**Theoretical Framework**

I have decided to create a composite indicator to measure and compare education systems between countries.

From my research, I found the following:

* Higher spending per student is associated with better outcomes for student later in life, especially for those from low-income families. This gives access to school resources the potential to reduce intergenerational poverty. (Jackson et al. 2015)
* Building on this, it has been found that a lack of qualified teachers is a sign of staffing problems (Ingersoll 2001), which leads to poorer student performance.
* Exposure to high staff turnover leads to worse educational outcomes by students, although the effects of this disruption can be reduced through proper management. (Gibbons et al. 2021)
* Gender equality within education is one of the most consistent factors in improving student outcomes. (Campbell 2021)

**Data Selection**

I used three datasets to construct this index:

* A dataset from Kaggle, available [here](#_https%3A%2F%2Fwww.kaggle.com%2Fcode%2Fn), which contained data on enrolment rates, out-of-school rates, completion rates, maths and reading scores, and literacy rates. All of these categories were separated between male and female.
* Education data from the World Bank, which contained education spending as a percentage of each countries’ GDP, split between primary, secondary and tertiary education and the expected years in education for males, females and the average.
* Data from the UNESCO Institute of Statistics, which contained data on the number of compulsory years of education, the number of free years of education, the teacher: pupil ratio of primary and secondary education, the percentage of teachers who are qualified, and the attrition rates for teachers for pre-primary, primary and secondary education.

**Imputation of Missing Data**

Every variable I selected, with the exception of compulsory schooling had at least one missing value. I addressed them as follows:

* For all variables gathered from the UNESCO Institute of Statistics, the World Bank and the Kaggle dataset, if a country had data for a particular variable for any year from 2011 to 2020, but not for 2021, I took an average of the years available instead, as while it would likely not be completely accurate, it would serve as good approximation.
* For the variables related to free schooling, I assumed that no data meant that the respective country had none. I felt this was a reasonable solution as less than 25 countries were missing this information and mostly occurred with countries with no formal schooling.
* I decided to drop the variables related to teacher attrition, due to a lack of data. Less than 1/3 of countries had this data available for any of the previous 10 years.
* For the ratio between qualified teachers to pupils and the percentage of current teachers who are currently qualified, I replaced missing values with the global average.

**Multivariate Analysis**

After carrying out multivariate analysis, some variables were very strongly correlated with each other. To address this, the following changes were made:

* Principle Component Analysis was used to combine the six completion rate variables. This resulted in one component that explained 95% of the variance.
* Principle Component Analysis was used to combine the out-of-school rate variables. This resulted in three components that explained a combined 95.5% of the variance.
* The expected years of education for males and females were changed into one variable that represented the difference between the two.
* This was repeated for the literacy rate variables.
* The average of the math and reading scores was used instead of the two being separate.

**Normalisation**

All variables were centred and scaled to be between -1 and 1.

**Weighting and Aggregation**

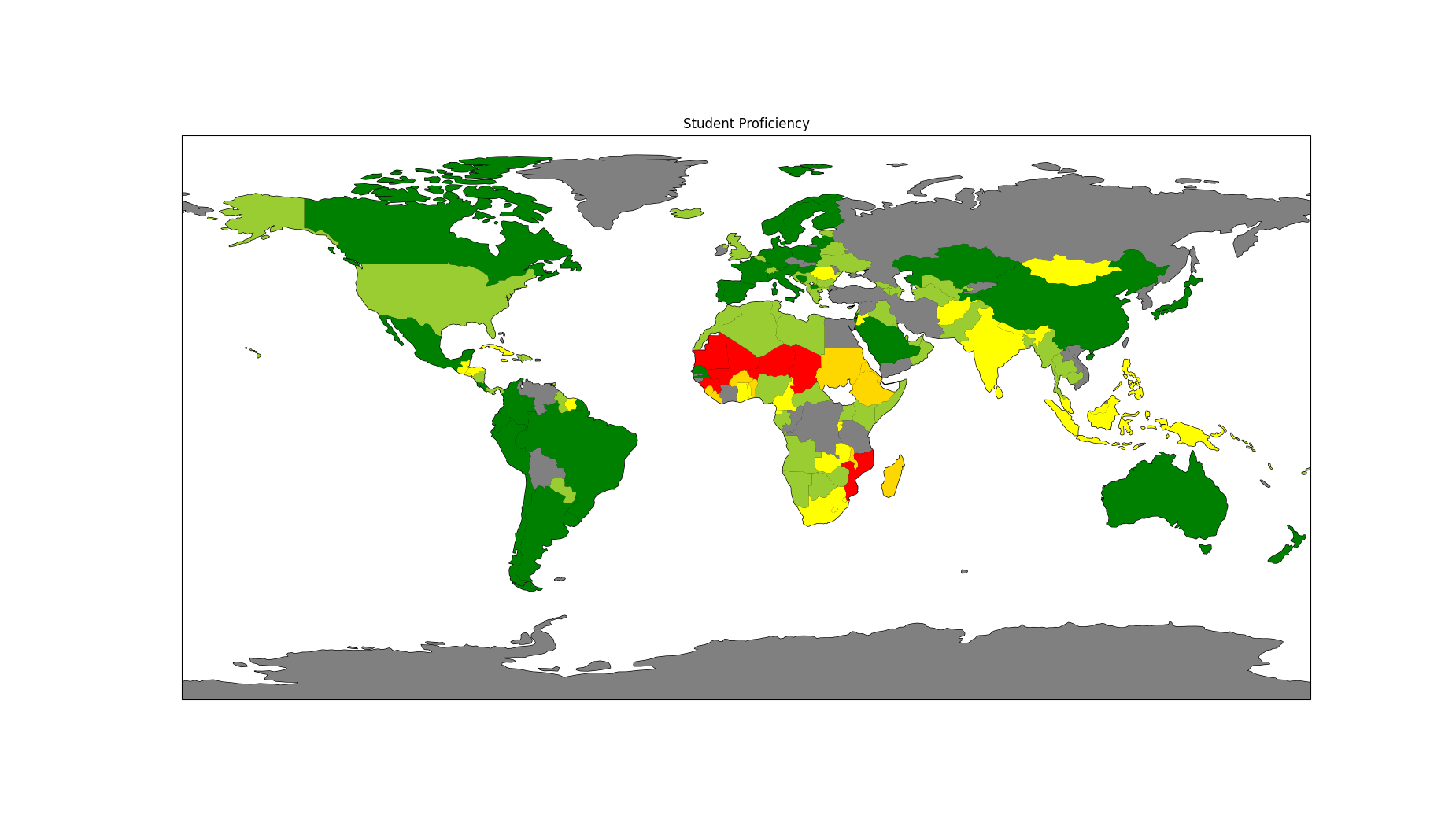
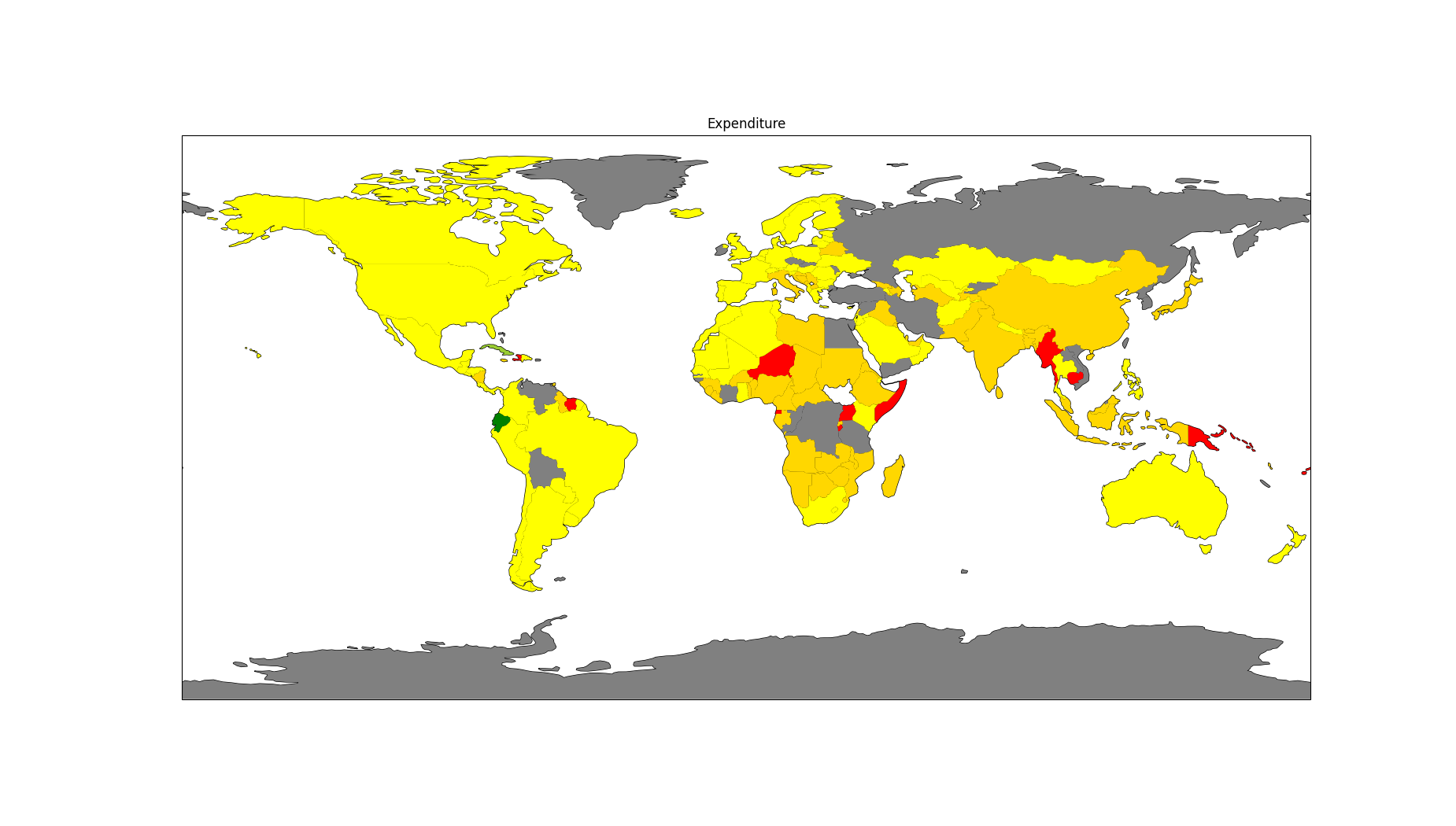
I decided to give each variable equal weighting as I could no find any research-backed way to rank the importance of one variable over another.

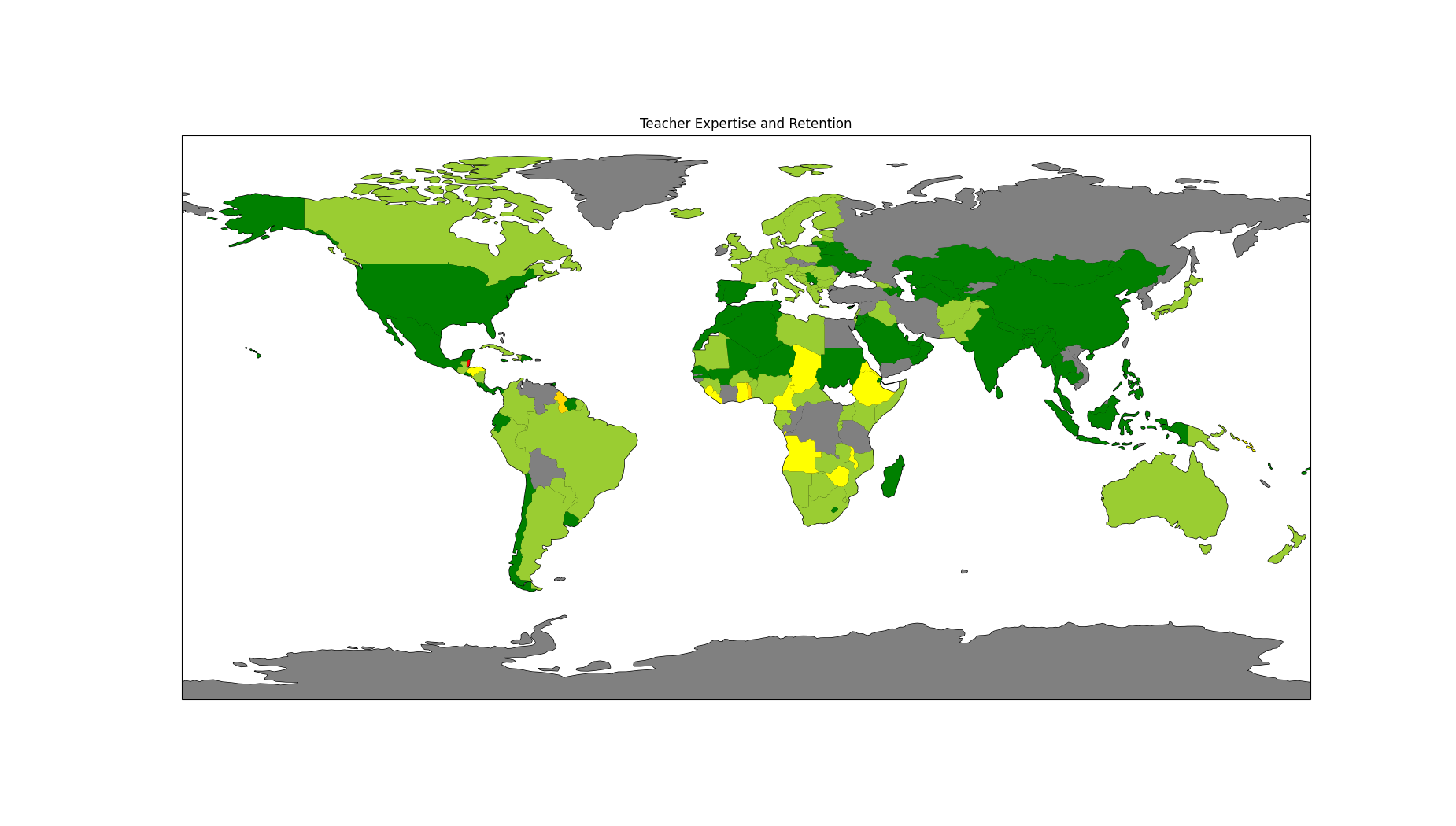
For aggregation, I divided the variables between five sub-indicators as follows:

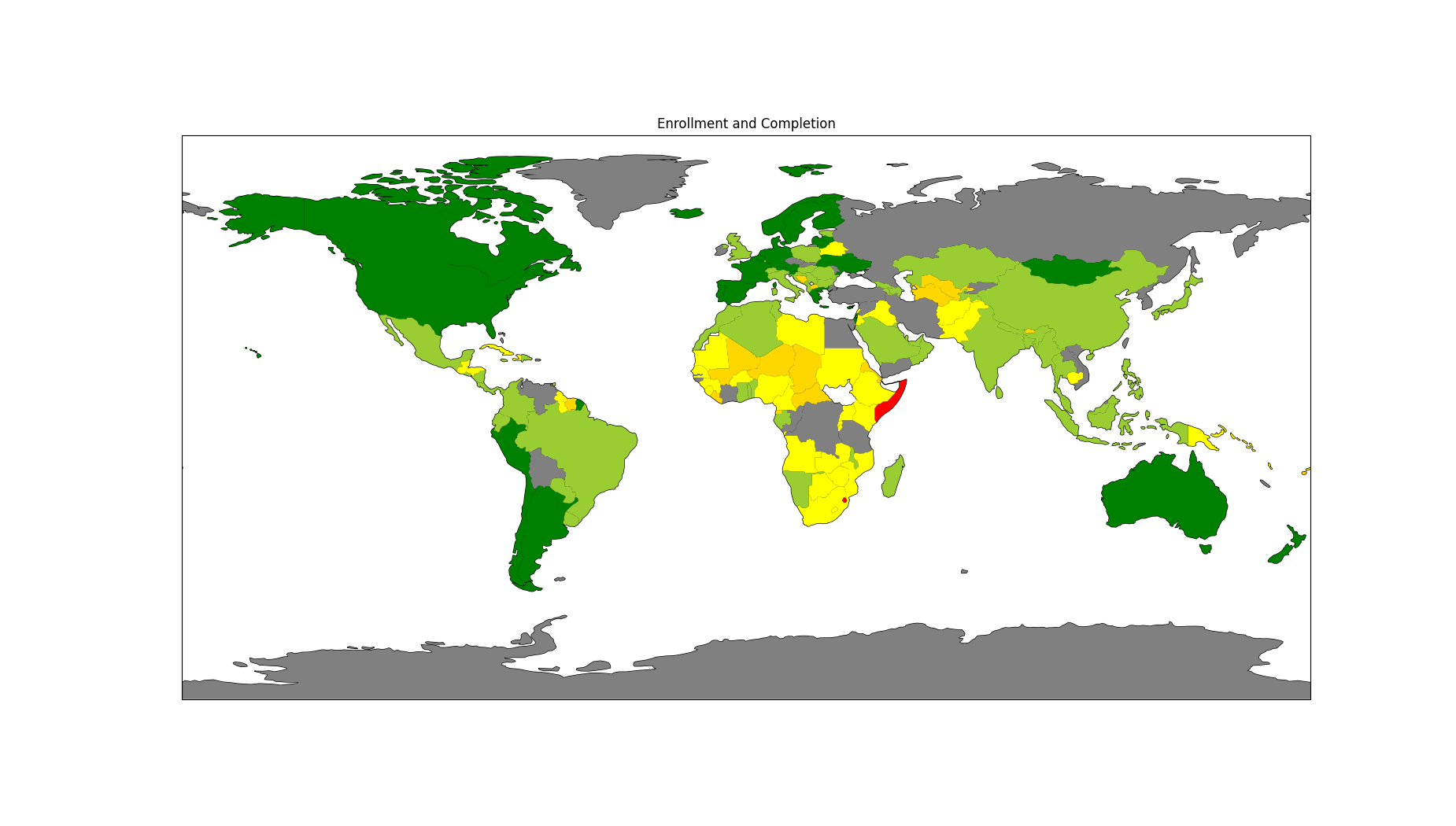
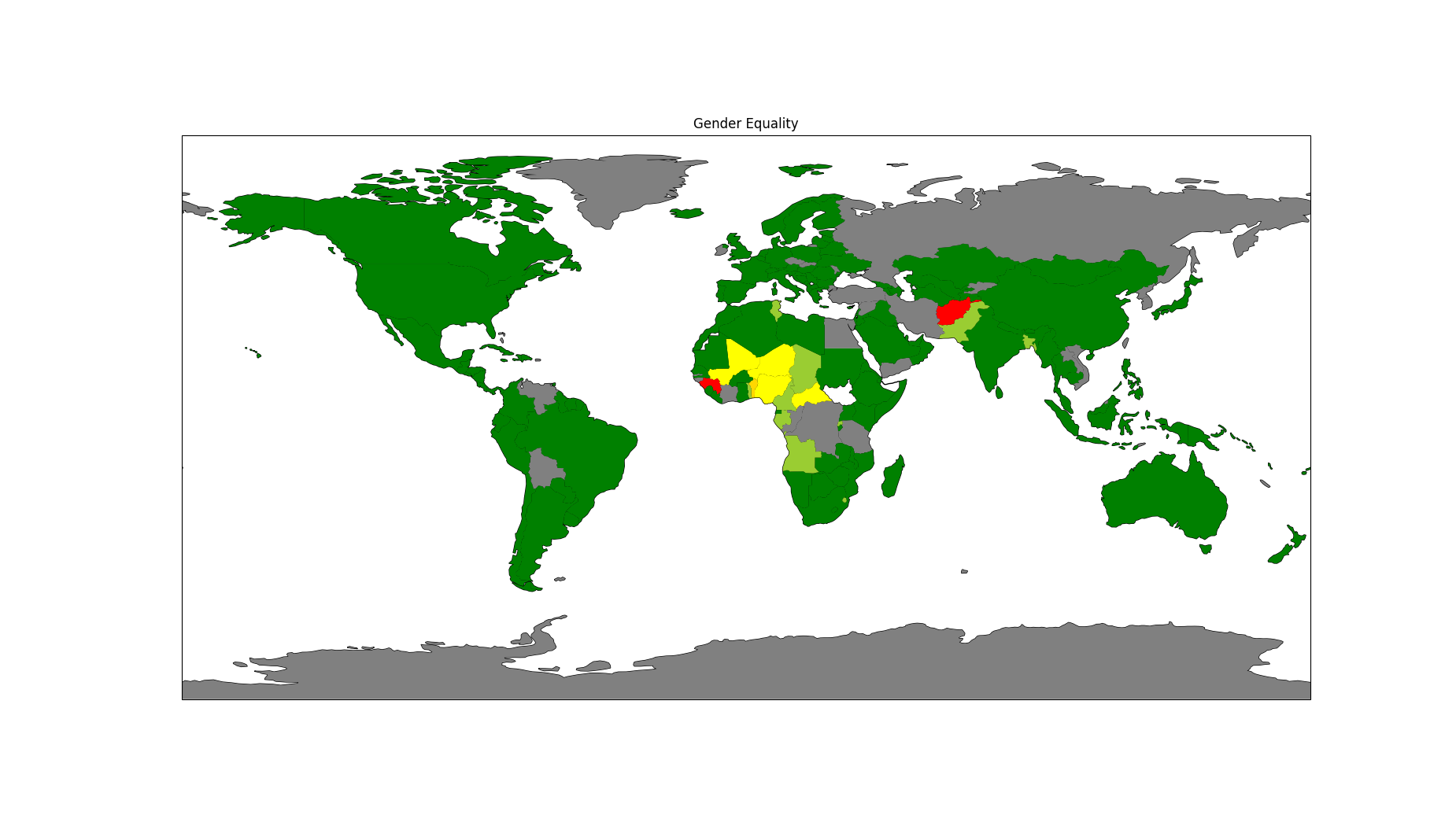
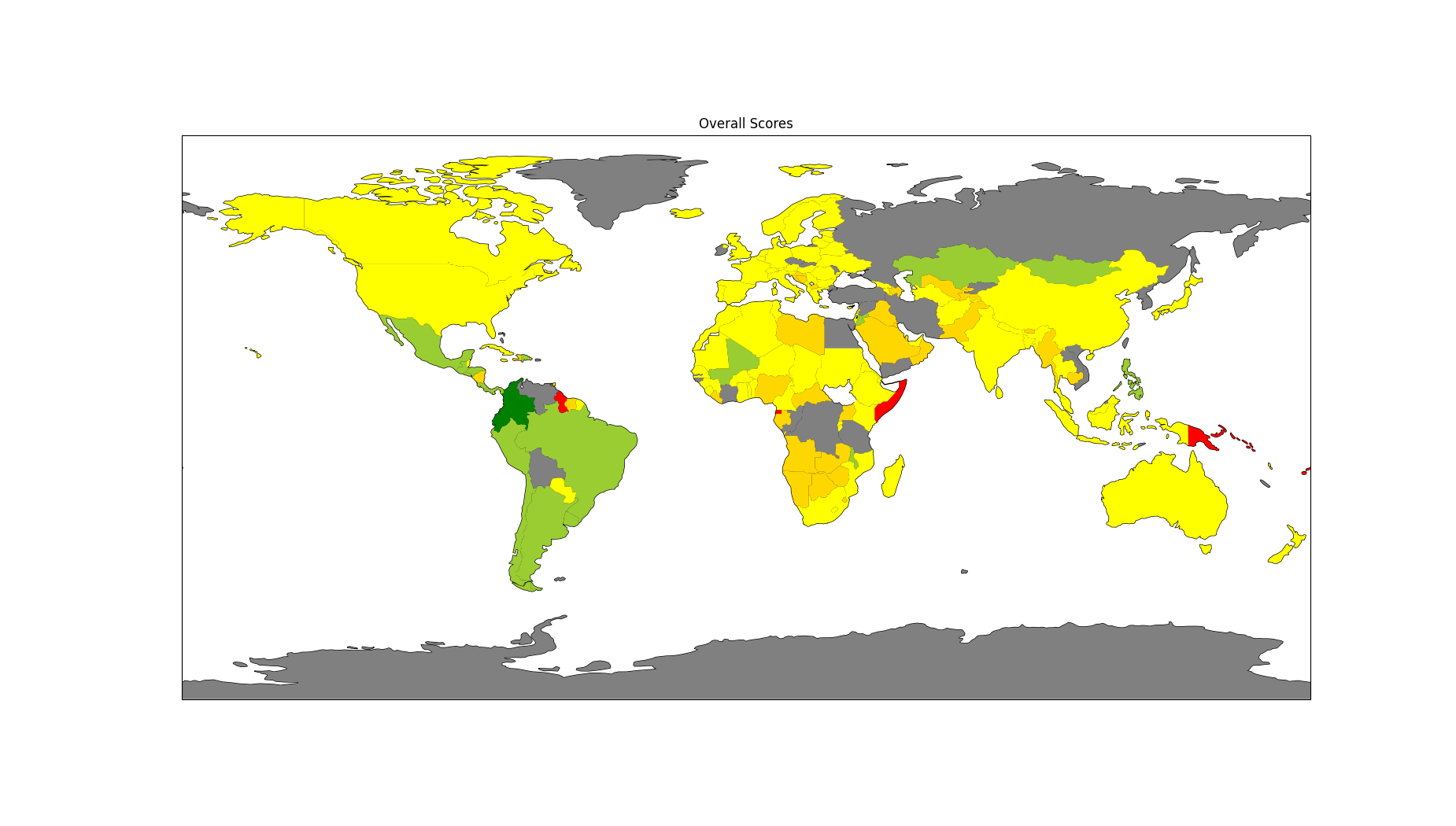
* Expenditure
  + Government expenditure on primary education as a % of GDP
  + Government expenditure on secondary education as a % of GDP
  + Government expenditure on tertiary education as a % of GDP
  + Number of years of compulsory primary and secondary education guaranteed in legal frameworks
  + Number of years of free pre-primary education guaranteed in legal frameworks
  + Number of years of free primary and secondary education guaranteed in legal frameworks
* Enrolment and Completion
  + Expected Years of School
  + Gross\_Primary\_Education\_Enrollment
  + Gross\_Tertiary\_Education\_Enrollment"
* Teacher Expertise
  + Pupil-qualified teacher ratio in pre-primary education
  + Pupil-qualified teacher ratio in primary education
  + Pupil-qualified teacher ratio in secondary education
  + Percentage of qualified teachers in pre-primary education, both sexes (%)
  + Percentage of qualified teachers in primary education, both sexes (%)
  + Percentage of qualified teachers in secondary education, both sexes (%)
* Student Proficiency
  + Grade\_2\_3\_Proficiency'
  + Primary\_End\_Proficiency
  + Lower\_Secondary\_End\_Proficiency
  + The three out-of-school-rate components
* Gender Equality
  + Expected Years in Education Delta
  + Literacy Rate Delta

**Links to other Indices**

Compared to other indices, my results are quite anomalous. My results place the best overall education systems in South America, with Columbia and Ecuador scoring very highly. The education component of the Human Development Index, by contrast, places these countries in the middle on a global scale, instead favouring North America, Europe and Australia. General groupings of countries with similar scores do not vary as widely, however.

**Visualisation of Results**



**References**

Campbell, J.A., (2021). The Moderating Effect of Gender Equality and Other Factors on PISA and Education Policy. Education Sciences [online], 11(1). Available from: <https://www.mdpi.com/2227-7102/11/1/10> [accessed 8th May 2025].

Gibbons, S., Scrutinio, V., Telhaj, S. (2021). *Teacher turnover: Effects, mechanisms and organisational responses* [online]. Available from: <https://www.sciencedirect.com/science/article/pii/S0927537121001147> [accessed 8th May 2025].

Ingeroll, R.M. (2001). *Teacher Turnover and Teacher Shortages: An Organizational Analysis* [online]. Available from: <https://www.gse.upenn.edu/pdf/rmi/TeacherTurnoverTeacherShortages-RMI-Fall-2001.pdf> [accessed 8th May 2025].

Jackson, C.K., Johnson, R.C., Persico, C. (2015). *The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms* [online]. Available from: <https://www.nber.org/system/files/working_papers/w20847/w20847.pdf> [accessed 8th May 2025].