Building a Linear Model to Explain University Ranking

Dylan Goldsborough and Geert Konijnendijk
Experimental Design and Data Analysis
Instructor: Eduard Belitser
University of Amsterdam
June 1, 2016

BUILDING A LINEAR MODEL TO EXPLAIN UNIVERSITY RANKING

Introduction

With the increasing globalization of the world, it is becoming significantly easier to study at any given university, regardless of one's country of origin. With the creation of this opportunity to study anywhere, the practice of ranking universities has taken off. Nowadays, many different institutions monitor the quality of universities worldwide, and publish annual rankings. Examples of these institutions are the Times Higher Education magazine, the Shanghai Ranking Consultancy and the Center for World University Rankings. These scores can have great effects on which universities are preferrably attended by students, and can even direct the flows of international students. In a study by Van Bouwel and Veugelers (2009), it was found that the Shanghai ranking has a great explanatory power in predicting where students will choose to attend university, as students tend to move from countries with low scores to countries with higher scores. With the availability of data on the quality of universities, the question can be asked what causes the quality of some universities to be higher than the quality of others. Is the spread around the world random, or are there traits exhibited by countries and regions that allow great universities to emerge? Are there economic factors that make a difference? And are there general traits of the university that affect the quality of education?

University ranking in general is not an easy assessment, demonstrated by the large number of different university rankings that disagree on the exact ordering of all universities. It is still disputed what the right way is to rank universities, as many rankings are heavily criticized on factors such as the indicators of quality they chose or their statistical inaccuracy (Dill & Soo, 2005). However, according to this study, the factors chosen to represent quality accross rankings are becoming increasingly in agreement with each other.

In general, there are a number of factors that are looked at to assess the quality of a school. One broad category of indicators are research achievements related to the institution. In this category, the quality of the school is judged by the number of academic awards won by alumni and staff, the number of papers annually published in renowned scientific journals, and how often researchers at this university are cited (Shanghai Ranking Consultancy, 2015). The second broad category regards teaching quality, and regards the staff-to-student ratio, the available funds for teaching, and the number of international students (Times Higher Education, 2016). Finally, some rankings attempt to supplement the resulting score with a reputation survey, performed on an annual basis (Times Higher Education, 2016). There is debate on whether variables related to the service to students should be included. An example is a score suggested by Murias, de Miguel, and Rodríguez

(2008), which includes availability of study places and ability to enroll in the students' chosen courses without oversubscription issues. Another survey also found that students themselves found the quality of service to them the most important factor in quality, instead of research achievements of the school (González-López, 2006). These factors are not without scrutiny: according to McAleer (2005), factors such as school size are often not taken into account in the right way, as school-wide statistics are used instead of per-student counts.

To determine a score from the different factors, two different methods are used (Dill & Soo, 2005). Some rankings, such as the Shanghai ranking, use and aggregate the raw scores, with weights for each factor. Others use a weighted aggregate of the rankings for each factor, and use this to determine their final ranking.

Only a limited amount of research has been done to answer the question what allows high quality education to emerge. Typically, studies focus on a smaller scale, often an area within the United States, or on high school education. A study by Marlow (2000) gives us some insight in the type of model that can be built to explain education quality. He built a linear multiple regression model to explain quality of public education in California. Factors he includes are the GDP per capita in the area, the population density, the student share of the population and the education spending per student. The variable population density is mostly interesting on a local scale, as it carries information about the neighborhood a school is located in. As universities draw students from all over the country this variable is not as meaningful, so for explaining university ranking this is not interesting. The education spending, student share of the population and GDP per capita, however, might be of interest.

There are several other factors that have not been utilized in this study, but could be interesting in the context of university quality. The interest rate on loans could be interesting, as tuition fees are expensive in many countries, and student loans obligatory as a result. It might also be interesting to see if an interaction is present between income per capita and the effect of loan interest rates, as higher income could lessen the need for loans, and thus lessen any effect that interest rate has. The percentage of the population with internet access could be an interesting operationalization of how modern the country is, which could also affect the quality of the universities in this country. According to McAleer (2005), one important factor in the quality of a university education is the size of the university. Finally, we can test if there are differences in the university quality based on the continent on which the university is located.

In this study we aim to build a model to explain the quality of university education. To produce models, we will use multiple regression, ANOVA and an ANCOVA. We use data from the Shanghai Ranking Consultancy as dependent variable. The country-specific explanatory variables used to construct the models are GDP per capita, education spending per student, the loan interest rate, the percentage of the population with internet access, and the continent on which the university is located. Furthermore, we will also include university-specific data, namely the ratio of men to women, the ratio of international students, the size of the university, and the staff-to-student ratio.

In Section , we describe how we build our dataset using data from different sources. Then, we describe the model building process, and the hypotheses associated with this model. The results are shown in Section , after which conclusions are drawn in Section . Finally, discussion points are then brought forward regarding the operationalization of the variables, and suggestions for future research in Section .

Methods

Dataset

In this study we use data from multiple sources. The aggregated score from the Academic Ranking of World Universities (Shanghai Ranking Consultancy, 2015) acts as a response variable. The Shanghai Consultancy dataset focuses heavily on achievements from students and teachers alike. The Shanghai score is based on performance of alumni, performance of staff, research output and per capita performance which in turn is based on a weighted average over the number of full-time academic staff of the previous criteria. A weighted average of these four criteria then forms the total score. The Shanghai dataset contains data from a number of years, the 2011 data was used for analysis. The 2011 data contains rankings for 500 universities.

The explanatory variables are obtained from the Times Higher Education World University Rankings (Times Higher Education, 2016) and the World Databank (The World Bank, 2016). The variables from the Times ranking are used since they contain detailed per-university information, namely the number of students, the male-to-female ratio, the student-to-staff ratio, and the number of interational students. The World Databank data gives different development indicators for each country world-wide, such as the spending on education per student, the interest rate, the GDP per capita, and the fraction of the population with internet access. Out of data from multiple years in the Times dataset, again the 2011 data was used. This data contains 200 universities. The data obtained from the World Databank is all from the year 2011. Each variable from the World Databank data contains some missing values for some of the 248 countries present in the dataset.

The obtained datasets were merged by matching them on country names and university names and adding a variable representing the continent a university is in, to be used in

ANOVA and ANCOVA. Over all datasets, data from 525 universities is available. After combining the datasets, there are observations with missing values. During the analysis as many observations as possible were kept, meaning that different numbers of observations were present depending on the variables present. The dataset containing all variables consists of 113 observations. That containing only the Shanghai score, student-to-staff ratio and continent variables has 167 observations. When we add the international student ratio,university size, government expenditure per student and GDP this becomes 121. Other combinations of variables were not used.

Variable Selection

We selected a pool of explanatory variables on the basis of the review of the literature. The potential independent variables selected were: the continent, the number of students, the male-to-female ratio, the student-to-staff ration, the interational students ratio, the government expenditure on education per student, the interest rate, the GDP per capita, and the fraction of the population with internet access. We also included the interaction between GDP and the interest rate, to account for the possibility of student loans playing a role in countries with a lower per capita income. As a response variable, the Shanghai score on a university basis was selected.

The variables representing internet access and government expenses per student appear to be collinear, as can be seen in Figure 4 in appendix A. As such, only one of the two ought to be included in our model. In our preliminary investigation, we found strong collinearity between the fraction of the population that has internet, and the education expenditure per student. This correlation is possibly present because both are likely to be higher in richer, more modern countries. We choose to drop the internet access, as the non-common variance in the education expenditure per student is likely more relevant. This is supported by a linear regression of both variables on the log of the Shanghai score: the internet access was not significant ($p = .425$), whereas the expenditure per student was highly significant ($p = .003$).

The response variable appears to follow an exponential distribution (Figure 3 in appendix A). As a result, we take the natural logarithm of the Shanghai score as our response variable in the models.

We performed both a step-up and step-down algorithm to produce a multiple linear model. The step-up and step-down algorithm produced differing models. We selected the linear model produced by the step-down algorithm, as it had a higher explanatory power ($R^2 = .216$) than the step-up model ($R^2 = .2098$), as well as two fewer explanatory variables. The produced model contains the following explanatory variables: number of students ('n'), staff-to-student ratio ('ssr'), national education expenditure per student ('eps'), GDP per capita ('gpd'), the national interest rate ('r'), and the international student ratio ('isr'). We calculated the Cook's distance for all datapoints, and detected no outliers. The interest rate was not significant in this model by a small margin, but was included to control for this variable in the model.

To evaluate the effect of the continent on the score, we first perform an 1-way ANOVA using the logarithm of the Shanghai score as the response variable, and the continent as the explanatory variable. In addition, we perform an ANCOVA that includes the continent as a factor, and the variables of the linear model where the assumption holds that all interactions between this variable and the continent are parallel. We do not consider country-specific variables for the ANCOVA, as some continents contain too few countries.

Hypotheses

For significance, a value of $\alpha < .05$ was assumed. For the multiple linear model, we formulate the following hypotheses:

1. $H_0 : \beta_{\text{ssr}} = 0$
   $H_1 : \beta_{\text{ssr}} \neq 0$
2. $H_0 : \beta_{\text{isr}} = 0$
   $H_1 : \beta_{\text{isr}} \neq 0$
3. $H_0 : \beta_{\text{eps}} = 0$
   $H_1 : \beta_{\text{eps}} \neq 0$
4. $H_0 : \beta_{\text{n}} = 0$
   $H_1 : \beta_{\text{n}} \neq 0$
5. $H_0 : \beta_{\text{gdp}} = 0$
   $H_1 : \beta_{\text{gdp}} \neq 0$
6. $H_0 : \beta_{\text{r}} = 0$
   $H_1 : \beta_{\text{r}} \neq 0$

For the ANOVA, we formulate the hypothesis $\mu_{\text{EU}} = \mu_{\text{AS}} = \mu_{\text{NA}} = \mu_{\text{OCE}}$. The ANCOVA tests the same hypothesis, but holding the student-to-staff ratio constant.

Statistic	Score	Students	Std:Stf	Internat.	F:M	GDP	Expend	Interest
Min.	9.32	2243	3.60	0.0101	0.1494	35142	19.99	0.500
1st Qu.	18.84	14604	10.00	0.1364	0.9231	49781	20.08	2.000
Median	23.74	21394	14.10	0.2346	1.0833	49781	20.08	3.250
Mean	27.66	23526	14.08	0.2872	1.0173	50469	24.95	2.906
3rd Qu.	31.06	29991	17.00	0.3889	1.1739	49781	32.01	3.250
Max.	70.44	62468	40.50	1.1739	1.5641	88003	39.86	7.737

Table 1
Dataset statistics.

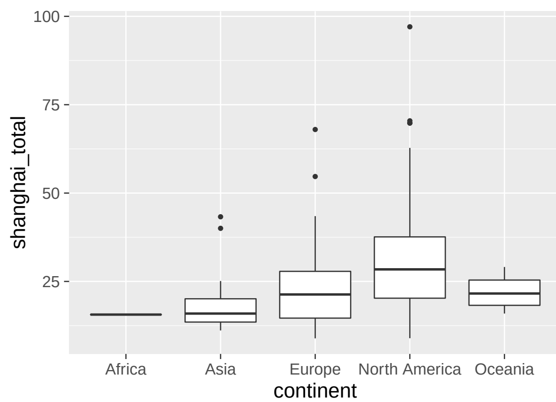
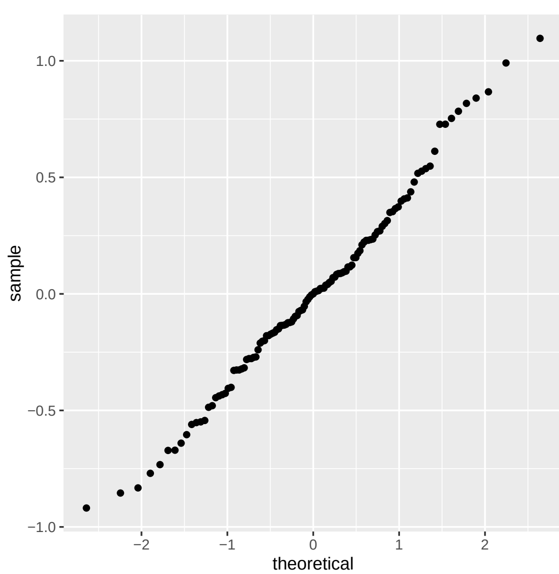
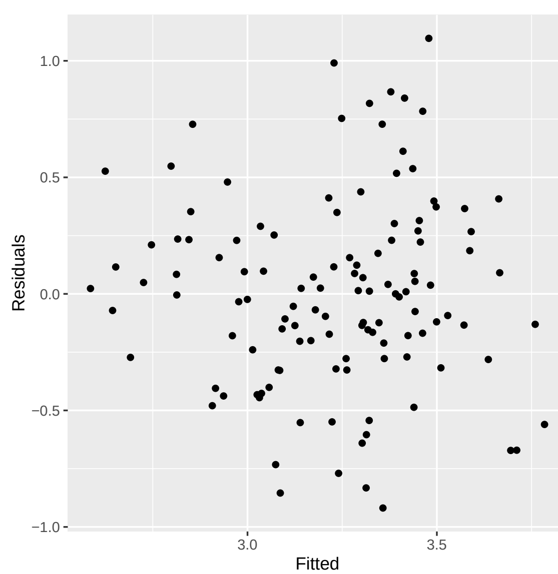


Figure 1
Boxplots of Shanghai score per continent



(a) *QQ-plot of full model residuals.*



(b) *Fitted against residuals of full model.*

Figure 2
Plots checking model assupmtions.

Results

Descriptive Statistics

The dataset used in ANOVA and ANCOVA contains data from universities 4 continents: 20 from Asia, 65 from Europe, 74 from North America and 8 from Oceania remain in the dataset after removing entries with missing values. Statistics about the dataset containing all variables can be found in Table 1. In this table we see that some variables are likely not normal because their mean and median do not overlap and the two quartiles are not equally far away from the median which could be a problem for ANOVA, ANCOVA and linear regression. Examples are the Shanghai score and the government expenditure per student.

As a first investigation into the dataset we plotted the histograms for all variables in Figure 3 in Appendix A. The distribution for most variables is not fully clear, but it can be assumed normal for the variables originating from the Times dataset. The World Databank variables are too sparse to assume a distribution for. The score response variable is clearly not normal, but most likely exponentially distributed.

Figure 1 shows box plots of the Shanghai score per continent using the dataset containing only Shanghai score and continent variables. In this figure North America appears to have the highest scores, followed by Europe and Oceania, but they are relatively close together. ANOVA can be used to detect if the scores are significantly different.

Assumptions

To verify that the built final linear model is valid, its assumptions need to be checked. The residuals of the model are distributed normally, as illustrated in Figure 2a. This was confirmed by a Shapiro test ($p = .5389$). To test for heteroscedasticity, the fitted values were plotted against the residuals. Figure 2b shows no particular shape, so we can assume the data is homoscedastic.

In addition to the normality and heteroscedacity assumptions, there should be no interaction between the different continents in order to use ANCOVA. Figure 5 shows a fit for each variable and each continent. Only for the student staff ratio variable the lines are reasonably parallel and can no interaction be assumed.

In ANOVA, the group size is sufficiently large in all cases except Oceania, which has size 8. This is a relatively small and should be taken into consideration.

Coefficient	Estimate	Std. Error	t-value	p-value
(Intercept)	3.558e+00	2.588e-01	13.746	$<$ 2e-16 ***
student:staff ratio	-2.760e-02	8.154e-03	-3.385	0.000976 ***
international student ratio	6.687e-01	2.376e-01	2.815	0.005749 **
expend per student	-3.173e-02	1.111e-02	-2.855	0.005109 **
number of students	1.319e-05	3.547e-06	3.720	0.000311 ***
GPD	1.112e-05	5.115e-06	2.175	0.031713 *
interest rate	-7.342e-02	4.231e-02	-1.735	0.085392

Table 2
*Signif. codes: '***' .001, '**' .01, '*' .05.*

Contrast	Diff.	Lower 95%	Upper 95%	p-value
Europe-Asia	0.17403870	-0.12245257	0.4705300	0.4257300
North America-Asia	0.46474087	0.17252272	0.7569590	0.0003369
Oceania-Asia	0.19765325	-0.28740417	0.6827107	0.7156506
North America-Europe	0.29070218	0.09359203	0.4878123	0.0010460
Oceania-Europe	0.02361455	-0.41082953	0.4580586	0.9989942
Oceania-North America	-0.26708762	-0.69862677	0.1644515	0.3778006

Table 3
Results of a Tukey post hoc analysis.

Hypothesis Testing

The results of the multiple linear regression model are summarized in Table 2. We found that the selected variables explained a signficant part of the variance in university score in the Shanghai ranking, $R^2 = .2563$, $F(6,114) = 7.891$, $p = 4.035\text{e-}07$. We find that increasing the number of staff members per student has a very significant negative correlation with the received score. The national education expenditure per student was also found to have a negative correlation with the university score. We find that an increasing number of students corresponds to an increase in the university score as well as the ratio of international students to regular students, and the national GDP per capita. The interest rate was not found to be significantly related to the Shanghai university score.

In the one-way ANOVA, we observed a significant main effect for continent on the logarithm of the Shanghai score, $F(3,163) = 8.1454$, $p = 4.365\text{e-}05$. In Table 3, the results of a Tukey post hoc analysis are summarized. We find that the mean logarithm of the Shanghai score are significantly different between North-America and Europe, and between

North-American and Asia. In both cases, North-America has the higher score.

In the ANCOVA, we found that there was a significant effect of the continent on the logarithm of the Shanghai score, after controlling for the student-to-staff member ratio, $F(3,162) = 7.2541$, $p = .0001349$. The covariate, the student-to-staff member ratio, was significantly related to the logarithm of the Shanghai score, $F(1,162) = 6.9792$, $p = .0090545$.

Conclusion

In this study we investigated what factors influence university quality. More specifically, per-univerisity and per-country variables that were suspected to influence ranking were collected by their relevance in literature and their effect on the Shanghai university ranking was investigated. Three different models were built, one using ANOVA, one using ANCOVA and one using multiple linear regression. Model assumptions were validated before testing the hypotheses. Student-to-staff ratio, international student ratio, government expenditure per student, university size and GDP per capita were found to be significant while the interest rate on loans was not. The continent a university is on was found to make a difference in Shanghai score, even when controlling for student-to-staff ratio.

In our model we found that having more staff members per student has a positive correlation with the Shanghai score. When there is more staff per student there is less pressure on staff members and students have better access to staff, intuitively increasing quality.

While seemingly counterintuitive, we can explain the negative effect of increased education spending per student on the quality of the best universities. Countries that spend a lot per student often have strong public universities, with regulations in place that limit the maximum tuition fee. This tuition fee, combined with state funding per student, is the only source of income for the university. Countries that do not invest as heavily into education do not have these caps on the tuition fee, and as a result universities in these countries can end up with a higher income per student than in countries with more funding. With more money to spend per student the quality of education and research can be improved, for example by allowing the university to spend more on modern facilities that facilitate research and learning.

The university size and international student ratio are also positively correlated with the Shanghai score. This is in accordance with the literature, which suggests that university size is an important factor, and also predicts a link between flow of international students and Shanghai ranking. This is probably an effect of universities that are perceived to have a high quality attracting more students, both from within a country and from abroad. University size could have an effect, as it is possible that larger universities are able to maintain more expensive resources. For example, an H-NMR spectrometer costs millions, and many universities do not own one. Owning one could improve the quality of education and research at this university.

GDP per capita was found to correlate positively with the Shanghai score. The GDP having a significant effect on Shanghai score can also be explained in a straightforward manner since universities in a country with a higher GDP are likely to be better equipped, both on staff and other aspects.

In both ANCOVA and ANOVA we found that some continents have better scores than others, ANOVA suggests that North America has the highest scores. North America has significantly higher scores than Europe and Asia, but not significantly higher than Oceania. This might be the case because Oceania's sample size is low. In ANCOVA we again find that the continent is significant after controlling for student to staff member ratio.

Discussion

A main discussion point in this study is the operationalization of the dependent variable. Quality of education is an abstract concept that is not measurable in one definite, objective way. We chose the Shanghai score, as opposed to another score, for its relative objectivity and it being used in the literature. However, the weights given to each of their criteria, as well as their choice of criteria, is subjective.

Another criticism is that we only look at the best universities, and not at an overall distribution of the quality of universities throughout the nation. This is a likely cause for the magnitude and direction of the effect of education spending, as a lower education spending per student likely benefits the few top universities, while disadvantaging most other institutions. For a better image we would consider all universities, but a dataset for such a study is much harder to come by.

A final problem is that not all countries are as strongly represented as others in the top 200 universities. Countries such as the United States of America and the Netherlands have a relatively strong presence in the top 200, whereas the entire continent of Africa is represented by one single entry, which had to be dropped in order to perform the ANOVA. Future studies would benefit from using a larger sample, in which various countries are represented in similar numbers.

Furthermore, a variable that might have been interesting but of which data was not available is the student ratio of a country's population. This variable is expected to be of interest because a larger student population could be an impulse to improve univeristy quality.

References

Dill, D. D., & Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher education*, *49*(4), 495–533.

González-López, I. (2006). Dimensions for evaluating university quality in the european space for higher education. *Electronic journal of research in educational psychology*, *4*(3), 445–468.

Marlow, M. L. (2000). Spending, school structure, and public education quality. evidence from california. *Economics of Education Review*, *19*(1), 89–106.

McAleer, M. (2005). The ten commandments for ranking university quality. *Journal of Economic Surveys*, *19*(4), 649–653.

Murias, P., de Miguel, J. C., & Rodríguez, D. (2008). A composite indicator for university quality assesment: The case of spanish higher education system. *Social Indicators Research*, *89*(1), 129–146.

Shanghai Ranking Consultancy. (2015). *Academic ranking of world universities 2011-2015.* Retrieved from `https://www.kaggle.com/mylesoneill/world-university-rankings`

The World Bank. (2016). *World development indicators.* Retrieved from `http://data.worldbank.org/indicator`

Times Higher Education. (2016). *World university rankings 2011-2016.* Retrieved from `https://www.kaggle.com/mylesoneill/world-university-rankings`

Van Bouwel, L., & Veugelers, R. (2009). Does university quality drive international student flows? *Available at SSRN 1538118*.

A Plots

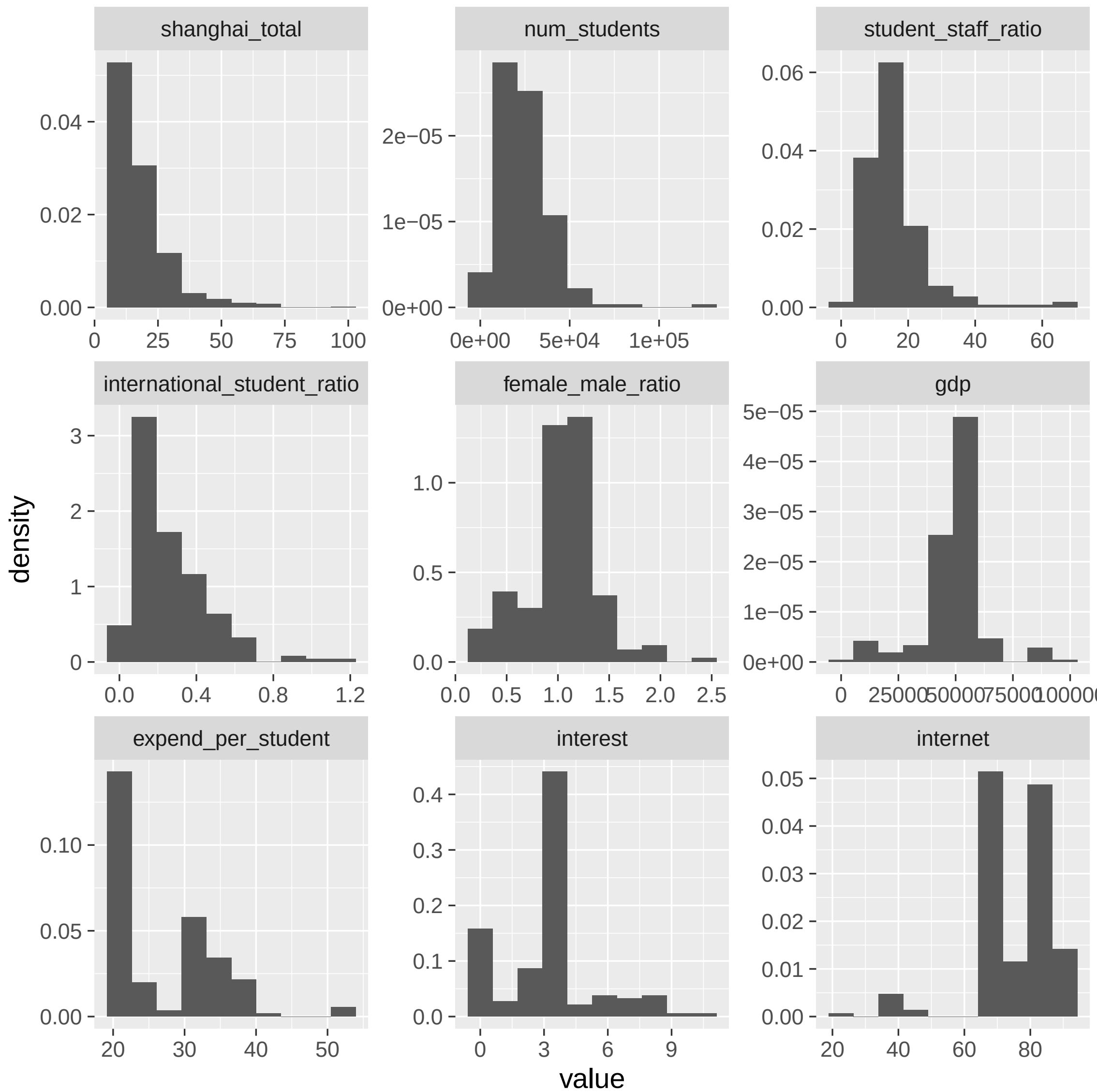


Figure 3
Histograms of the dataset.

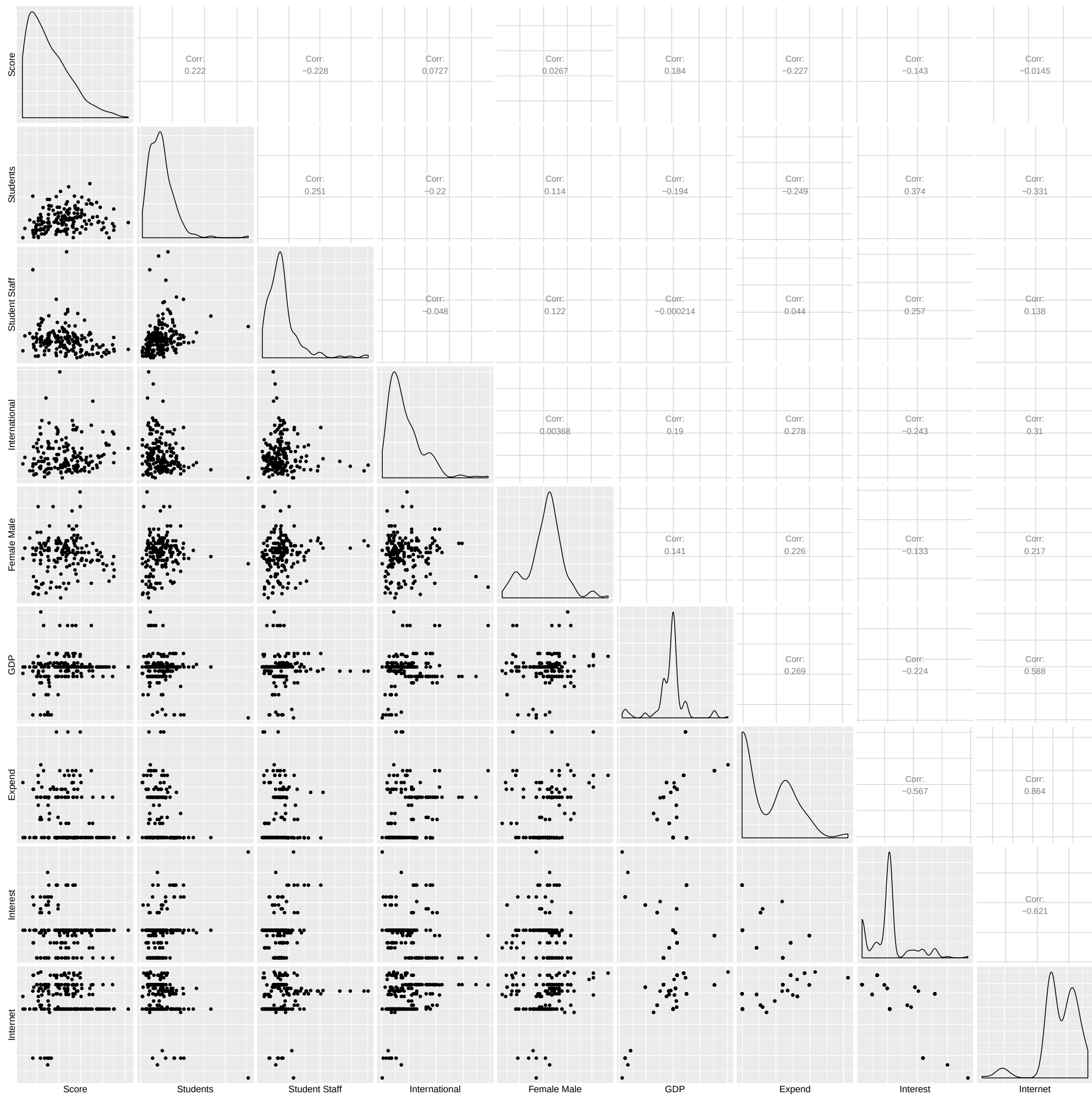


Figure 4

Pairwise plots of the dataset with correlation.

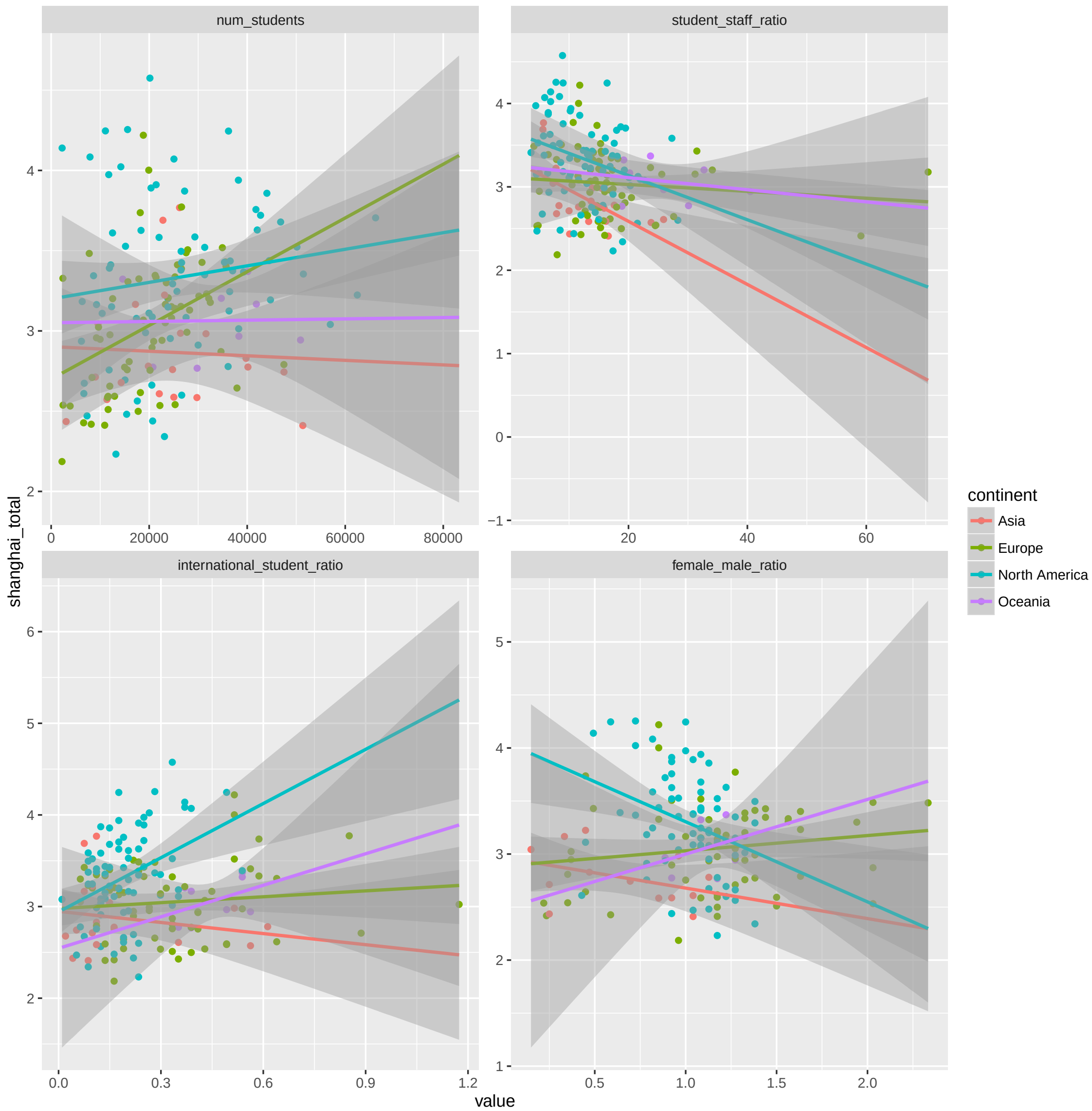


Figure 5

Interaction between continents.