7COM1079-0901-2024 - Team Research and Development Project

Final report title: Exploring the relation Between GDP per Capita and Happiness: An Analysis of the 2017 World Happiness Report

Group ID: A117

Dataset number: DS008

Prepared by: [Sravani Kumari Palem – 22067322,

Chaithanya Krishna Nuthalapati – 23067825,

Sai Ganesh Aavula – 24000424,

Gopala Krishna Reddy Avuluri – 23036787,

Gopala Krishna Pothumarthi – 23080478]

University of Hertfordshire

Hatfield, 2024

Table of Contents

1. Introduction 3
   1. Problem statement and research motivation 3
   2. The data set 3
   3. Research question 3
   4. Null hypothesis and alternative hypothesis (H0/H1) 3

1. Background research 4
   1. Research papers (at least 3 relevant to your topic / DS) 4
   2. Why RQ is of interest (research gap and future directions according to the literature) 4

1. Visualisation 4
   1. Appropriate plot for the RQ *output of an R script (NOT a screenshot)* 5
   2. Additional information relating to understanding the data (optional) 5
   3. Useful information for the data understanding 5

1. Analysis 5
   1. Statistical test used to test the hypotheses and output 5
   2. The null hypothesis is rejected /not rejected based on the p-value 5

1. Evaluation – group’s experience at 7COM1079 6
   1. What went well 6
   2. Points for improvement 6
   3. Group’s time management 6
   4. Project’s overall judgement 6
   5. Comment on GitHub log output 6

1. Conclusions 7
   1. Results explained. 7
   2. Interpretation of the results 7
   3. Reasons and/or implications for future work, limitations of your study 7

1. Reference list 7

Harvard (author, date) format.

1. Appendices 7
2. R code used for analysis and visualisation. 7
3. GitHub log output. 9

**1. Introduction**

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

Happiness is a mix of many factors, including social, economic, and cultural influences. Understanding what makes people happy is important for governments and policymakers, as it will helps them create better programs to improve people’s lives. One big question is how much a country’s economy affects happiness. This study looks at GDP per capita, a common way to measure a country’s economy, and how it connects to happiness scores. Research (Helliwell et al., 2017) shows that economy does play a crucial role in happiness, but it’s not clear how strong this connection is. That’s why we’re exploring this topic further.

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

The dataset is from the 2017 World Happiness Report, which ranks 155 countries by their happiness levels. Happiness scores are based on a survey where people rate their lives. It includes data on factors like economy GDP per capita, social support, life expectancy, family and freedom. These factors help explain why some countries are happier than other countries and provide a good starting point for studying the connection between a country’s economy and happiness score.

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

Is there a correlation between economy GDP per capita and happiness scores across countries?

* Independent variable (Interval/Measurement data): Economy GDP per capita
* Dependent variable(Interval/Measurement data): Happiness Score

To find out, we use visualization plots and apply Pearson’s correlation test. This will help us to measure the strength and direction of the relationship between GDP per capita and happiness scores.

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

**Null Hypothesis (H0):** There is no significant correlation between GDP per capita and happiness scores across countries. Any observed relationship is purely due to chance.

**Alternative Hypothesis (H1):** There is a significant correlation between GDP per capita and happiness scores across countries. This suggests that changes in GDP per capita are associated with variations in happiness levels.

Pearson’s correlation test will be used to evaluate these hypotheses. If the p-value is greater than 0.05, we fail to reject the null hypothesis, indicating no significant relationship between the variables. Otherwise, we reject the null hypothesis, supporting a correlation.

1. **Background research**
   1. **Research papers (at least 3 relevant to your topic / DS) (200 words)**

The World Happiness Report dataset is often used in studies to understand what influences happiness and how it relates to economic and social factors.

**De Neve and Oswald (2012)** studied the connection between happiness and future income using data from siblings. They found that people with higher life satisfaction and positive emotions are more likely to earn more in the future. This shows that happiness doesn’t just depend on income—it can also influence it.  
**Stevenson and Wolfers (2013)** examined income and happiness across countries, focusing on GDP per capita. They found that people in richer countries tend to be happier, showing a positive link between income and happiness. However, they also noticed that the effect of income on happiness becomes smaller as income increases, meaning money isn’t the only factor for well-being.  
**Helliwell, Huang, and Wang (2017)** studied how social factors affect happiness using the same data. They discovered that trust, strong social ties, and good governance are important for happiness, proving that many things beyond money matter.  
  
These studies highlight how useful the dataset is for understanding the different factors that influence happiness and ways to improve well-being globally.

* 1. **Why RQ is of interest** **(research gap and future directions according to the literature)** **(100 word**s)

This research looks at the connection between GDP per capita and happiness scores, a topic of growing interest in economics and social studies. Studies like Stevenson and Wolfers (2013) find a positive link, but others suggest that happiness stops increasing as incomes rise and depends more on trust and good governance. This shows that happiness is complex and needs to be studied from different perspectives like family, health etc. Using data from the 2017 World Happiness Report, this study re-examines the relationship. Better understanding these factors can help policymakers create balanced strategies to improve well-being for people across the world.

1. **Visualisation**
   1. **Appropriate plot for the RQ** (**50 words)**

The scatter plot is ideal for showing the relationship between GDP per capita (X-axis) and happiness scores (Y-axis). Each data point represents a country, while the linear trend line (red) highlights the positive correlation. The plot includes an informative title, axis labels, and units, clearly visualizing the strength and direction of the relationship.

**Titles and Axes:**

Title**:** Correlation between GDP per Capita and Happiness Score

X-axis: GDP per Capita (Units: Scaled)

Y-axis: Happiness Score (0-10)

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

**Scatter Plot**: The scatter plot shows the relationship between GDP per Capita and Happiness Score. Each point represents a country, with the blue dots highlighting the observed data. The red trendline indicates a positive correlation, suggesting that countries with higher GDP per Capita generally have higher happiness scores.

**Histogram:** The histogram depicts the distribution of Happiness Scores across all countries. It reveals a normal distribution, with most countries scoring between 4 and 6. This suggests that while happiness varies, it is generally centred around moderate levels. The overlaying bell curve confirms the suitability of the data for correlation analysis.

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

The scatter plot reveals a clear positive correlation between GDP per capita and happiness scores, as countries with higher GDP tend to have higher happiness scores. The histogram shows a normal distribution of happiness scores, with the majority of countries falling between a moderate range of 4 to 6, indicating consistent global happiness levels.

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

The Pearson's product-moment correlation test was selected to evaluate the relationship between GDP per capita and happiness scores because both variables are continuous and follow a linear pattern, as visualized in the scatter plot. This test measures the strength and direction of the linear association between the two variables. The test produced a correlation coefficient of 0.8125, indicating a strong positive relationship, with a highly significant p-value (< 2.2e-16), confirming the appropriateness of the test for the research question.

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

Based on the p-value (< 2.2e-16), which is far below the significance level of 0.05, we reject the null hypothesis. This means there is sufficient evidence to support the alternative hypothesis that GDP per capita and happiness scores are significantly correlated. The Pearson correlation coefficient of 0.8125 indicates a strong positive relationship between the two variables. Higher GDP per capita is generally associated with higher happiness scores, suggesting that economic performance is an important factor in national well-being. However, this result does not imply causation and should be interpreted alongside other factors influencing happiness

1. **Evaluation – group’s experience at 7COM1079**
   1. **What went well (75 words)**

Our group successfully analysed the dataset, ran statistical tests, and created visualization results to answer the research question. We worked well together, collaboration between the group members was smooth and efficient, dividing tasks effectively. Using GitHub, we maintained version control, which ensured seamless integration of our contributions. Using R for data analysis and visualization went smoothly. Clear communication among the team members enabled us to solve problems, stay focused on research goals, and achieved the project results.

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

While our project was successful and reached project goals, we identified a few areas for improvement. We could have conducted a more in-depth exploration of additional factors influencing happiness, such as family, health or governance. However earlier planning for the data visualization and analyzation has saved the time. Communication among the group members was effective, though we could have been improved by scheduling more meetings. Finally, we ensuring equal contribution from all members would make the workload more balanced.

* 1. **Group’s time management (50 words)**

Our group managed time effectively by setting clear deadlines and milestones for each task. Early planning allowed for smooth progress and clear results. Although, there were moments of delay due to overlapping personal schedules, we have managed smooth and effective communication among the group members to meet the project goals.

* 1. **Project’s overall judgement (50 words)**

Overall, the project was a success, as we met our objectives and gained valuable insights into the relationship between GDP per capita and happiness. The results were well-supported by statistical evidence, and the group’s collaborative effort was evident in the quality of the analysis and the clarity of the presentation.

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

**Commit Message**: "added correlation between GDP per Capita and Happiness Score"  
**Explanation**: This commit introduces a key aspect of the project, establishing the correlation between GDP per capita and happiness score. It forms the foundation of the analysis and enhances the study’s rigor by emphasizing the importance of this relationship.

**Commit Message**: "added histogram"  
**Explanation**: The addition of a histogram visualizes the distribution of key variables. This helps better understand the data’s shape and outliers, aiding in deeper statistical analysis and improving data interpretation.

**Commit Message**: "research question corrected"  
**Explanation**: This commit is pivotal in clarifying the research question, aligning the focus of the project with the actual investigation. A well-defined research question ensures the study’s direction and provides clarity to the audience regarding the project's purpose.

1. **Conclusions**
   1. **Results explained (75 words)**

The analysis showed a strong positive correlation (r = 0.8125, p < 0.001) between GDP per capita and happiness scores, as seen in the scatter plot, leading us to reject the null hypothesis. The histogram of happiness scores showed a normal distribution, with most countries having moderate happiness levels and fewer with very high or low scores. These results confirm that economic performance plays a significant role in happiness, but other factors like governance, social support, health, also matter.

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

The findings highlight that higher GDP per capita correlates with increased happiness, reinforcing the role of economic factors in well-being. The normal distribution of happiness scores suggests that the relationship applies broadly across countries, though diminishing returns occur at higher income levels. This implies that beyond a certain point, non-economic aspects, such as trust and freedom, are critical for further improvements in happiness. These insights can help policymakers focus on comprehensive well-being strategies.

* 1. **Reasons and/or implications for future work, limitations of your study (50 words)**

Future research should analyse non-economic factors, such as governance and mental health, to better understand their impact on happiness. The study’s reliance on the 2017 dataset and limited variables is a constraint. Including data from multiple years and exploring diverse factors could provide a deeper understanding of global happiness trends.

1. **Reference list**

Harvard (author, date) format.

* Stevenson, B. and Wolfers, J., 2013. Subjective well-being and income: Is there any evidence of satiation?. *American Economic Review*, *103*(3), pp.598-604.
* Helliwell, J.F., Huang, H. and Wang, S., 2017. The social foundations of world happiness. *World happiness report*, *8*, pp.8-46.
* Stevenson, B. and Wolfers, J., 2013. Subjective well-being and income: Is there any evidence of satiation?. *American Economic Review*, *103*(3), pp.598-604.

1. **Appendices**
2. R code used for analysis and visualisation ***(not included in the word count)***

Analysis.R code with the appropriate statistics to test the hypotheses.

**Plot1: Scatter Plot**

#add the data set

data<-read.csv("./Assests/2017.csv")

# Check for missing values in the relevant columns

sum(is.na(data$Happiness.Score)) # Missing values in Happiness.Score

sum(is.na(data$Economy..GDP.per.Capita.)) # Missing values in Economy..GDP.per.Capita.

plot(data$Economy..GDP.per.Capita., data$Happiness.Score,

main = "Correlation between GDP per Capita and Happiness Score",

xlab = "GDP per Capita (Units: Scaled)",

ylab = "Happiness Score (0-10)",

pch = 19,

col = "blue")

# Add a trend line (linear regression line)

abline(lm(Happiness.Score ~ Economy..GDP.per.Capita., data = data), col = "red")

legend("topleft", legend = c("Data Points", "Regression Line"),

col = c("blue", "red"), pch = c(16, NA), lty = c(NA, 1), lwd = 2)

# Pearson correlation

correlation\_result <- cor.test(data$Happiness.Score, data$Economy..GDP.per.Capita., method = "pearson")

print(correlation\_result)

**Plot2: Histogram**

#add the data set

data<-read.csv("./Assests/2017.csv")

print(data)

hist(data$Happiness.Score,

main = "Histogram of Happiness Score with Normal Curve Overlay",

xlab = "Happiness Score",

ylab = "Density",

col = "lightblue",

border = "black",

freq = FALSE) # Set freq = FALSE for density plot

# Add the normal distribution curve

x\_vals <- seq(min(data$Happiness.Score)-1, max(data$Happiness.Score)+1, length = 100)

y\_vals <- dnorm(x\_vals, mean = mean(data$Happiness.Score, na.rm = TRUE),

sd = sd(data$Happiness.Score, na.rm = TRUE))

lines(x\_vals, y\_vals, col = "red", lwd = 2)

1. GitHub log output.

GitHub <URL:-> <https://github.com/SaiGaneshAavula/A117.git>

commit d3db2fc370a6ea892d13022b4b7223bc3ea2bd0c

Author: Chaithanyakrishna852 <cn24aal.herts.ac.uk>

Date: Sun Jan 5 17:33:47 2025 +0000

added why RQ is interesting

commit 05ac1349ecc9dabf3e12d8a455a5803eb7347de8

Author: Chaithanyakrishna852 <cn24aal.herts.ac.uk>

Date: Thu Jan 2 18:58:16 2025 +0000

added description for null and alternative hypothesis

commit 8af6fbaa1f549ceb11e17a62f15e6cfc9c8e079a

Author: SaiGaneshAavula <sa24aja.herts.ac.uk>

Date: Thu Jan 2 14:23:31 2025 +0000

Explaining how we are going to answer RQ.

commit 90f8b0488328189abf50e5802b132335cc500ad2

Author: SaiGaneshAavula <sa24aja.herts.ac.uk>

Date: Thu Jan 2 13:58:58 2025 +0000

data set description added

commit 9eef1ec818ac89c078799098090241fe1ad53b15

Author: SaiGaneshAavula <sa24aja.herts.ac.uk>

Date: Thu Jan 2 13:50:01 2025 +0000

problem statement and research motivation added

commit 015f36500bbd523c594e79044538b0ace179cc8b

Author: SaiGaneshAavula <sa24aja.herts.ac.uk>

Date: Thu Jan 2 13:42:03 2025 +0000

research question corrected.

commit 389580d76a5b9c0f0ff042117a0aca0c9159121b

Author: Sravani-Kumari-Palem <sp24aau@herts.ac.uk>

Date: Thu Jan 2 13:19:10 2025 +0000

research question added

commit 9b2ddbd53e16a4d5ec94a5c4a7c048704d14a6bd

Author: Sravani-Kumari-Palem <sp24aau@herts.ac.uk>

Date: Wed Jan 1 18:32:44 2025 +0000

Doc added and group Id added in doc

commit ba6323f6dfd0e89db4f1f1e6f1357661198b078a

Author: Sravani-Kumari-Palem <sp24aau@herts.ac.uk>

Date: Fri Nov 22 08:47:26 2024 +0000

added normal distributed curve

commit 1540d97331feef6114bcda03bbc89bc86b5e43ac

Author: ga23acb <ga23acb@herts.ac.uk>

Date: Fri Nov 22 08:37:00 2024 +0000

added histogram

commit e7d38477aa0b6e67482fbb158311dd27a57032a2

Author: ga23acb <ga23acb@herts.ac.uk>

Date: Thu Nov 21 20:26:32 2024 +0000

created visualisation file and dataset reading

commit 7558c37b965429100b1105fdc1fd0750dc3f24e2

Author: Gopalakrishnapothumarthi <gp24aas@herts.ac.uk>

Date: Thu Nov 21 08:02:08 2024 +0000

Pearson correlation test

commit 3830d04e45cb708e94ed1ac05cf2364a51377f12

Author: Gopalakrishnapothumarthi <gp24aas@herts.ac.uk>

Date: Thu Nov 21 07:50:06 2024 +0000

Check for missing values in the relevant columns

commit f14a4d2c94b20763d49d68a544fe4accdb1a00f8

Author: SaiGaneshAavula <sa24aja@herts.ac.uk>

Date: Wed Nov 20 23:36:50 2024 +0000

Adding linear regression line

commit b208da929ecc3f56a85b421d2f1bc46dc597e664

Author: SaiGaneshAavula <sa24aja@herts.ac.uk>

Date: Wed Nov 20 23:34:38 2024 +0000

Correlation between GDP per Capita and Happiness Score

commit 0d2fbf5a2a1f902599b1f25e1bb5d01203fe2c72

Author: Chaithanyakrishna852 <cn24aal@herts.ac.uk>

Date: Wed Nov 20 22:17:04 2024 +0000

Reading the dataset

commit 10386d46ac439485fef0bd6916a758b92cb7f609

Author: Chaithanyakrishna852 <cn24aal@herts.ac.uk>

Date: Wed Nov 20 15:53:44 2024 +0000

added first comment

commit cb30d1592cad1697445b2f86dbe8f970f5ecd813

Author: Sravani-Kumari-Palem <sp24aau@herts.ac.uk>

Date: Wed Nov 20 15:44:58 2024 +0000

correlation file created

commit 1ca39b8290cd3f73d11ea87ced90daa45c930579

Author: Sravani-Kumari-Palem <sp24aau@herts.ac.uk>

Date: Wed Nov 20 13:22:11 2024 +0000

Dataset added

(END) Intial Commit with Dataset