

Factors related to Geographical Literacy

Shruti Hegde
University of Maryland
College Park, MD
shegde17@terpmail.umd.edu

Himanshi Manglunia
University of Maryland
College Park, MD
himanshi@terpmail.umd.edu

ABSTRACT

Previous research work has investigated that there is interest to determine whether specific countries are associated with lower geographical literacy.

In line with those studies we are trying to explore whether the country is actually related to geographical literacy rate or there are other factors which may be related. In order to examine the factors that might be associated to a person's geography literacy, we performed multiple linear regression.

Author Keywords

Geographical literacy; Multiple regression; predictor and outcome variables;

INTRODUCTION

In order to determine the Geographical literacy of people in a specific country, there have been studies which tested the knowledge of a sample of participants. The results of these tests seem to be shocking. 74% of survey participants from Britons can't find Greece on a map, and 10% confused Poland with America [1]. 65% of people surveyed in UK mistakenly believe that Britain is made up of four countries, rather than the correct three [3]. After more than three years of combat, nearly two-thirds of Americans couldn't locate Iraq on a map [4]. Why do we care about these statistics? We do because Geographic literacy is important. Why is it important? It is important in today's globally interconnected society, that people understand the world around them, for better decision making [5]. It is necessary to know where the people we interact with everyday, come from, and how we interconnect with each other [6].

Through this study, we have tried to determine if there are factors which are related to a person's Geographical literacy by conducting a survey. The results of the survey were passed through a Multiple Regression Model and the most significant factor which is associated with a person's Geographical literacy was determined.

DATA COLLECTION

For the survey, we approached participants to answer pre decided Geographical literacy questions. The sample for this study are the students of University of Maryland who participated in the survey. The data was collected on Monday, Friday and Sunday from 2.30 pm

to 5.30 pm. The passers by near the Stamp students union and the McKeldin Library. The survey took about 3-4 minutes per participant. There were a total of 97 participants surveyed.

THE SURVEY QUESTIONS

The questions asked in the survey are as follows:

Demographics :

1. Age
2. Education Level
3. Country of origin
4. Taken Geography in school

Common Geography:

1. Latitude/Longitude
2. Largest country
3. Largest ocean
4. Highest mountain
5. Full form of DC in Washington DC

Locate following countries and state on a world map:

1. India, China, Japan, Argentina, USA, Brazil, Russia, UK, Germany, South Africa, Ethiopia, Australia, Thailand, Egypt.
2. Maryland

Challenges faced

This study uses the Multiple Linear Regression as the aim is to understand the values of the Quiz score on Geographic awareness with respect to multiple predictor variables. The response variable is the 'Total score variable'. The predictor variables that we considered are age, education level, countries and have you studied Geography in school. As the data points collected for our study were less and sparse, the levels for the categorical variable Country of origin has been reduced into *continents* using countrycode library in R. The final predictor variables are age, Education level, continents and have you studied Geography in school.

DESCRIPTIVE STATISTICS

The graph in Figure 1 depicts the distribution statistics of the continent, which is a categorical type of variable. These values were computed using R.

There were 21 participants from Africa, 29 from Americas (North and South together) and a majority, 37 from Asia. Figure 2 shows the distribution of Education Level for the sample. Minority of the participants, 38 were Graduate students, and majority, 59 were Undergrad students.

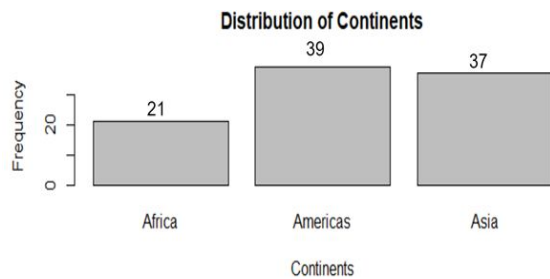


Figure 1. Distribution of the Continents

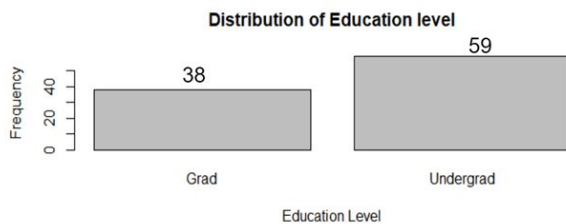


Figure 2. Distribution of previous Geographic Education

Figure 3 shows that 41 people had previous Geographical education in school and 56 of the surveyed had no Geographical education before.

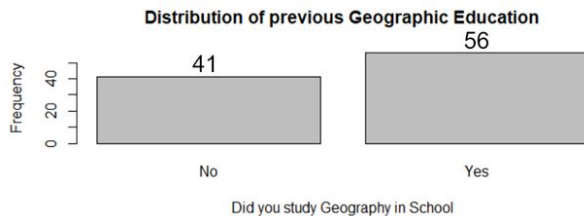


Figure 3. Distribution of previous Geographic Education

Using the box plot in Figure 4, we can see that the mean age of the participants is 22.05, with a standard deviation of 3.22. The mean score is 14.36.

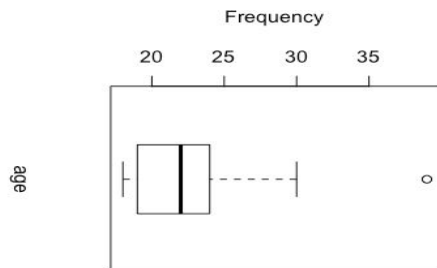


Figure 4. Distribution of age

The box plot in Figure 5 shows the distribution of Scores obtained by the participants. The mean score is 14.36, with a standard deviation of 4.83. It can be

observed that the score is left skewed, and most of the participants tend to have better than average score.

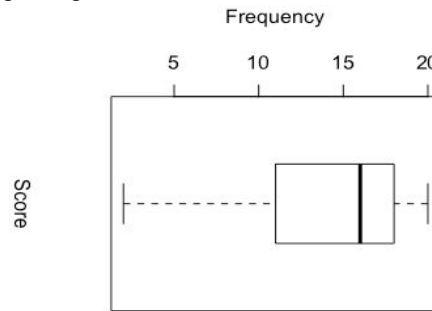


Figure 5. Distribution of the Score obtained

MULTIPLE REGRESSION

In this study we analyzed the data in order to see what goes into a person's geographical knowledge.

Full model = age + continent + Education level + did you take geography

Checking Assumptions

Before we build the model, we check if the assumptions of multiple linear regression hold true. Multiple regression methods depend on the following four assumptions:

1. the residuals of the model are nearly normal,
2. the variability of the residuals is nearly constant,
3. the residuals are independent, and
4. each variable is linearly related to the outcome.

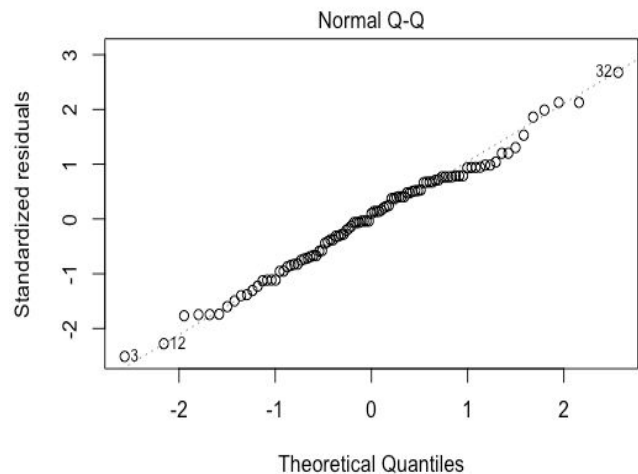


Figure 6. A normal probability plot of the residuals

A normal probability plot of the residuals is helpful in identifying observations that might be outliers. A normal probability plot of the residuals is shown in Figure 6. While the plot exhibits some minor irregularities, there are no outliers that might be cause for concern. It satisfies the first condition i.e the residuals of the model are nearly normal.

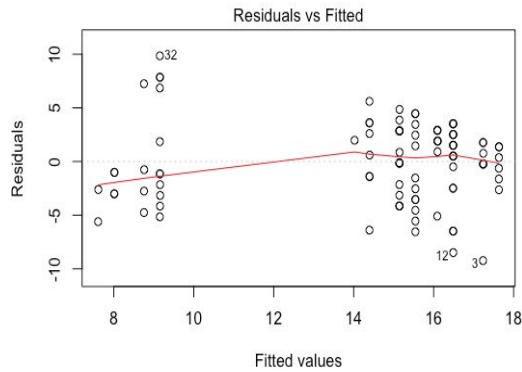


Figure 7. Plot of the residuals against their corresponding fitted values

A plot of the residuals against their corresponding fitted values is shown in Figure 7. The variance of the residuals is approximately constant. There are no obvious deviations from constant variance. The red line is approximately horizontal at 0 which satisfies the constant variability condition.

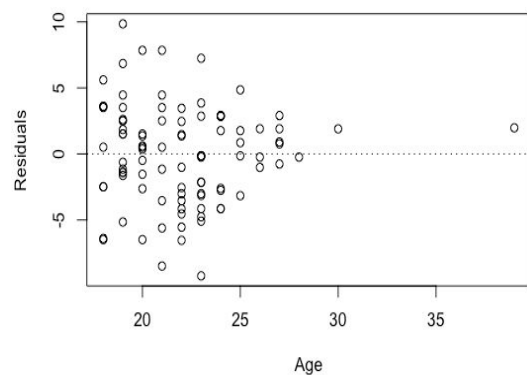


Figure 8. Plot of residual vs predictor variable 'Age'

The residuals don't appear to be evenly distributed and has an outlier in Figure 8. The predictor variable 'Age' doesn't appear to be linearly related to the response variable.

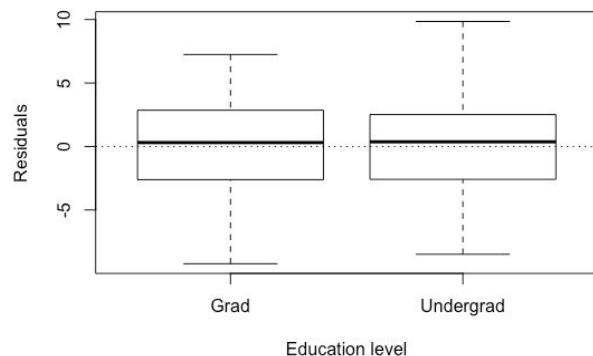


Figure 9. Plot of residual vs predictor variable 'Education level'

For the two-level Education variable, we are guaranteed not to see any remaining trend, and instead

we check that the variability doesn't fluctuate across groups, which it does not here.

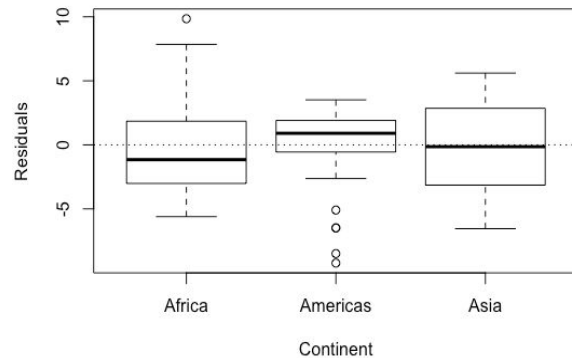


Figure 10. Plot of residual vs predictor variable 'Continent'

In the case of the Continent variable, we see a little less variability in the Americas group than the Africa and Asia groups. There appears to be non-constant variance across Continent variable.

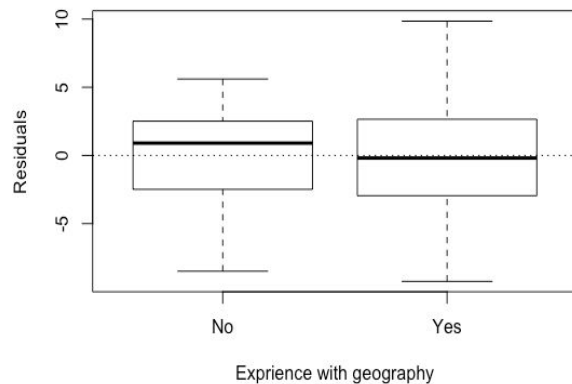


Figure 11. Plot of residual vs 'Did you take geography in school'

In the case of the 'Did you take geography in school' variable, we see less variability in the 'No' group than the 'Yes' group. There appears to be non-constant variance across this variable.

Full model

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.578038	4.491140	1.687	0.095
Age	0.001449	0.168920	0.009	0.993
ContinentAmericas	8.479390	1.096311	7.734	1.35e-11
ContinentAsia	6.387709	1.051270	6.076	2.82e-08
Education.levelUndergrad	0.402238	1.139496	0.353	0.725
Did.you.take.Geography.in.school.Yes	1.143869	0.878959	1.301	0.196

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.8 on 91 degrees of freedom
Multiple R-squared: 0.4146, Adjusted R-squared: 0.3824
F-statistic: 12.89 on 5 and 91 DF, p-value: 1.767e-09

Figure 12. Summary statistics of the full model

Using the `lm()` , we built our full model and applied `summary()` function on our full model to get the results as in figure 12.

Hypothesis Testing

I. Significance for overall model

Question of interest: Is the regression relation significant? Are one or more of the independent variables in the model useful in explaining variability in total score?

Null hypothesis : There is no relation between the independent (predictor) variables and total score.

Alternate Hypothesis: At least one of the independent variables is useful in explaining/predicting the total scores.

Ho: $\beta_1 = \beta_2 = \beta_3 = 0$

H_A : At least one $\beta_j \neq 0$ (for $j = 1, 2, 3$)

The significance level (α) for testing is 0.05. The F statistic is 12.89 and the p-value is $1.767e-09$. Since the p-value < 0.001 , we reject the null hypothesis. The data provides convincing evidence that this model is significantly better than the intercept model. At least one of the independent variables is useful in predicting the total score in the model.

II. Significance of a Single Independent Variable in the Model

H_0 : $\beta_i = 0$ when other explanatory variables are included in the model.

H_A : $\beta_i \neq 0$ when other explanatory variables are included in the model

where i is an independent variable in the model

Since the p value (0.993) is greater than 0.05, the variable 'age' doesn't contribute to the model and is not significant when other independent variables are included in the model. Similarly 'Education level' and 'Did you take geography' variables are not significant when other independent variables are included in the model as the p values are high. The p values for continentAmericas and ContinentAsia are $1.35e-11$ and $2.82e-08$ respectively. Since values are less than 0.05, the variable 'Continent' is significant when other independent variables are included in the model.

Model selection

The R function `step()` was used to perform predictor variable selection. We started with the full model and performed backward elimination on the dataset.

The best model is the one with the smallest AIC score. From the results in Table 1, the best model is the one that includes the predictor variable continent.

Model	AIC score
Score ~ Age + Continent + Education Level + Did you take geography in school	264.79
Score ~ Continent + Education Level + Did you take geography in school	262.79
Score ~ Continent + Did you take geography in school	261.04
Score ~ Continent	260.62

Table 1. Result for model selection using step()

RESULTS AND DISCUSSION

Call:

```
lm(formula = Score ~ Continent, data = geo)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-8.7179 -2.7143  0.2821  2.2821 10.2857
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.7143     0.8236   10.581 < 2e-16 ***
ContinentAmericas 8.0037     1.0215    7.835 7.07e-12 ***
ContinentAsia   6.3668     1.0311    6.174 1.68e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 3.774 on 94 degrees of freedom

Multiple R-squared: 0.4034, Adjusted R-squared: 0.3907

F-statistic: 31.78 on 2 and 94 DF, p-value: $2.86e-11$

Figure 13. Summary statistics of the best model

Figure 13 shows the summary statistics of the best model. Adjusted R^2 for the best regression model is 0.3907 which means that 39.07% variability in total scores is explained by the continent variable.

Interpretations

Intercept: Participants from Africa on average have a total score of 8.71 points

Slope: Participants from the continent Americas on average score 8 points more than participants from the continent Africa.

Participants from continent Asia on average score 6.36 points more than participants from continent Africa.

LIMITATIONS

1. **Convenience Sampling:** Participants were students at UMD and they were easily accessible. There were only two buildings in the campus where the data collection was performed. The sampling was not uniformly distributed across the University of Maryland

campus because of shortage of time. This bias can be eliminated by approaching a randomly selected set of participants.

2. **External factors:** As the study was done without taking prior appointments from participants, it's possible that the students were impatient to participate in the study due to other commitments, and they had to rush while answering their questions, thereby not giving their cent percent effort to answer the survey questions. This can be eliminated by scheduling appointments for interviews.
3. **International Travel:** The students who travelled internationally would have been better at answering geographical questions, as they have traveled across the world as compared to students who have been living in the same state/country throughout.
4. **Unconscious bias:** As the participants were selected by the researchers who were conducting the study themselves, who knew the hypothesis, maybe they tried to approach people from a certain country more unconsciously, which may have biased the results.
5. **Survey questions bias:** The survey questions were not distributed uniformly across all the continents. If a particular participant belonged to a continent with the higher number of questions, he would have gotten a higher score than the participant who belonged to the continent with lower number of questions.

REFERENCES

1. (2014, March 21). Where in the world? - Retrieved November 29, 2018 from <https://www.dailymail.co.uk/travel/article-2586028/Britons-bad-geography-revealed-74-Britons-Greece-map.html>
2. (2015, Oct 16). US students are really bad at Geography - Retrieved November 29, 2018 from <https://www.usnews.com/news/articles/2015/10/16/us-students-are-terrible-at-geography>
3. (2008, Nov 19). Geography: Why in the world do we know so little - Retrieved November 16, 2018 from <https://www.theguardian.com/education/2008/nov/19/bad-at-geography>
4. (2006, May 4). CNN.com - Study: Geography Greek to young Americans - May 2, 2006. Retrieved October 19, 2018, from <http://www.cnn.com/2006/EDUCATION/05/02/geog.test/>
5. (2012, Jan 19). Why is Geo-literacy important? - Retrieved November 28, 2018 from <https://www.nationalgeographic.org/media/why-is-geo-literacy-important/>
6. (2012, Sept 26). The importance of Geo-Literacy - Retrieved November 30, 2018 from https://blog.education.nationalgeographic.org/2012/09/26/the_importance_of_geo-literacy/
7. (2009, May 6). World: Americans Fail Geography, but are other nationalities better? - May 6, 2009. Retrieved November 16, 2018, from <https://www.rferl.org/a/1068259.html>
8. (2018, July 17). Jimmy Kimmel asks Americans to name countries on map with. Retrieved October 30, 2018, from <https://www.news.com.au/entertainment/tv/jimmy-kimmel-reveals-simple-test-so-many-americans-just-cant-pass/news-story/e4a43b076266459f9a0683bc2c23bc0f>
9. (n.d) Colorful World map Retrieved October 28, 2018 from <https://www.vectorstock.com/royalty-free-vector/simple-colorful-world-map-vector-3554502>