of 0.0003. This confirms my initial assumption that workforce size does not significantly affect ESG scores.

**Clustering Analysis**

In a subsequent analysis, I sought to identify patterns or groupings in the dataset using clustering techniques. Initially, I applied the elbow method to determine the optimal number of clusters, but no clear KKK-value emerged. As an alternative, I used the Silhouette Coefficient to identify the appropriate number of clusters. After performing KKK-means clustering, I visualized the results in a scatter plot, labeling data points by industry.

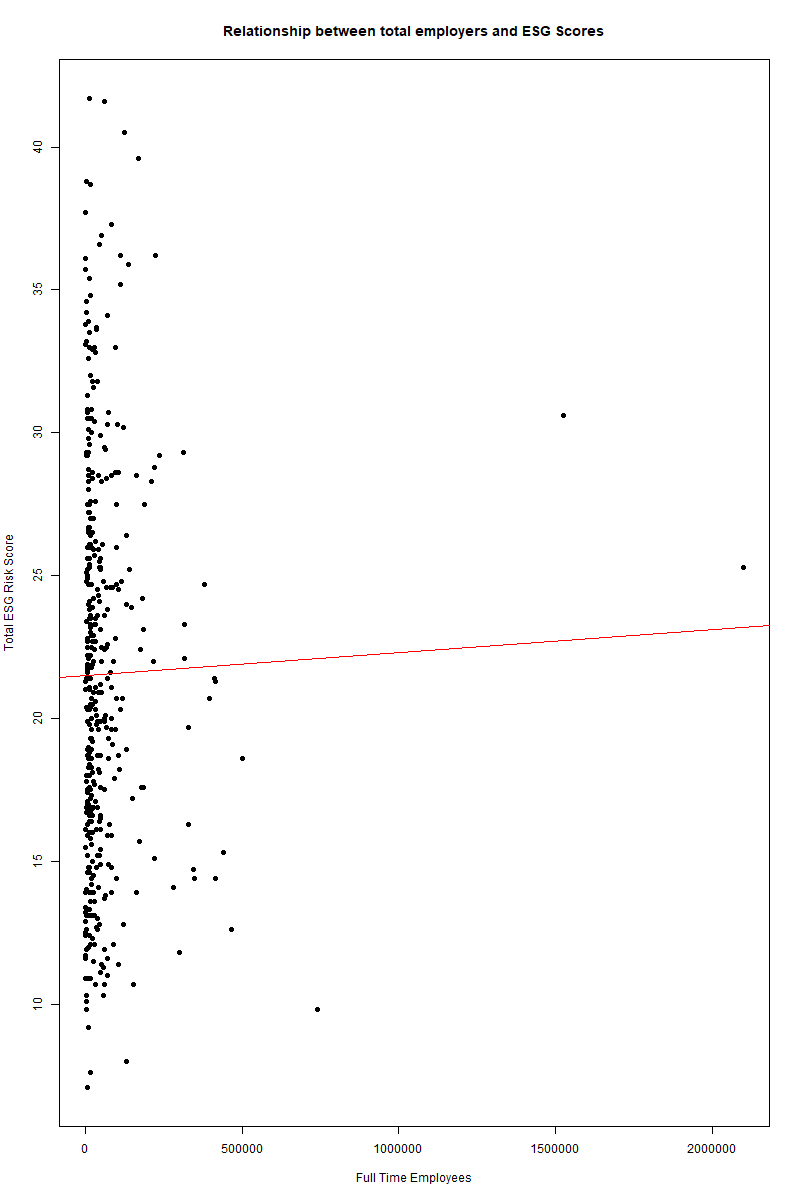
The clustering analysis revealed that ESG scores varied widely across industries, with no consistent clustering pattern for specific sectors. This suggests that industry affiliation alone does not reliably predict ESG performance.

**Next Steps**

1. **Industry-Specific Analysis**: Investigate the performance of individual industries in greater depth to identify any unique trends or outliers.
2. **Regression on Other Variables**: Explore additional columns in the dataset to determine whether other factors show a stronger correlation with total ESG scores.
3. **Expanded Clustering Analysis**: Revisit the clustering approach using additional features or alternative clustering algorithms to uncover deeper insights.

A screen shot of a computer

Description automatically generated



Industry Specific Analysis

In the image below I performed a box plot of the R-Squared Values of each sector speficall. From what I can tell the industry is not a noticeable contributing factor for the total amount of workers and the ESG score. 