7COM1079-0901-2025 - Team Research and Development Project

Final report title: A Comparison of Mean CWUR(Center for World University Rankings) Scores Between Universities in Asia and Europe.

Group ID: A 28

Dataset number: DS 276 (cwurData.csv),
https://www.kaggle.com/datasets/mylesoneill/world-university-rankings/data?select=cwurData.csv

Prepared by: Khant Zwe Naing (24073150),
Adeleye Francis Adewuyi(24092208), Muhammad Taha Mirza(24093676),
Mercy Toluwani Idowu(24160330), Jatin Kumar(24056471).

University of Hertfordshire
Hatfield, 2025

Table of Contents

5. Evaluation – group's experience at 7COM1079
   5.1. What went well
   5.2. Points for improvement
   5.3. Group's time management
   5.4. Project's overall judgement
   5.5. Comment on GitHub log output

6. Conclusions
   6.1. Results explained.
   6.2. Interpretation of the results
   6.3. Reasons and/or implications for future work, limitations

7. Reference list
   Harvard (author, date) format.
8. Appendices
   A. R code used for analysis and visualisation.
   B. GitHub log output.

1. Introduction

1.1. Problem statement and research motivation **(100 words)**

The Centre for World University Rankings (CWUR) is an autonomous institution that ranks universities worldwide on criteria including educational quality, alumni performance, and research capabilities. University rankings offer an administrative and academic perspective on performance and can assist policymakers and aspirant students. It has been made evident via research that geography and economic inequality impact regional university quality (Altbach, 2016; Marginson, 2017). It has also been made clear via research that discrepancies exist among budgetary allocation, pedagogy, and global outreach for Asian and European universities and can have implications for global rankings (Shin and Kehm, 2013). The document focuses on differing performances of Asian and European universities.

1.2. The data set **(75 words).**

This research uses a CWUR dataset, which has been produced by a third-party institution, categorizing universities worldwide in terms of education quality, employment of their graduates, research production, and worldwide reputation. This data set has variables like institution, world rank, publications, region, and CWUR Score. For this research, Region (Asia and Europe) will be the independent variable, with CWUR Score as the dependent variable. This data set provides a sufficient sample for a comparative study between Asia and Europe.

1.3. Research question **(50 words).**

The research question for this study is: "Is there a difference in the mean

CWUR (Centre for World University Rankings) Score between universities in Asia and universities in Europe?" The dependent variable is CWUR Score, the independent variable is Region, and this is a comparison of means/medians question.

1.4. Null hypothesis and alternative hypothesis (H0/H1) **(100 words)**

To address this research question, the study defines two hypotheses.
H0 (Null Hypothesis): There is no difference in the mean CWUR Score between universities in Asia and universities in Europe.
H1 (Alternative Hypothesis): There is a difference in the mean CWUR Score between universities in Asia and Europe.

Because CWUR Score is continuous and the region variable has two groups, a comparison-of-means test is required. Depending on normality, this will be either a t-test or Mann-Whitney U test.

2. Background research

2.1. Research papers (**200 words)**

Several studies in recent years used the CWUR data set to examine regional inequalities in university rankings. Kumar and Rajani in 2025 compared India's top 10 universities with world leaders in various university systems, including CWUR. The authors found that CWUR rankings showed a small performance divide between Indian and foreign universities, proposing that CWUR lean towards emerging regions in their rankings, given their emphasis on education quality and job readiness.

Zeng, in a study from 2024, analysed a period of a decade for CWUR rankings for top Asian universities like Tsinghua, NUS, and Tokyo. He found that though these universities experienced improvement in QS and THE rankings, their performance in CWUR rank improvement was relatively modest. This, of course, has to do with the distinct set of priorities in CWUR.

A study published by World Bank authors Demirgüç-Kunt and Torre in 2022 analysed how country-level CWUR rankings differ in a European and Central Asian perspective. Their results showed that, in CWUR rankings, European nations, especially the Netherlands, scored higher than Asian nations. This indicates a comparative advantage in CWUR ranking in favor of Europe.

Cumulatively, these studies suggest that CWUR reflects true but subtle regional variations, making it a good data set for investigating differences in average performance between Asia and Europe.

2.2. Why RQ is of interest? **(100 word**s)

Although studies of world rankings are common, there are very few that compare regions based on CWUR data. Most studies were done using QS,

THE, and/or ARWU data. There are also different metrics used in CWUR rankings, including alumni job and education quality, which result in different values for different world regions. Recent studies indicate that Asian institutions are making progress in certain world rankings, yet their improvement in CWUR rankings remains insignificant. This raises a question of whether there are significant differences in the average scores of Asia and Europe aggregated from CWUR data.

3. Visualisation
   3.1. Appropriate graphs for the RQ (**50 words)**

- Main plot: boxplot of CWUR (Centre for World University Rankings) Score by Region comparing Asia and Europe.
- Additional plot: histogram of CWUR Scores with normal curve overlay to inspect skewness and normality.
- Both plots are generated in R, with clear English titles, axis labels and legend.

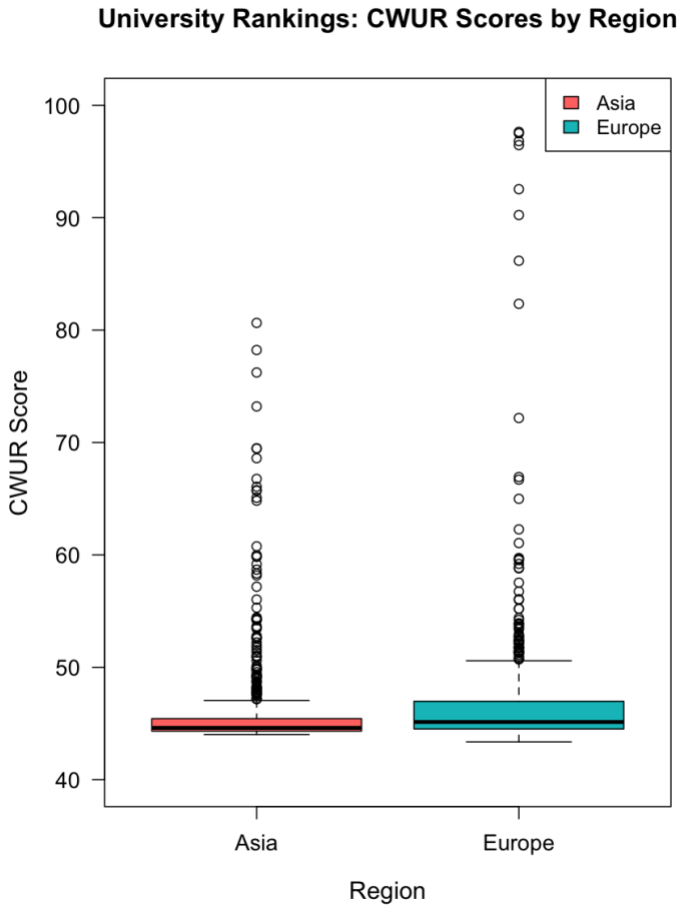

Figure 1 Boxplot of CWUR University Scores by Region (Asia and Europe).

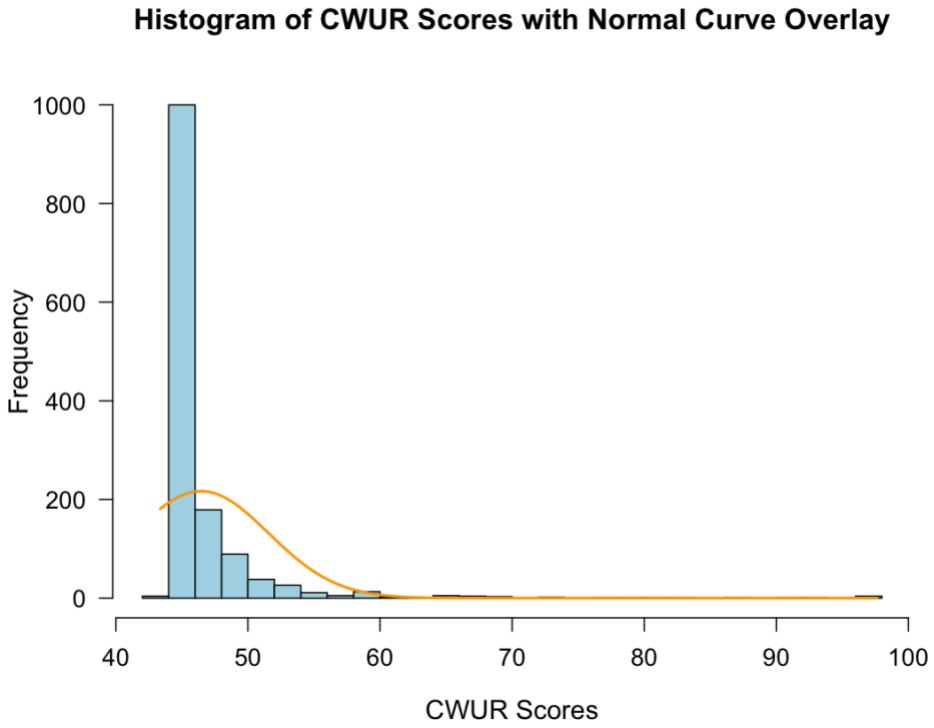

Figure 2 Histogram of CWUR (Centre for World University Rankings) Scores with normal curve overlay..

3.2. Additional information relating to understanding the data **(50 words)**

- Boxplot: compares CWUR Scores between Asia and Europe by showing medians, spread and outliers.
- Histogram: shows the overall distribution of CWUR Scores and how far it departs from the normal curve.

3.3. Useful information for the data understanding **(50 words)**

- The boxplot shows that European universities have a higher median CWUR Score and wider spread than Asian universities.
- Both regions contain high-scoring outliers, but most scores cluster in the mid-range.
- Because the distribution is skewed, medians are more informative than means and justify using a Mann-Whitney U test.

4. Analysis

4.1. Statistical test used to test the hypotheses and output **(75 words)**

This research compares the differences in CWUR Scores between two independent groups of institutions, those in Asia and those in Europe. Since

CWUR Score is a continuous data type, it would be analyzed using means and/or median values. Also, from a graph consisting of a histogram with a normal curve, it appears that this data distribution has a significant skew, which suggests that a non-parametric test alternative to the t-test, using the Mann-Whitney U test, would be most relevant in this analysis. The value of U for Score by Region was found to be about 146,660.5, with a p-value of 1.16×10^-19.

Statistical Test	Value
Wilcoxon rank-sum (Mann–Whitney U) test statistic (W)	176208
p-value	1.68×10^{-15} ($p < 0.05$)

Table 1 Statistical Test and Values

4.2. The null hypothesis is rejected /not rejected based on the p-value (100 words)

The null hypothesis states that there is no difference in CWUR (Centre for World University Rankings) Scores between universities in Asia and universities in Europe. The Wilcoxon rank-sum (Mann–Whitney U) test produced a test statistic of W = 176208 with a p-value of approximately 1.68×10^{-15}, which is far below the 0.05 significance level. Therefore, the null hypothesis is rejected. Although the research question refers to mean CWUR Scores, the distribution is skewed with outliers, so the Mann–Whitney U test appropriately compares the distributions (medians). The boxplot indicates that European universities tend to have higher CWUR Scores than Asian universities.

5. Evaluation – group's experience at 7COM1079

5.1. What went well **(75 words)**

First, our group formulated a research question that was not suited to the structure of the CWUR data set and suggested research question types. After receiving some feedback on booking slot sessions, we realized that our original research question was not suited for visualization and subsequent statistical testing. As a result, we changed our strategy and chose a research question that compared means/medians, Type 2, which suited our CWUR Score variable and our testing of two independent regions.

5.2. Points for improvement **(75 words)**

- Our initial research question was not well aligned with the CWUR dataset and required revision after feedback.
- More structured planning at an earlier stage would have reduced last-minute edits.

- Revisiting lecture slides and recorded sessions significantly improved our understanding of research question types.

- This helped us investigate the dataset more effectively and produce clearer, more meaningful visualisations aligned with the analysis requirements.

5.3. Group's time management (**50 words)**

There has been improvement in time management as the project progressed. The early activities were undertaken in a reactive manner, but then there has been a focus on tasks including data handling, R programming, literature review, and writing. Although there were a few analyses and editing tasks close to deadlines, deadlines were met for all necessary plot preparation, testing, and report writing**.**

5.4. Project's overall judgement (**50 words)**

On the whole, this project served its purpose. We focused on a specific research question using the prescribed CWUR dataset, produced relevant visualizations, and identified a proper test for our data that was non-parametric and matched our data distribution. This project has boosted our knowledge in forming a research question, identifying a statistical test, and presenting data in an organized way.

5.5. Comment on the GitHub log output **(50 words)**

Appendix B shows the full GitHub log. Key commits:

1. Commit Message: ["Did some testing in testing.R, Produced a final code, done data cleaning by grouping countries into two regions"] : Cleaned CWUR data and prepared regions for analysis.

2. Commit Message: [created histogram with Normal Curve Overlay, FinalVLSCode.R]: Cleaned CWUR data and prepared regions for analysis.
3. Commit Message: ["created Mann–Whitney U test (Wilcoxon rank-sum test), FinalVLSCode.R"] : Applied the appropriate non-parametric test to answer the research question.

6. Conclusions

6.1. Results explained (**75 words)**

A comparison of CWUR Scores among Asian and European universities was carried out. As depicted from the box plot, CWUR Scores for European universities are higher than those of Asian universities. Moreover, it was shown that CWUR Scores among European universities were not normally distributed because the histogram depicted skewed distribution. A result from the Wilcoxon rank-sum test, also known as Mann-Whitney U-test, showed that

$W = 176208$ and $p \approx 1.68 \times 10^{-15}$.

6.2. Interpretation of the results (**75 words)**

The very small p-value means we reject the null hypothesis and conclude that CWUR Scores for Asian and European universities are significantly different. The boxplot suggests that European universities tend to achieve higher CWUR Scores than Asian universities in this dataset. This finding is consistent with literature that reports regional differences in higher education performance and resources. It suggests that regional context, such as funding models and historical academic development, may influence how universities perform in global ranking systems like CWUR.

6.3. Reasons and/or implications for future work, limitations of your study (**50 words)**
This study is limited to one ranking system, one time, and a specific selection of Asian and European countries, so the results may not generalise widely. Future work could include other regions, compare multiple rankings, add further variables, and investigate which factors drive regional score differences.

7. Reference list

Centre for World University Rankings (CWUR) (2024) *World University Rankings – Methodology*. Available at: https://cwur.org

(Accessed: 12 December 2025).

Altbach, P.G. (2016) *Global perspectives on higher education*. Baltimore: Johns Hopkins University Press.

Marginson**, S.** (2016) 'Global university rankings and global higher education', *European Journal of Education*, 51(1), pp. 45–58.

Shin, J.C. and Kehm, B.M. (eds) (2013) *Institutionalization of world-class university in global competition*. Dordrecht: Springer.

Huang, F. (2022) 'The rise of East Asian universities in global rankings', *Higher Education Quarterly*, 76(4), pp. 864–883.

Zeng, Y**.** (2024) 'A study on the competitiveness of Asian universities using different world university rankings', *Proceedings of the 2nd International Conference on Social Psychology and Humanity Studies*, pp. 162–173.

Demirgüç-Kunt, A. and Torre, I**.** (2020) *Measuring human capital in Europe and Central Asia*. Policy Research Working Paper No. 9458. Washington, DC: World Bank.

Kumar, A. and Rajani, C. (2025) 'Indian higher education in global rankings: a critical review', *Higher Education Policy Studies*, 2(4), pp. 4526–4540.

O'Neill, M**.** (2016) *World University Rankings*. Kaggle dataset. Available at: https://www.kaggle.com/datasets/mylesoneill/world-university-rankings/data?select=cwurData.csv (Accessed: 12 December 2025).

8. Appendices

A.

R code used for analysis and visualisation ***(not included in the word count)***
Analysis.R code with the appropriate statistics to test the hypotheses.

R File name - FinalVlsCode.R

```
# 7COM1079 Final Report-R Code


library(readr)
library(dplyr)
cwur <- read_csv("cwurData.csv")

# Define region groupings using the full lists
asia_countries <- c(
  "Japan", "Israel", "South Korea", "Singapore",
      "China",
  "Taiwan", "Hong Kong", "Thailand", "Malaysia",
      "India",
  "Turkey", "Saudi Arabia", "Iran", "Lebanon",
      "United Arab Emirates"
)

europe_countries <- c(
  "United Kingdom", "Switzerland", "France",
      "Sweden", "Italy",
  "Germany", "Netherlands", "Finland", "Norway",
      "Denmark",
  "Belgium", "Spain", "Ireland", "Austria",
      "Portugal",
  "Czech Republic", "Greece", "Hungary", "Poland",
      "Iceland",
  "Slovenia", "Estonia", "Croatia", "Slovak
      Republic",
  "Bulgaria", "Lithuania", "Romania", "Cyprus",
      "Serbia"
)

# Add Region column and keep only Asia/Europe with
      real scores
asia_europe <- cwur %>%
  mutate(
    Region = case_when(
      country %in% asia_countries   ~ "Asia",
      country %in% europe_countries ~ "Europe",
      TRUE                          ~ NA_character_
    )
  ) %>%
```

```
  filter(Region %in% c("Asia", "Europe"),
         !is.na(score))

# Make Region an ordered factor (Asia, then Europe)
asia_europe$Region <- factor(asia_europe$Region,
                             levels = c("Asia",
     "Europe"))

print(table(asia_europe$Region))

# Create output folder (if it does not exist)
if (!dir.exists("outputs")) dir.create("outputs")

# ----------------------------
# BOXPLOT
# ----------------------------
png("outputs/boxplot.png", width = 1000, height =
     1200, res = 150)
par(mar = c(5, 6, 5, 4))
colors <- c("#ff7771", "#00bfc3");

boxplot(score ~ Region,
        data    = asia_europe,
        main    = "University Rankings: CWUR Scores
      by Region",
        xlab    = "Region",
        ylab    = "CWUR Score",
        col     = colors,
        notch   = FALSE,
        ylim    = c(40, 100),
        outline = TRUE,
        las     = 1,
        cex.lab = 1.2,
        cex.axis= 1.1,
        cex.main= 1.3)
legend("topright",
       legend = c("Asia", "Europe"),
       fill   = colors)
dev.off()

# ----------------------------
# HISTOGRAM + NORMAL CURVE
# ----------------------------
# Compute mean and sd using the cleaned Asia/Europe
     data
m <- mean(asia_europe$score)
s <- sd(asia_europe$score)

# Export to PNG (inside outputs folder)
png("outputs/hist_normal_curve.png", width = 1200,
     height = 900, res = 150)
```

```
par(mar = c(5, 6, 5,4))  # margins: bottom, left,
     top, right

# Histogram of real scores (frequency on y-axis)
h <- hist(asia_europe$score,
          breaks  = 30,
          freq    = TRUE,
          main    = "Histogram of CWUR Scores with
     Normal Curve Overlay",
          xlab    = "CWUR Scores",
          ylab    = "Frequency",
          col     = "lightblue",
          las     = 1,
          cex.lab = 1.2,
          cex.axis= 1.1,
          cex.main= 1.3)
# drawing the curve overlay
x_vals <- seq(min(asia_europe$score),
              max(asia_europe$score),
              length = 200)
y_vals <- dnorm(x_vals, mean = m, sd = s)
y_vals <- y_vals * diff(h$breaks)[1] *
     length(asia_europe$score)
lines(x_vals, y_vals, lwd = 2, col = "orange")

dev.off()

# ----------------------------
# Wilcoxon test
# ----------------------------
test_result <- wilcox.test(score ~ Region, data =
     asia_europe)
print(test_result)

sink("outputs/test_results.txt")
cat("Statistical Test Used: Mann-Whitney U Test
     (Wilcoxon rank-sum)\n")
cat("Test Statistic:", test_result$statistic, "\n")
cat("P-value:", test_result$p.value, "\n")
sink()
```

B. GitHub log output.

```
ommit c9ed4bfd831158e666e413018aa35c7e15ff5f96
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Fri Dec 12 02:22:33 2025 +0000

    deleted old testing plots and made some changes in final
code file

commit 57365fc29389d4a2af4faaa022105b4909669ab8
```

```
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Fri Dec 12 00:24:19 2025 +0000

    created Mann-Whitney U test (Wilcoxon rank-sum test),
FinalVLSCode.R

commit 4b1a9624ce791fc9a1dfbd90c834d9d580849fb0
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Fri Dec 12 00:14:39 2025 +0000

    created histogram with Normal Curve Overlay,
FinalVLSCode.R

commit 54fe704982e5b9a1018e1c58d29874045dbd2102
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Fri Dec 12 00:10:02 2025 +0000

    created boxplot and saved it to outputs folder,
FinalVlsCode.R

commit 66aaadd93f10d6baead2f0a743da98119973a150
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Fri Dec 12 00:07:58 2025 +0000

    created region variables and an ordered factor in
finalvls code

commit 6187fd8ccb73b158bc54b1c3a1e91c39e64f4019
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Thu Dec 11 22:47:10 2025 +0000

    Deleted unnecessary file

commit 1deab8c106ba0d735c78e8aa6ad3c3f20ff9a3d2
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Thu Dec 11 22:45:06 2025 +0000

    Did some testing in testing.R, Produced a final code,
done data cleaning by grouping countries into two regions

commit c3c83c8180a17ddc575029bbaccc8b8bc9724930
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Thu Dec 11 20:11:03 2025 +0000

    updated report file to section 5

commit 9f493cd21701bb1b525fede455543019d5047ce3
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Thu Dec 11 18:33:21 2025 +0000

    updated report file to section 4.2

commit 81e7cf2eff4764f76eb124d14ce76d3a207f7627
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Thu Dec 11 17:07:26 2025 +0000
```

```
    added demo feedback folder which also include our first
question which was rejected, the reason we had to another
research question

commit 4b3b14ec19c5d002a44ce4b46c41f0ae2587c8b9
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Thu Dec 11 16:48:05 2025 +0000

    code was changed to use package ggplot which is under
tidyverse make the plots more clear

commit f608f0d99c274d5ef984c2780cff7bfbeae18cc4
Author: Adeleye Adewuyi <aa24auo@herts.ac.uk>
Date:   Thu Dec 11 14:33:53 2025 +0100

    Deleted readme.txt

commit 5c2f1250b52ef6e642ea655e05e135e68fe3a953
Author: sixcodessometimes <aa24auo@herts.ac.uk>
Date:   Thu Dec 11 14:31:33 2025 +0100

    Updated the Heading on Readme

commit 3d186a3988e24475e4f24d821bb9ac0b9843879a
Author: sixcodessometimes <aa24auo@herts.ac.uk>
Date:   Thu Dec 11 14:30:21 2025 +0100

    Updated Readme.md

commit 0e381ca8074c158bc9e444a5014b2c7d3f3b1a74
Author: sixcodessometimes <aa24auo@herts.ac.uk>
Date:   Thu Dec 11 14:27:55 2025 +0100

    Changed Readme file from .txt to .md

commit f1a1165dd5495969cb8884aa7d37f8812352ec99
Author: sixcodessometimes <aa24auo@herts.ac.uk>
Date:   Thu Dec 11 14:03:27 2025 +0100

    Up-to-date Readme file

commit d70b800ba1193c4f39c80cbcbd02ea59f56e7608
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Thu Dec 11 10:26:31 2025 +0000

    testing values in plots

commit 61c0685e695aa7cca946d5d89d590e8bfef0de37
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Thu Dec 11 07:30:09 2025 +0000

    report file updated

commit 0ff037a5447880877aa92871a02bab19d2e0ef71
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Thu Dec 11 04:57:42 2025 +0000

    added git brief instruction file for other team members
```

```
commit 367ed285394903e500372aad5a143ed635b80d2c
Author: jadstrike <jadstrike@gmail.com>
Date:   Thu Dec 11 04:34:05 2025 +0000

    name changes to the file '7COM1079_Final
report_GpA28.docx

commit 6409b0067e9e2af809de600af2b4ecfc7fee39ce
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Thu Dec 11 04:26:14 2025 +0000

    report file added

commit 635a4a8ef33b8d9ab495228f89fd93b1b9a37dee
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Thu Dec 11 04:22:03 2025 +0000

    visualization changes related to comparison of means
after getting feedback

commit 0d6561558218b808ce085109011541e0b0872d04
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Wed Nov 26 14:06:32 2025 +0000

    updated code

commit d690debcf7b29e1ea7efd31ae5477fe54c4ef3d4
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Wed Nov 26 03:28:36 2025 +0000

    visualization slide added

commit 789ce5d7544432c30a5d77373573a1ed323bd902
Merge: ae21528 fabe21d
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Wed Nov 26 01:32:55 2025 +0000

    Merge branch 'main' of
https://github.com/sixcodessometimes/TeamProj

    merge readme file

commit ae215287c17e82d6092457274e26269fb617517c
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Wed Nov 26 01:32:23 2025 +0000

    initial visualization

commit fabe21d62a74a00767096c83f932128f801e1cc2
Author: Muhammad Taha Mirza <mm24aom@herts.ac.uk>
Date:   Wed Nov 26 01:18:23 2025 +0000

    added name

commit 02bf03874d8412773d4c99ddb7e08a22ecde6d76
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Tue Nov 25 22:28:14 2025 +0000
```

```
    added name

commit 44a59a41c271bd11484249abbc95d9e1d94728fd
Author: sixcodessometimes <aa24auo@herts.ac.uk>
Date:   Tue Nov 25 23:21:28 2025 +0100

    Add Names

commit fed05a06efa2f75de32a5558d2eeb029a47a101b
Author: Dare <adetomisinoluwaleye@gmail.com>
Date:   Tue Nov 25 23:17:10 2025 +0100

    Add Names

commit 9592ab587942a32fe19420a7ec86ec8147be6b2a
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Tue Nov 25 21:52:53 2025 +0000

    added readme

commit 17f791e7c310d62c1f181d48a64aef2ffbec52fb
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Tue Nov 25 21:50:16 2025 +0000

    research question slide added

commit 2dcfbc4a135feaf71f037d073f5d9602b43b2ac5
Author: KhantZweNaing <kn24abc@herts.ac.uk>
Date:   Tue Nov 25 21:21:04 2025 +0000

                initial_setup
```