Table of Contents

[**1.** **Introduction** 3](#_Toc187104808)

[**1.1.** **Problem Statement and Research Motivation** 3](#_Toc187104809)

[**1.2.** **The Dataset** 3](#_Toc187104810)

[**1.3.** **Research Question** 4](#_Toc187104811)

[**1.4.** **Null and Alternative Hypotheses** 4](#_Toc187104812)

[**2.** **Background Research** 4](#_Toc187104813)

[**2.1.** **Research Papers Review** 4](#_Toc187104814)

[**2.2.** **Research Gap and Interest** 5](#_Toc187104815)

[**3.** **Visualization** 5](#_Toc187104816)

[**3.1.** **Appropriate Plot for the RQ Output of an R Script** 5](#_Toc187104817)

[**3.2.** **Additional Information Relating to Understanding the Data** 6](#_Toc187104818)

[**Figure 4: Histogram of world rankings showing the distribution across ranks.** 7](#_Toc187104819)

[**3.3.** **Useful Information for Data Understanding** 7](#_Toc187104820)

[**4.** **Analysis** 7](#_Toc187104821)

[**4.1.** **Statistical Test Selection** 7](#_Toc187104822)

[**4.2.** **Hypothesis Testing Results** 8](#_Toc187104823)

[**5.** **Evaluation** 8](#_Toc187104824)

[**5.1.** **What went well** 8](#_Toc187104825)

[**5.2.** **Points for improvement** 8](#_Toc187104826)

[**5.3.** **Group’s time management (50 words)** 9](#_Toc187104827)

[**5.4.** **Project’s overall judgment** 9](#_Toc187104828)

[**5.5.** **Changes to group since Assignment 1** 9](#_Toc187104829)

[**5.6.** **Comment on the GitHub log output** 10](#_Toc187104830)

[**6.1.** **Results Explained** 10](#_Toc187104831)

[**6.2.** **Results Interpretation** 10](#_Toc187104832)

[**6.3.** **Future Work and Limitations** 10](#_Toc187104833)

[**7.** **Reference List** 11](#_Toc187104834)

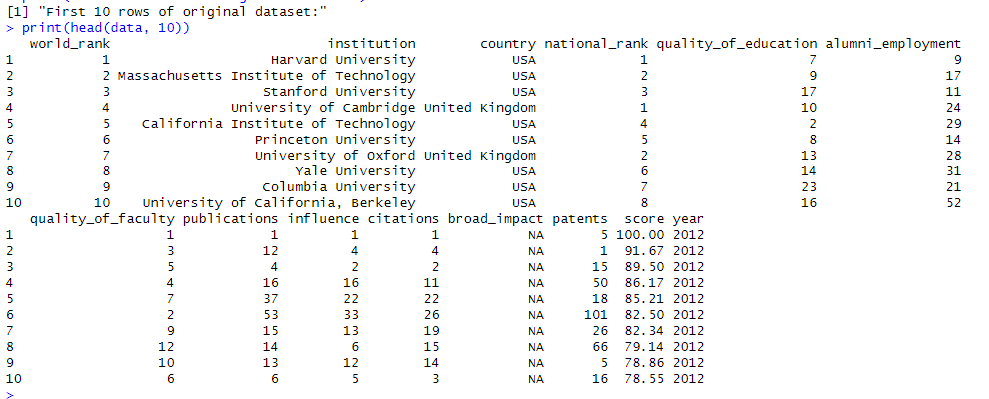
[**8.** **Appendices** 12](#_Toc187104835)

1. **Introduction**
   1. **Problem Statement and Research Motivation**

In an increasingly interconnected educational market, universities around the world are under increasing pressure to improve their global competitiveness and draw in top talent. World university rankings are used by many institutions as strategic growth tools and important performance indicators (Estrada-Real & Cantu-Ortiz, 2022). There is a knowledge gap for institutional development, too, as little is known about how elements like educational quality affect these rankings. Using the CWUR dataset, this study attempts to examine the connection between university rankings and educational quality measurements, offering data-driven insights to assist institutions in strengthening their performance, aligning with international standards, and improving their strategy.

* 1. **The Dataset**

The "cwurData.csv" dataset (2012–2015), which contains 14 performance metrics and detailed annual university rankings, is used in this study. With an emphasis on world rank (a dependent variable) and educational quality (an independent variable), each record depicts a university's position in the world. The dataset, which includes top universities globally, provides a thorough understanding of ranking criteria and academic performance. The study offers a strong basis for assessing how educational quality affects international university rankings by exploring these links.



**Figure 1: Load the Dataset and Print 10 Sample Data**

* 1. **Research Question**

This study studies the relationship between 2012 and 2015 between university rankings (global rank) and educational quality. In order to evaluate the influence of educational standards on rankings within the global higher education landscape, it uses the CWUR dataset to investigate the relationship between educational quality metrics and worldwide university classifications.

**Methodology**

Regression modeling, correlation analysis, and data visualization techniques will all be used in our statistical study in R. Using statistical tests and smoothing techniques, the study investigates the extent and nature of the correlation between world rankings and quality of education scores across a number of years.

* 1. **Null and Alternative Hypotheses**

**H0 (Null Hypothesis):** In the CWUR dataset from 2012 to 2015, there is not an apparent connection between a university's world ranking position and its quality of education score (correlation coefficient = zero).

**H1 (Alternative Hypothesis):** According to the CWUR dataset from 2012 to 2015, there is a substantial link (correlation coefficient is not equal to zero) between a university's world ranking position and its quality of education score. This implies that universities with better (lower) international rankings are probably those with greater quality of education ratings. To ascertain the type and degree of this link, statistical tests will be undertaken.

1. **Background Research**
   1. **Research Papers Review**

Global ranking datasets have been used in a number of research on university rankings to examine the variables affecting institutional performance. Estrada-Real and Cantu-Ortiz (2022) created automated predictive models utilizing a 10-year dataset from the QS World University Rankings. After examining performance criteria such as faculty ratios, citations, and academic repute, they discovered that these factors had a big impact on organization rankings. Their research demonstrates how colleges can raise their rankings through strategic planning based on data-driven insights.

Using principal component analysis, Robinson-Garcia et al. (2023) analyzed data from several international university rankings, such as ARWU, THE, and QS. Based to their research, research production and citation impact had a greater influence on rankings than teaching quality or student happiness, even if different ranking systems had different methodologies. This confirms the notion that global rankings significantly weight research metrics.

In order to find patterns, Bublyk et al. (2023) conducted a thorough statistical analysis of the QS World University Rankings, concentrating on correlation analysis and smoothing techniques. They highlighted faculty ratios and citation metrics as important factors influencing university performance. According to their findings, institutions can use rankings data strategically while being aware of its limits.

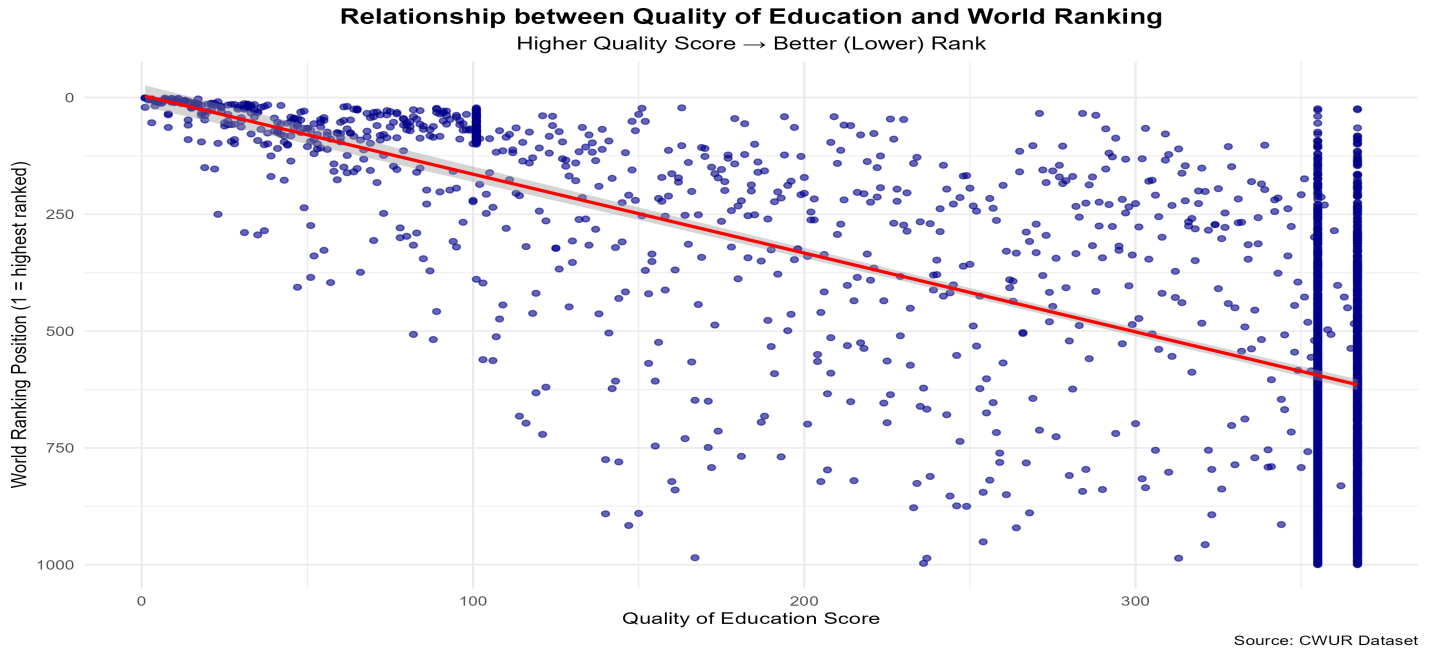
The foundation for comprehending the connection between university rankings and elements like quality of education is provided by these studies, which offer insightful information about how specific results criteria influence global rankings.

* 1. **Research Gap and Interest**

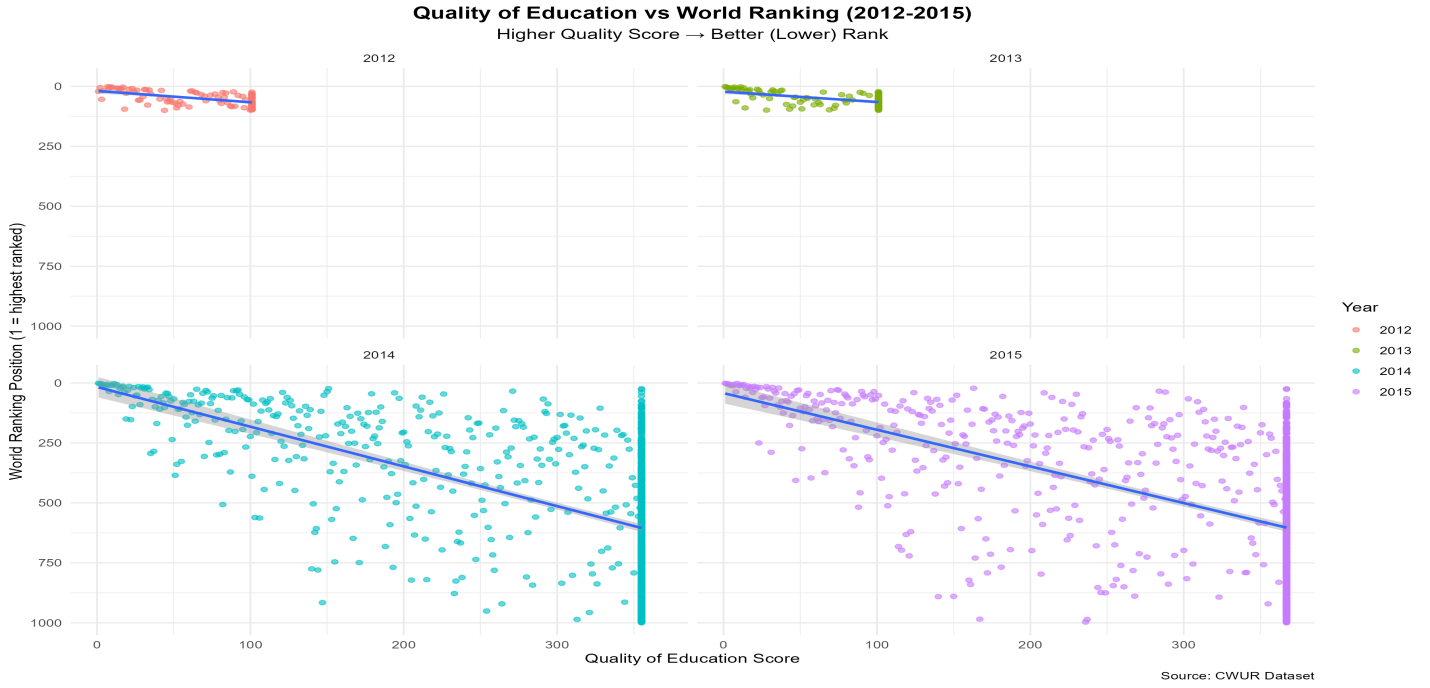
The research exposed is quite interesting since it fills a vacuum in the literature, which mostly concentrates on research metrics and the influence of citations on university rankings. Even though recent research highlights these elements, it ignores more comprehensive facets of student outcomes and educational quality. More thorough evaluation frameworks are required in order to effectively inform strategic decision-making, take institutional diversity consideration, and integrate teaching quality and student achievement. Beyond merely research-driven measures, future studies should focus on establishing balanced assessment techniques that offer universities with useful information to raise their rankings and overall educational success.

1. **Visualization**
   1. **Appropriate Plot for the RQ Output of an R Script**

To illustrate the hyperlink between educational quality and university world rank from 2012 to 2015, a scatter plot with a regression line was chosen. Understanding the ranking method is aided by the inverted y-axis, which plots higher for lower rank values (higher rankings). To show the uncertainty of the prediction, confidence intervals are submitted.



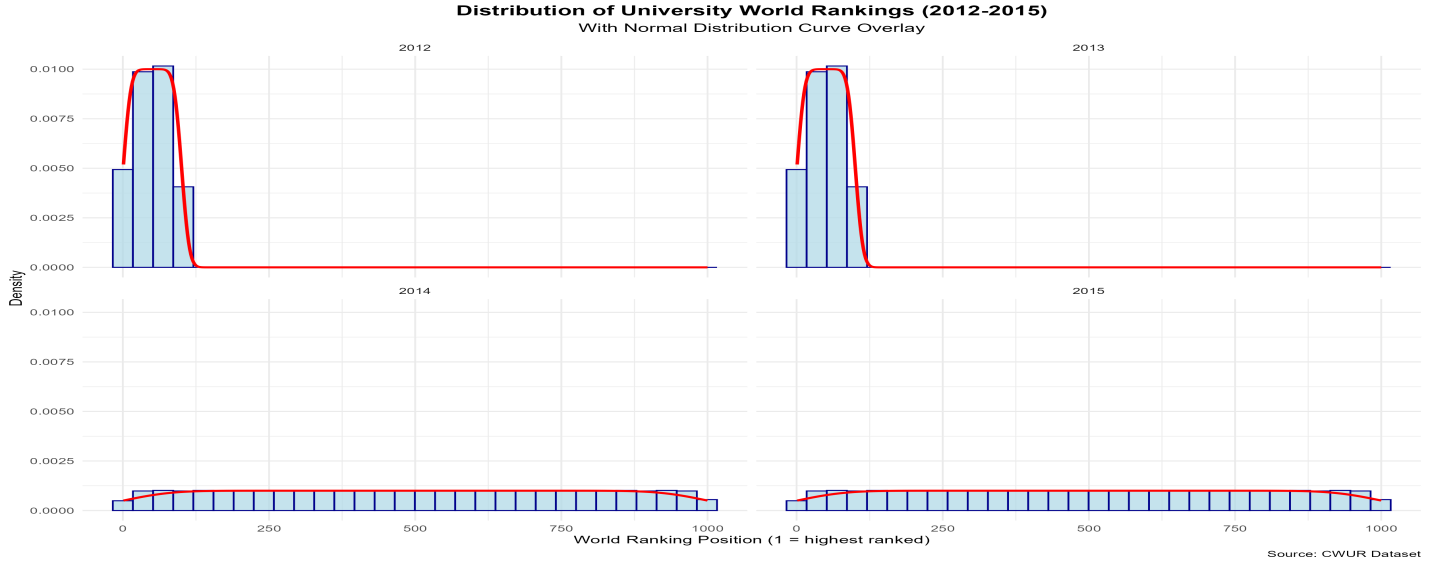
**Figure 2: Scatter plot showing the relationship between education quality and world rank.**

****

**Figure 3: Scatter plot by year, visualizing trends across the years (2012-2015).**

* 1. **Additional Information Relating to Understanding the Data**

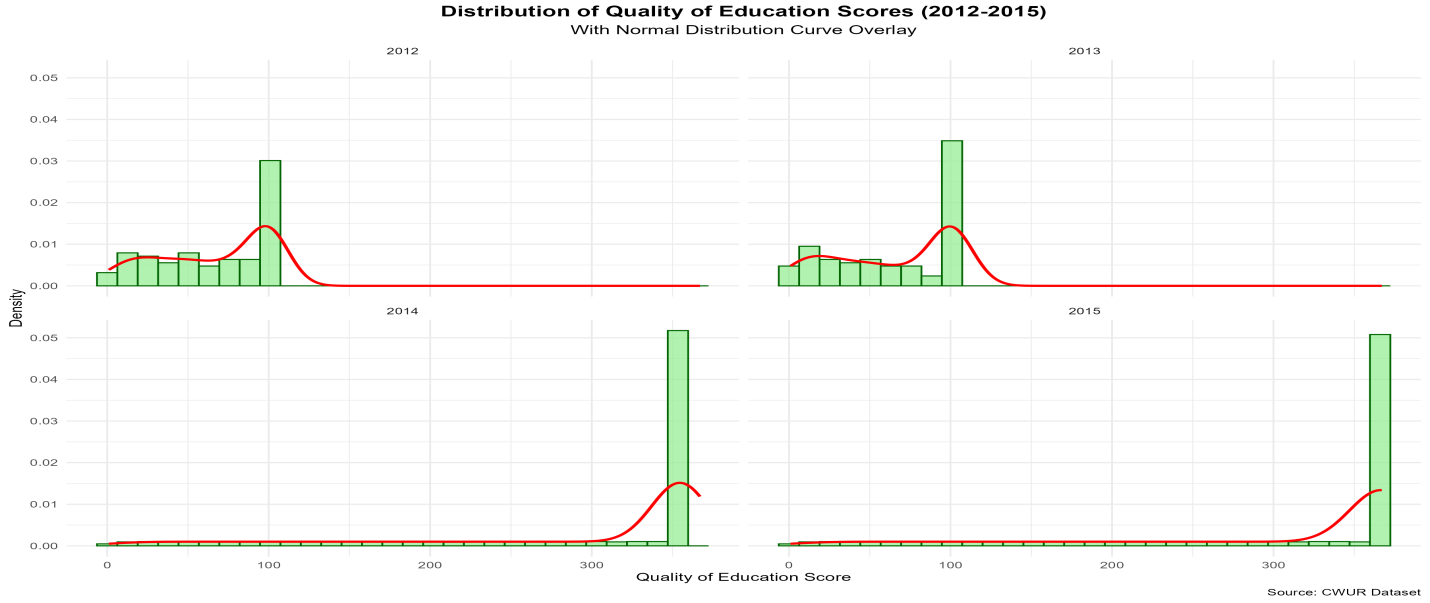
Fewer colleges are in the top or bottom rankings, while the majority are grouped in the center, according to the world rankings histogram. Considering Spearman's correlation works better with non-linear data, this distribution served as the basis for the examination of the relationship between ranks and educational quality.



**Figure 4: Histogram of world rankings showing the distribution across ranks.**

* 1. **Useful Information for Data Understanding**

The scatter plot with the regression line suggests that world rank and the quality of education are positively correlated. Though there is a lot of variety, better (lower) ranks are typically associated with higher quality schooling. This implies that rankings are influenced by a variety of factors, even though educational quality is among the most significant.



**Figure 5: Histogram of education quality scores to examine the spread of scores across universities.**

1. **Analysis**
   1. **Statistical Test Selection**

Because our data presented a non-normal distribution, as indicated by the Shapiro-Wilk test (p-value < 0.05), we used Spearman's rank correlation analysis to examine the association between educational quality and world rankings from 2012 to 2015. This non-parametric test is perfect for university ranking data analysis since it efficiently assesses the direction and strength of correlations between ranked variables without requiring a normal distribution, thereby rendering it especially appropriate for our study question.

* 1. **Hypothesis Testing Results**

We reject the null hypothesis based on the statistical analysis, that is backed by several important facts. There is a high positive association between world rankings and academic performance, as indicated by the Spearman's correlation coefficient of 0.633 and p-value < 2.2e-16. With an R-squared value of 0.4572, our regression model indicates that 45.72% of the variance in rankings can be explained by educational quality. The relevance of the model is further supported by the F-statistic of 1851.38 (p < 2.2e-16). These findings provide compelling evidence that better (lower) world ranking positions are correlated with higher educational quality.

1. **Conclusions**
   1. **Results Explained**

Higher ranks are linked to higher educational quality, according to the statistical analysis of the CWUR dataset from 2012 to 2015, which revealed a substantial positive relationship (rs = 0.633, p<2.2e-16) between university rankings and educational quality scores.

According to the regression model, 45.72% of the variation in rankings can be determined by educational quality, indicating that other factors may also play a role. Over the course of four years, the association was constant, and the Shapiro-Wilk test confirmed the use of Spearman's correlation by confirming non-normal data.

* 1. **Results Interpretation**

According to the findings, university rankings are positively impacted by educational quality, supporting the study question on how it affects standing internationally. Even while the link is strong, educational quality only partially explains the ranking variance, indicating that other important characteristics like reputation, citations, and research output are also important. To succeed in global rankings over the long run, universities looking to increase rankings should implement a holistic approach that enhances different performance categories, as well as the quality of education.

* 1. **Future Work and Limitations**

The study has a number of drawbacks, such as its exclusive focus on one ranking system (CWUR), its constrained timeframe of 2012–2015, and its limited use of criteria to determine educational quality. More quality indicators, a comparison of different ranking systems, longer time periods, institutional and regional variances, and an examination of the causal relationship between ranking changes and increases in educational quality should all be included in additional research.

1. **Reference List**
2. Bublyk, M., Vysotska, V., Fernandes, E., & Dziuba, R. (2023). World Universities Strategic Analysis Based on Data from the QS World University Rankings. CEUR Workshop Proceedings, 3171, 817-837.
3. Estrada-Real, A. C., & Cantu-Ortiz, F. J. (2022). A data analytics approach for university competitiveness: the QS world university rankings. International Journal on Interactive Design and Manufacturing, 16(3), 871-891.
4. Robinson-Garcia, N., Torres-Salinas, D., & Herrera-Viedma, E. (2023). Mining university rankings: Publication output and citation impact as their basis. Research Evaluation, 32(1), 1-17.
5. Moed, H. F. (2017). A critical comparative analysis of five world university rankings. Scientometrics, 110(2), 967-990.
6. Hazelkorn, E. (2019). Rankings and the Public Good Role of Higher Education. International Higher Education, 99, 7-9.
7. Liu, N. C. (2015). The Story of Academic Rankings. International Higher Education, 54, 2-3.
8. Vernon, M. M., Balas, E. A., & Momani, S. (2018). Are university rankings useful to improve research A systematic review. PloS one, 13(3), e0193762.
9. Altbach, P. G. (2015). The dilemmas of ranking. International Higher Education, 42, 2-3.
10. Docampo, D., & Cram, L. (2017). Academic performance and institutional resources: a cross-country analysis of research universities. Scientometrics, 110(2), 739-764.
11. Salmi, J. (2020). Excellence initiatives to create world-class universities: Do they work? International Higher Education, 100, 17-19.
12. Olcay, G. A., & Bulu, M. (2017). Is measuring the knowledge creation of universities possible: A review of university rankings. Technological Forecasting and Social Change, 123, 153-160.
13. Safón, V. (2019). Inter-ranking reputational effects: an analysis of the Academic Ranking of World Universities (ARWU) and the Times Higher Education World University Rankings (THE) reputational relationship. Scientometrics, 121(2), 897-915.
14. **Appendices**

**### Appendix A: R Code**

# ============================================

# University Rankings Analysis

# Research Question: How does the quality of education influence university rankings

# (world rank) between 2012 and 2015?

# ============================================

# Load required libraries

suppressWarnings({

library(ggplot2)

library(dplyr)

library(tidyr)

})

# Set random seed for reproducibility

set.seed(123)

# Load and examine data

data <- read.csv("H:/assisgnment/hertfordshire assignment/Krishna stoke/Dataset/archive (1)/cwurData.csv")

# Display initial data information

print("First 10 rows of original dataset:")

print(head(data, 10))

# Dataset Structure

print("\nDataset Structure:")

str(data)

# Summary of Original Dataset

print("\nSummary of Original Dataset:")

summary(data)

# Create focused table for study period

data\_table <- data %>%

filter(year >= 2012, year <= 2015) %>%

select(world\_rank, institution, country, quality\_of\_education, year) %>%

arrange(year, world\_rank) %>%

head(20)

print("\nFirst 20 rows of data for 2012-2015 (selected columns):")

print(data\_table)

# Clean and preprocess data

data\_clean <- data %>%

filter(year >= 2012, year <= 2015) %>%

drop\_na(quality\_of\_education, world\_rank) %>%

mutate(

world\_rank = as.numeric(world\_rank),

rank\_category = factor(cut(world\_rank,

breaks = c(0, 100, 250, 500, 1000),

labels = c("Top 100", "101-250", "251-500", "501+"),

include.lowest = TRUE

))

) %>%

filter(between(quality\_of\_education,

quantile(quality\_of\_education, 0.25) - 1.5 \* IQR(quality\_of\_education),

quantile(quality\_of\_education, 0.75) + 1.5 \* IQR(quality\_of\_education)

))

# Generate yearly statistics

summary\_by\_year <- data\_clean %>%

group\_by(year) %>%

summarise(

mean\_world\_rank = mean(world\_rank),

mean\_quality\_education = mean(quality\_of\_education),

correlation = cor(quality\_of\_education, world\_rank, method = "spearman"),

.groups = 'drop'

)

print("\nSummary Statistics by Year:")

print(summary\_by\_year)

# Create scatter plot

scatter\_plot <- ggplot(data\_clean, aes(x = quality\_of\_education, y = world\_rank)) +

geom\_point(aes(color = factor(year)), alpha = 0.6) +

geom\_smooth(method = "lm", formula = y ~ x, se = TRUE) +

scale\_y\_reverse() +

facet\_wrap(~year) +

labs(

title = "Quality of Education vs World Ranking (2012-2015)",

subtitle = "Higher Quality Score → Better (Lower) Rank",

x = "Quality of Education Score",

y = "World Ranking Position (1 = highest ranked)",

color = "Year",

caption = "Source: CWUR Dataset"

) +

theme\_minimal() +

theme(

plot.title = element\_text(hjust = 0.5, face = "bold", size = 14),

plot.subtitle = element\_text(hjust = 0.5, size = 12),

axis.title = element\_text(size = 11)

)

# Save scatter plot

ggsave("scatter\_plot\_by\_year.png", scatter\_plot, width = 12, height = 8, dpi = 300)

# Create histogram for world rankings distribution

hist\_plot <- ggplot(data\_clean, aes(x = world\_rank)) +

geom\_histogram(aes(y = after\_stat(density)),

bins = 30,

fill = "lightblue",

color = "darkblue",

alpha = 0.7) +

geom\_density(color = "red",

linewidth = 1) +

facet\_wrap(~year) +

labs(

title = "Distribution of University World Rankings (2012-2015)",

subtitle = "With Normal Distribution Curve Overlay",

x = "World Ranking Position (1 = highest ranked)",

y = "Density",

caption = "Source: CWUR Dataset"

) +

theme\_minimal() +

theme(

plot.title = element\_text(hjust = 0.5, face = "bold", size = 14),

plot.subtitle = element\_text(hjust = 0.5, size = 12),

axis.title = element\_text(size = 11)

)

# Save histogram

ggsave("histogram\_worldrank.png", hist\_plot, width = 12, height = 8, dpi = 300)

# Create histogram for quality of education distribution

edu\_hist\_plot <- ggplot(data\_clean, aes(x = quality\_of\_education)) +

geom\_histogram(aes(y = after\_stat(density)),

bins = 30,

fill = "lightgreen",

color = "darkgreen",

alpha = 0.7) +

geom\_density(color = "red",

linewidth = 1) +

facet\_wrap(~year) +

labs(

title = "Distribution of Quality of Education Scores (2012-2015)",

subtitle = "With Normal Distribution Curve Overlay",

x = "Quality of Education Score",

y = "Density",

caption = "Source: CWUR Dataset"

) +

theme\_minimal() +

theme(

plot.title = element\_text(hjust = 0.5, face = "bold", size = 14),

plot.subtitle = element\_text(hjust = 0.5, size = 12),

axis.title = element\_text(size = 11)

)

# Save education histogram

ggsave("histogram\_education.png", edu\_hist\_plot, width = 12, height = 8, dpi = 300)

# Statistical Analysis

# Shapiro-Wilk Normality Test

shapiro\_test <- shapiro.test(data\_clean$world\_rank)

print("\nShapiro-Wilk Normality Test Results:")

print(shapiro\_test)

# Correlation analysis by year

cor\_by\_year <- data\_clean %>%

group\_by(year) %>%

summarise(

correlation = cor(quality\_of\_education, world\_rank, method = "spearman"),

p\_value = cor.test(quality\_of\_education, world\_rank,

method = "spearman", exact = FALSE)$p.value,

.groups = 'drop'

)

print("\nCorrelation Analysis by Year:")

print(cor\_by\_year)

# Regression analysis

model <- lm(world\_rank ~ quality\_of\_education + factor(year), data = data\_clean)

# Create diagnostic plots

png("model\_diagnostics.png", width = 1200, height = 1200, res = 150)

par(mfrow = c(2, 2))

plot(model)

dev.off()

# Create regression summary

model\_summary <- summary(model)

coef\_table <- as.data.frame(coef(summary(model)))

colnames(coef\_table) <- c("Estimate", "Std\_Error", "t\_value", "p\_value")

write.csv(coef\_table, "regression\_summary.csv")

# Print final results

cat("\nComprehensive Analysis Results:\n")

cat("====================================\n")

# Normality Test

cat("1. Data Distribution Analysis:\n")

cat(" - Normality Test: ", ifelse(shapiro\_test$p.value > 0.05,

"Data appears normally distributed",

"Data does not appear normally distributed"), "\n")

cat(" - Shapiro-Wilk p-value:", format.pval(shapiro\_test$p.value), "\n\n")

# Correlation Analysis by Year

cat("2. Correlation Analysis by Year:\n")

print(cor\_by\_year)

# Regression Model Performance

cat("\n3. Regression Model Performance:\n")

cat(" - R-squared:", round(summary(model)$r.squared, 4), "\n")

cat(" - Adjusted R-squared:", round(summary(model)$adj.r.squared, 4), "\n")

cat(" - F-statistic:", round(summary(model)$fstatistic[1], 2), "\n")

# Full Regression Model Summary

cat("\nFull Regression Model Summary:\n")

print(summary(model))

### Appendix B: GitHub Log Output

[Include GitHub log output here]