

Achieving Gender Equality: Strategies for Advancing SDG 5

By: Jordan Chang, Ivan Chepelev, Abeera Fatima,
and Alisa Iskakova

What factors are
related to higher
gender equality
scores?

Inspiration for our Guiding Question

With our guiding question, we aim to explore the progress of different countries regarding the 5th SDG goal: gender equality, and discern the distinguishing factors that set successful countries apart from others. By delving into various socio-economic, cultural, and political factors, we seek to uncover the underlying mechanisms that drive progress in this crucial area to identify the strategies and initiatives most effective in fostering gender equality. Ultimately, our goal is to contribute to the ongoing global discourse on gender equality and empower communities to enact meaningful change by identifying the best practices and actionable insights to support progress towards gender equality.



Question 1

Research Question 1

Research Question One: What are the common characteristics found in countries with high gender equality scores that are lacking in countries with low gender equality scores?

- Our first research question aimed to uncover shared characteristics among countries that have achieved SDG 5 and determine whether countries lacking these characteristics exhibit lower gender equality scores.
- If such characteristics are found, we could recommend countries with lower scores to prioritize improvement in these areas in order to ultimately enhance their gender equality scores.

Data and Methods

Data:

UNICEF's 2023 State of the World's Children report:

- Gdp per capita
- Gini Coefficient
- Basic Drinking Water Services

Lost lots of variables during the data cleaning process.

Methods:

2-sample, 1-sided hypothesis test

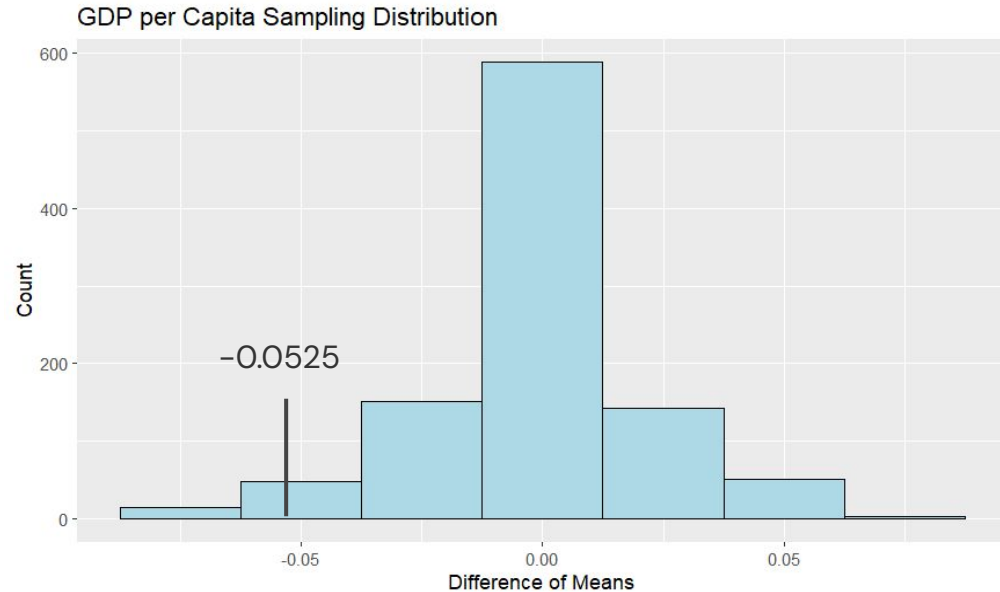
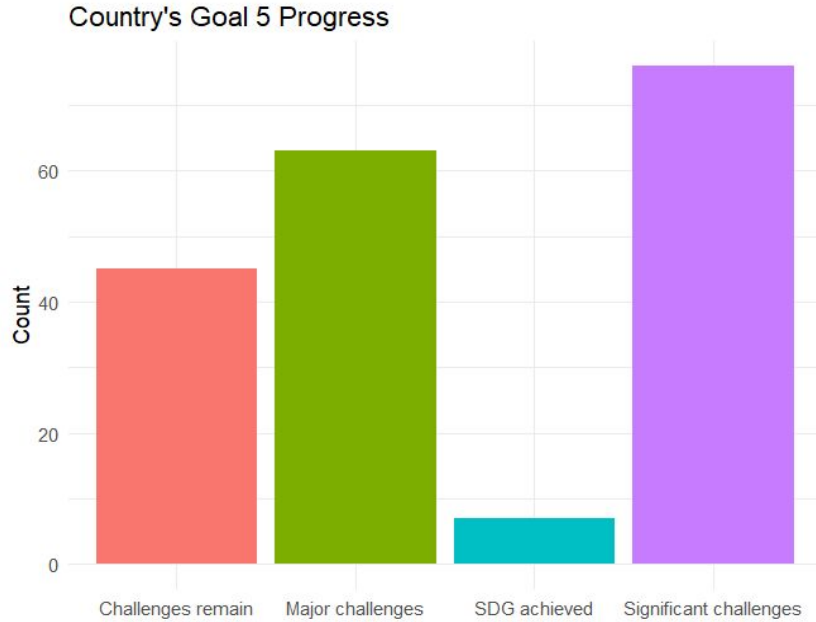
- First, we split the countries into two different groups (Achieved SDG vs Hasn't yet)
- Then, we compared the mean (μ) test score for each variable between the groups and observed the p-value.
- If $p\text{-value} < 0.05$ (α), we found there to be a significant difference in the mean test scores between the two groups.



Visualizations and Results

Observed Test Statistic: -0.0525

P-value: 0.064



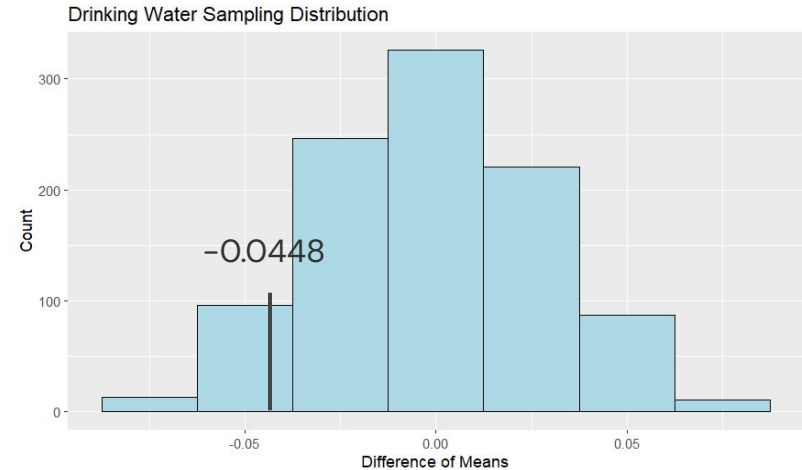
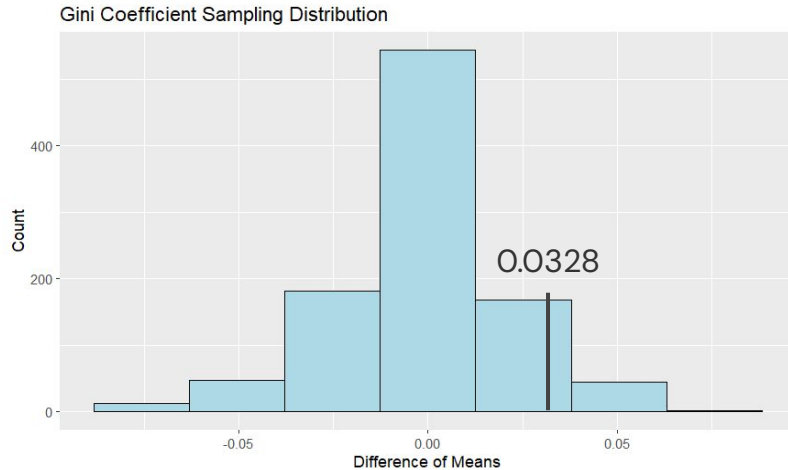
Visualizations and Results

Observed Test Statistic: 0.0328

P-value: 0.215

Observed Test Statistic: -0.0448

P-value: 0.109



Analysis

- Predetermined **significance (α) level of 0.05**
- Because calculated p-values are all higher than this significance level we cannot “reject the null hypothesis”. This means, based on the data, we cannot confidently conclude any difference between the observed variables and gender equality.
- **Null hypothesis:** $H_0: \mu_H = \mu_L$,
 - μ is the portion of high gender equality countries in a group which is divided by high/low variable scores.
- GDP per capita and gender equality: **p-value = 0.064**
 - This is a “low” p-value, providing low/moderate evidence against the null hypothesis. However it is still not enough to reject the null hypothesis without **p-hacking**.
 - Type II error? (false-negative)
- Gini Coefficient and gender equality: **p-value = 0.215**
 - This is the highest p-value, providing the least evidence against the null hypothesis, implying there is little to no difference high and low gender equality and gini coefficients.

Question 2

Research Question 2

Research Question Two: Is there a relationship between each country's achievements in potentially related SDG goal scores and their gender equality scores?

- Our second research question investigates whether greater progress in potentially related SDG goals means possibly higher gender equality scores.
- The findings will illustrate the impact of other SDGs on gender equality and guide us in identifying areas for improvement that could subsequently improve gender equality.

Data and Methods

Data:

Scores and current status for the following SDG's

- No Poverty
- Zero hunger
- Quality education
- Decent work and economic growth
- Reduced inequalities
- Peace and justice and strong institutions

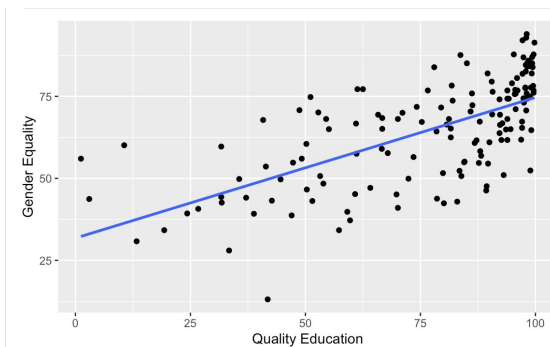
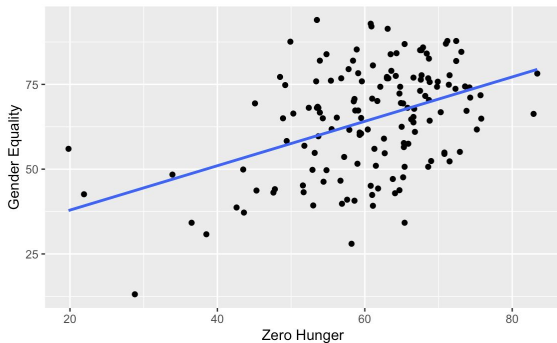
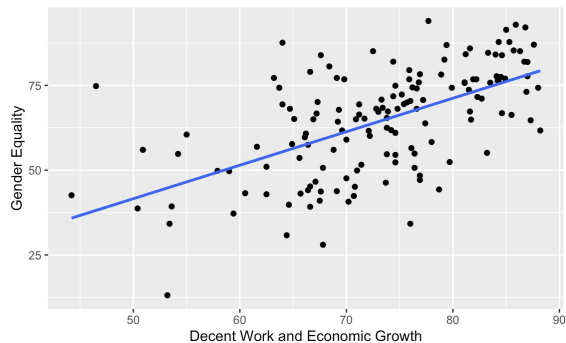
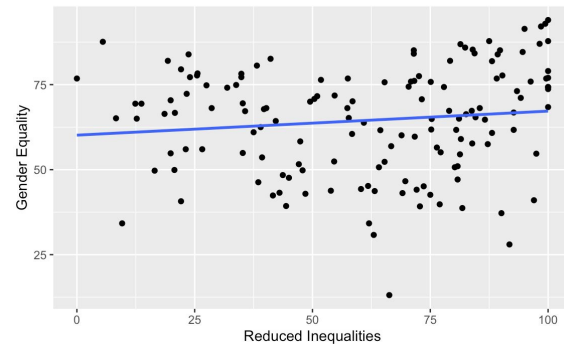
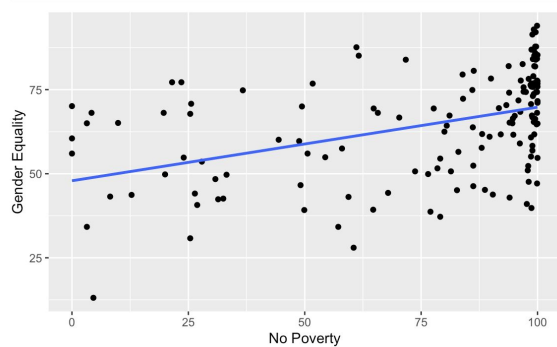
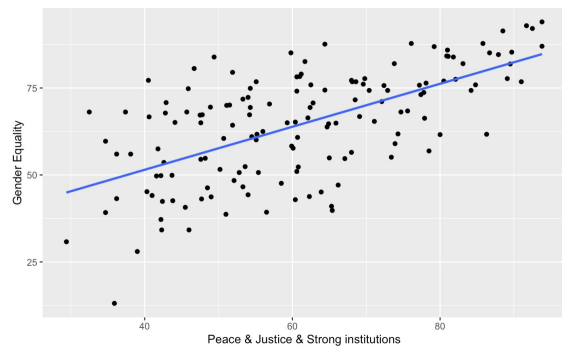
Methods:

Linear regression

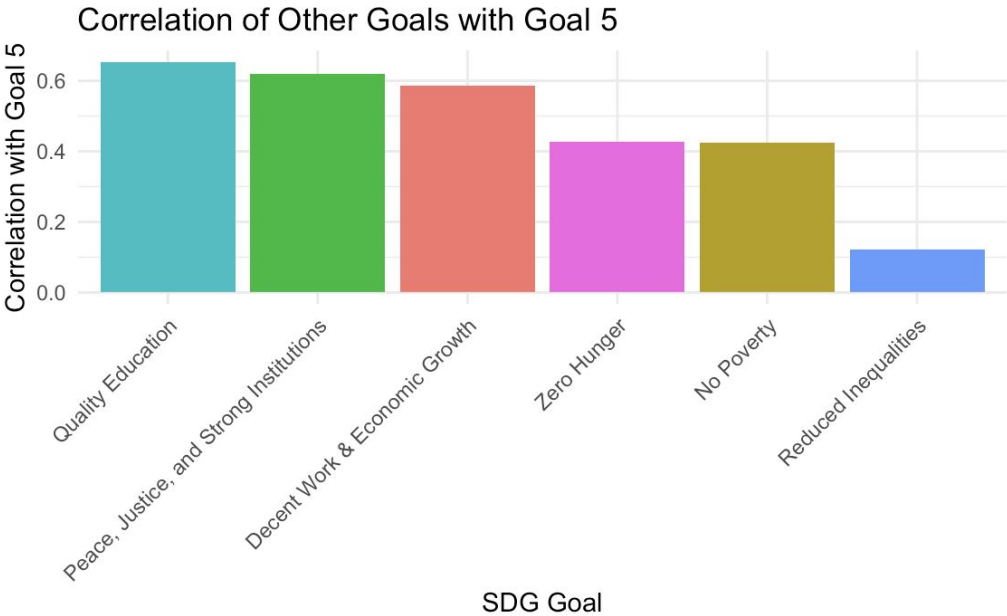
- We will use linear regression to fit a linear model on pairs of dataset points consisting of the SDG gender equality score and the score for a different SDG.
- Then we measured the linear association with r as a measurement of the direction and strength of the linear relationship between the two scores



Visualizations



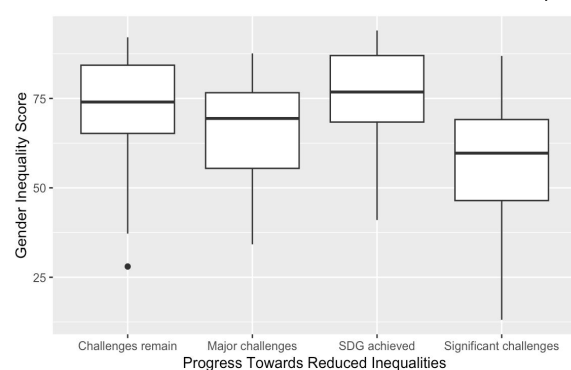
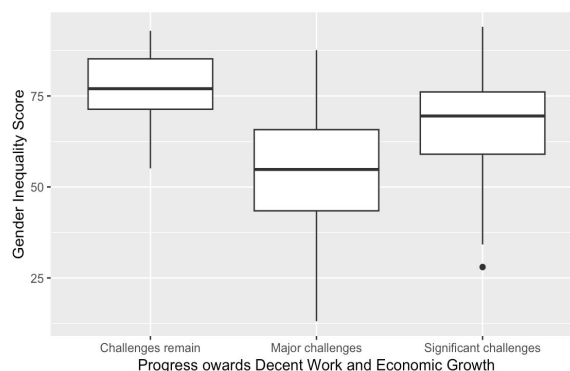
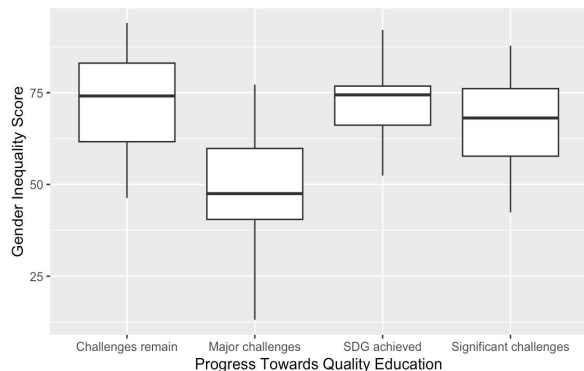
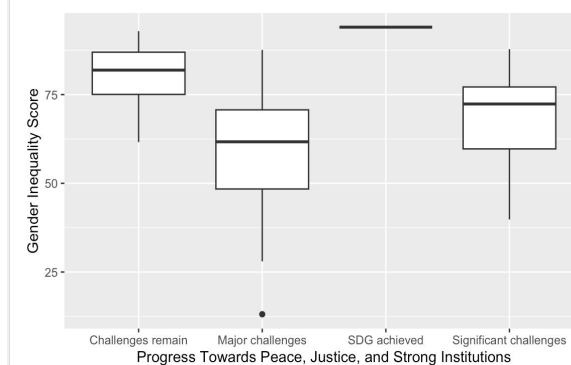
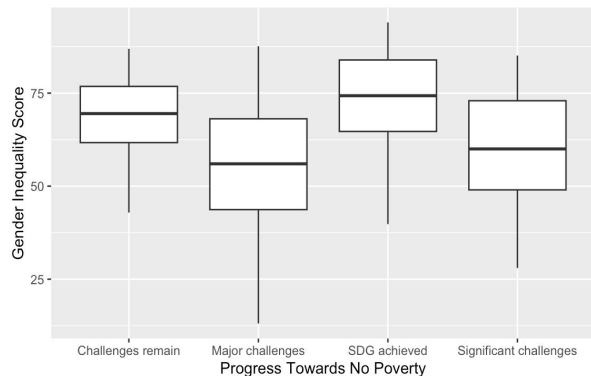
Visualizations



Goal<chr>	Correlation<dbl>
Quality Education	0.6541692
Peace, Justice, and Strong Institutions	0.6208285
Decent Work & Economic Growth	0.5855760
Zero Hunger	0.4265678
No Poverty	0.4259239
Reduced Inequalities	0.1214813

Goal<chr>	RMSE<dbl>
Zero Hunger	6.230121
Peace, Justice, and Strong Institutions	8.201461
Decent Work and Economic Growth	8.693385
Quality Education	17.208176
Reduced Inequalities	24.803779
No Poverty	27.924712

Visualizations



For the boxplots that show the SDG achieved category, we see that countries in that category have a higher median gender equality score than countries that haven't achieved it yet.

Results

Linear Regression models

- The linear regression models of the different SDG scores compared to Goal 5 illustrate that they all have a positive slope of generally the same magnitude. The only outlier seems to be Goal 10 (Reduced Inequalities) which has a lower magnitude than the rest

R-values

- Upon calculating the R values, we can see that Goal 10 has an R value of 0.1214813, which is significantly lower than the other goals. The ones with the highest R value are Goal 4 (Quality Education), followed by Goal 16 (Peace, Justice, and Strong Institutions)

RMSE values

- Calculating their respective RMSE values tells us that No Poverty and Reduced Inequalities have the highest RMSE values while Zero Hunger has the lowest

Boxplots

- By dividing the countries by their progress in each of these other goals, we see that countries that have less challenges left in these other areas have a higher median gender equality score

Analysis

- From the linear regression models, we can take away that generally, a higher score in other goals is associated with a higher score in Goal 5.
- We can quantify this relationship by calculating the R and RMSE values. Comparison of those tells us that high Quality Education and Peace, Justice, and Strong Institutions are associated the most with high Gender Equality.
- Goal 10 (Reduced Inequalities) has the lowest R value, indicating that a high score in this goal is not closely related to a high score in Goal 5. However, it also has the highest RMSE value, indicating that the prediction error for this goal is relatively large.
- Generally, we can take away from these visualizations that countries can potentially raise their score in Goal 5 by taking steps to raise their scores in all other goals, but especially Goal 4 (Quality Education) and Goal 16 (Peace, Justice, and Strong Institutions)

Question 3

Research Question 3

Research Question 3: Can we predict the gender equality status based on the metrics determined relevant in question 1 and question 2?

- Our third research question was driven by the desire to investigate the possibility of creating a predictive model using classification trees.
- This model was aimed to determine whether it is feasible to predict the gender equality status based on other metrics
- We wanted to utilize the insights gathered from our analysis of factors related to gender equality in the first two inquiries.

Data and Methods

Data:

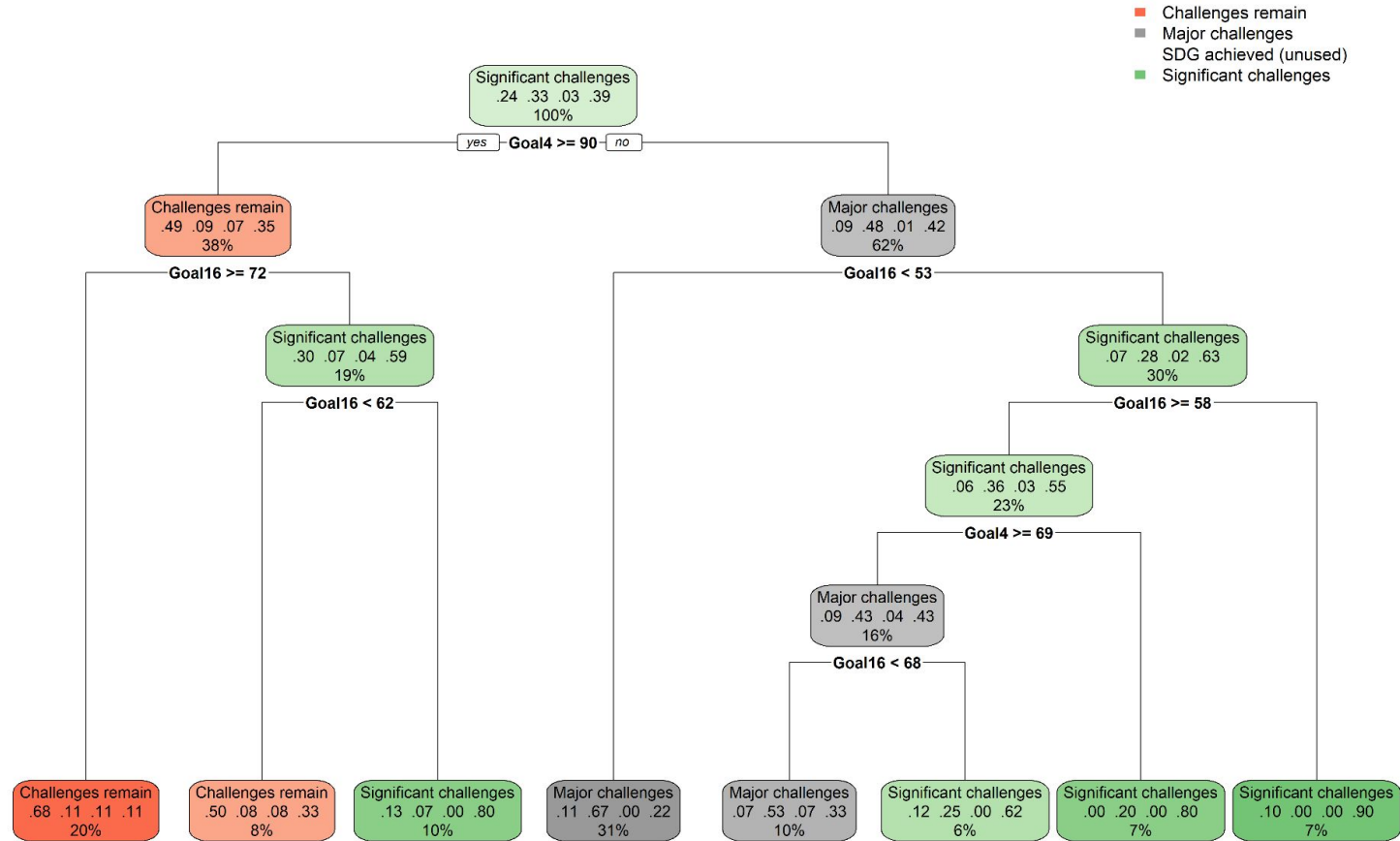
- We used statistics that were determined relevant from question 1 and question 2 to build a classification tree
- From question 2 we used to following metrics
 - Quality education (Goal 4)
 - Peace, Justice and Strong Institutions (Goal 16)
 - Decent Work & Economic Growth (Goal 8)
- Question 1 showed no significant links between the metrics explored in that question and the Gender Equality goal

Methods:

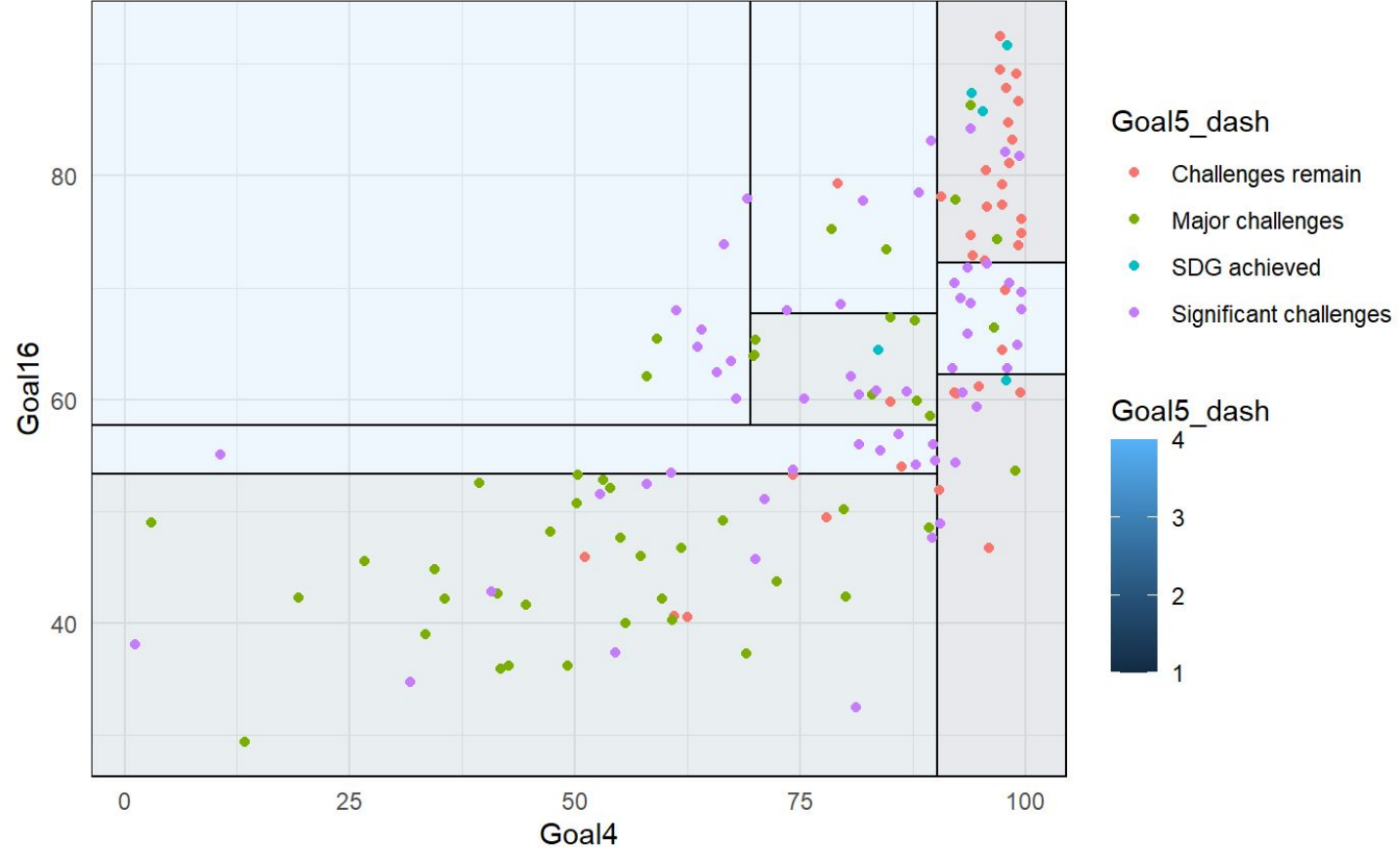
- We built a classification tree to predict gender equality status using key parameters identified in question 2.
- We split the data into a training and a testing dataset
 - We fitted the tree to the training dataset and then measured its accuracy using the testing dataset
- We created a graph to visualize partitions between every pair of goals used to fit the tree
- We created a confusion matrix to calculate the accuracy of the tree



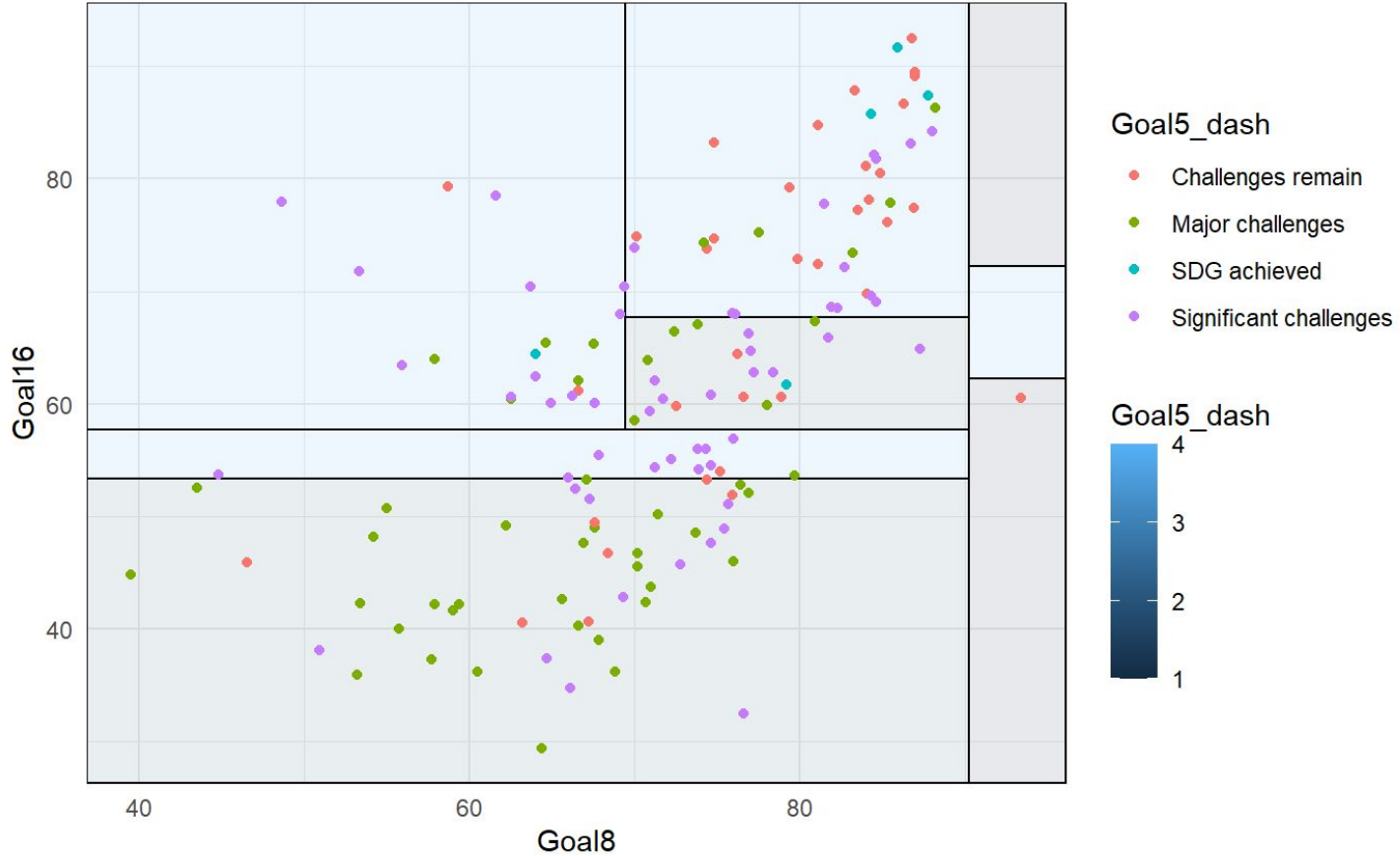
Visualizations



Visualizations



Visualizations



Results

- As a result of fitting we get a tree with 7 splits and 8 terminal nodes
- Based on the labels on the terminal nodes produced by the rpart.plot package we can judge the purity of the terminal nodes, which indicates the effectiveness of the split
- The tree is missing one of the categories we tried to predict
- We computed the following confusion matrix

predictions	Challenges remain	Major challenges	SDG achieved	Significant challenges
Challenges remain	5	1	1	5
Major challenges	2	8	0	5
SDG achieved	0	0	0	0
Significant challenges	1	2	0	6

Analysis

- Since the tree result is missing one of the predicted categories, namely “SDG achieved” we can say that the tree is unable to classify countries into the “SDG achieved” category
 - The reason for this, is the lack of data on the countries classified as “SDG achieved” in our dataset
 - At no split was “SDG achieved” the vote winner, so it doesn’t appear even in the terminal nodes
- Using the confusion matrix we can calculate the accuracy of the tree to be 53%
 - Since we are predicting 4 categories the tree is 28% more likely to predict the correct category than random guessing
- Therefore we can conclude that Quality education (Goal 4), Peace, Justice and Strong Institutions (Goal 16) and Decent Work & Economic Growth (Goal 8) are good predictors of Gender Equality (Goal 5)

Final Conclusions

Big picture:

Generally, the following SDG scores are most associated with a high Goal 5 score:

- Quality education (Goal 4)
- Peace, Justice and Strong Institutions (Goal 16)
- Decent Work & Economic Growth (Goal 8)

Countries with low Gender Equality should be encouraged to focus on the above 3 goals as having a higher score in these goals is associated with a higher Gender Equality score.

Limitations:

- Besides other SDG Goals, we could not determine any common characteristics shared by countries that achieved Goal 5 since there was no significant difference in the gdp per capita, gini coefficient, and basic drinking water services of countries with achieved and not achieved Goal 5 scores.
- Due to a lack of data or classification tree could not predict the SDG achieved status and we only achieved a 53% accuracy rate.
- A big limitation we faced was a lot of NA values across data sets.

Final Project

THANK YOU

RESOURCES

Photos

- [Gender equality SDG Picture](#)
- [State of the World Children's Report](#)
- [SDG Poster](#)
- [Gender equality poster](#)