

# Learning analysis: Exploring the interactions and demographics between different demographics

Adanna V. Obibuaku

21/11/2021

## 1. Indroduction

### 1.1 Motivation

This report focuses on the interactions and performance of an online course called “Cyber Security: Safety at Home, online and in Life” between different demographics. The type of demographics this report focuses on includes:

- Age range
- Employment status
- Region
- Gender
- Highest education level.

The report looks closely the demographics of pupil learning the course, their interactions, engagement and performance. Additionally, it explores reasoning behind the performance of each group; trying to identify constraints a demographics may have for engagement or learning the online course.

The report includes two models:

1. The first model takes uses two data sets; demographics data (age, gender etc). and usage of data which gives an insight about student engagements (number of weekly activities/quizzes etc), to predict the performance of a user.
2. The second model makes usues of data which provides the average performance of a specific demographics across different runs of the module to predict the expected performance of a specific demographics for the next run of the module

## 1.2 Background

The data is collected by the future learn website where the online course is hosted. It consists of 7 different run of the online course produced between the years 2016 - 2018.

## 2. Methodology

### 2.1 Monitoring attendance between different demographics

Here we are looking at the amount of pupils enrolled within the online course, between different demographics.

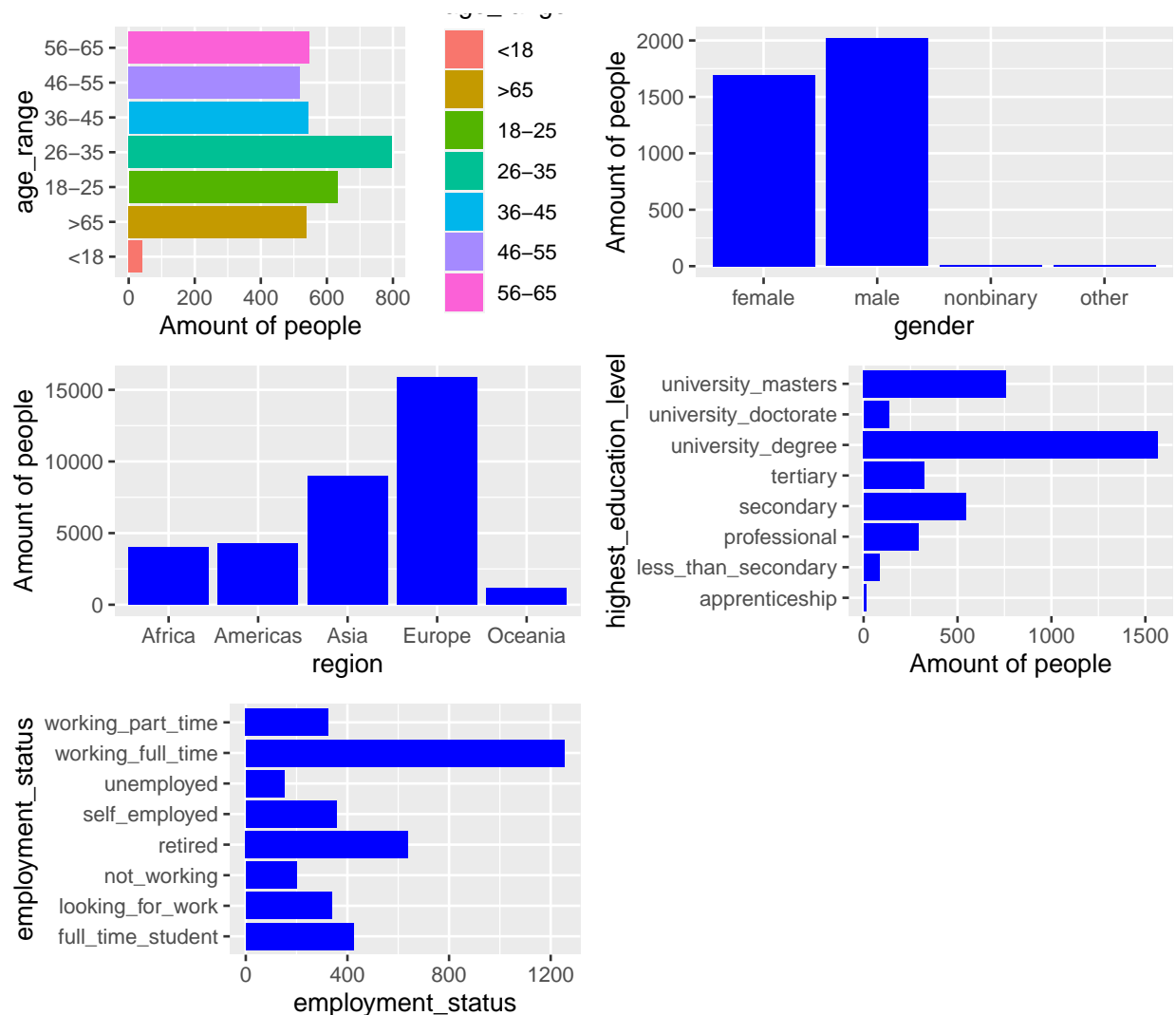


Figure 1: Bar graph illustrating the number of enrollments between different demographics

This graphs is a summary of the total 7 different runs of the course.

From the data it is found the highest number of pupils enrolled in the online course are aged 26-35. While, the lowest people less than 18. More male pupils are enrolled in the course more than female. Approximately 2000 males and 1750 females enrolled. In terms of locations, Europe has the highest number of pupils enrolled in this course, the second highest being pupils from Asia. Lowest, is Oceania. Pupils enrolled in this course, mostly have at least a university degree and are working full time.

Overall, the data illustrates that a significant amount of people who choose to enroll in this course appears to be between the ages 26-35, working full time with a high university degree from Europe.

## **2.2 Monitoring performance between different demographics**

### **2.2.1 Illustration of the performance across each run of the module**

The graph is an illustration of the demographics of pupils performance across different runs of the online course.

It appears that females pupils performed better than male pupils on runs 2, 4, 5 and 7. Male pupils performed better on runs 1, 3 and 6. Pupils who identify as non-binary tend to perform better than male and female pupils. However, there the population of non binary pupils within the data, is not large enough to support this claim.

Pupils from countries in European and Oceania performed better than pupils from countries Asia, Africa and Americas. European pupils appears to perform consistently the same across the 7 runs of online course, while Oceania significantly performed better than all the other regions on run 4 of the course.

Those who are under 18 appear to significantly perform the best. However given thht the population of pupils under 18 within the data is small, this claim is not supported. Therefore, we disregard this data. Pupils aged 56-65 and 65> tend to perform better.

There is no significant distinction in who performs between pupils with different forms of highest education. However, those with a highest form of education being a university degree appear to perform consistently through each run.

Across each run, there does not appear a significant distinction in performance between pupils with different employment status. However, pupils who are retired perform at a consistent rate through each run of the course.

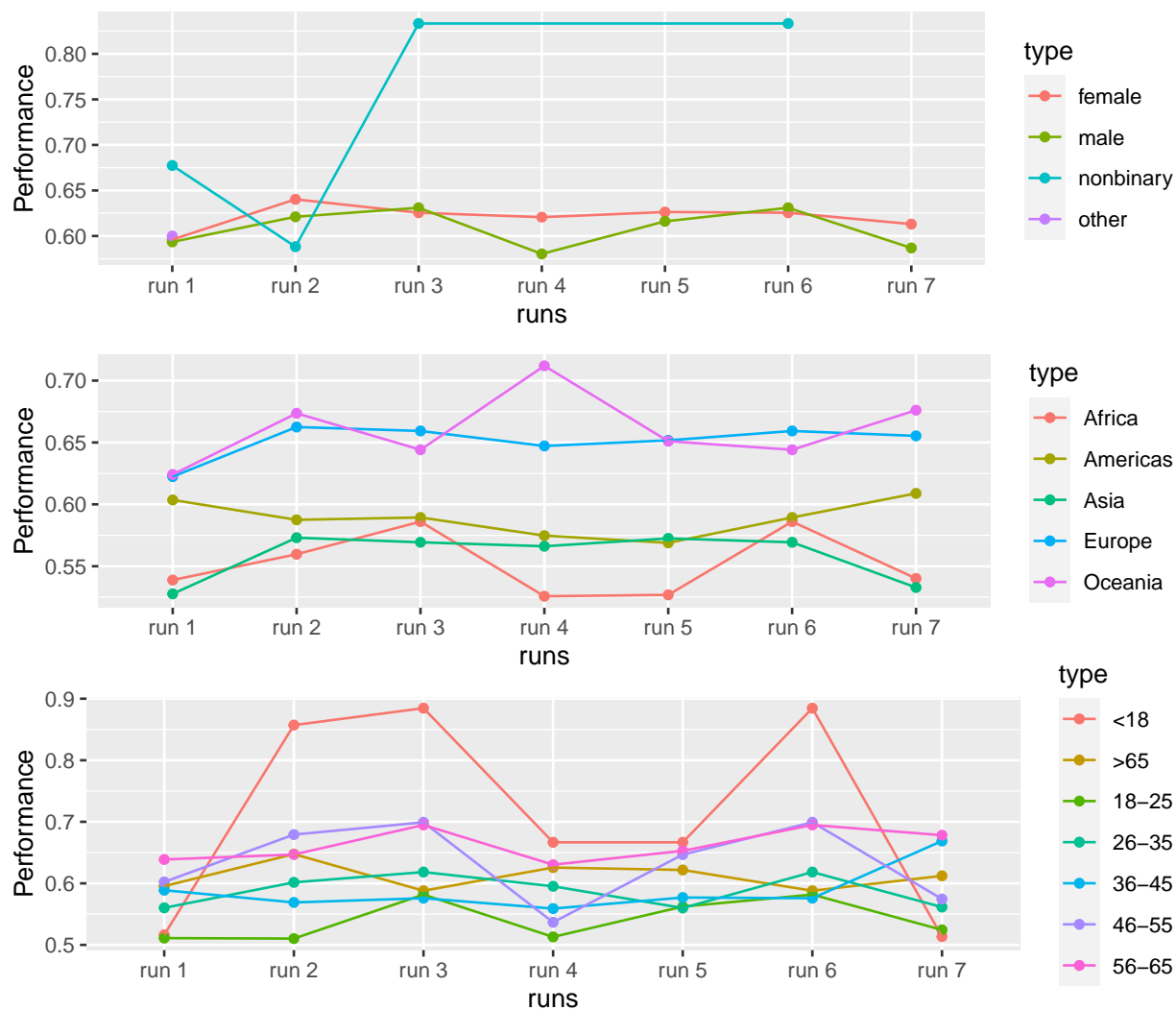


Figure 2: The average performance across different runs of the module compared between different demographics

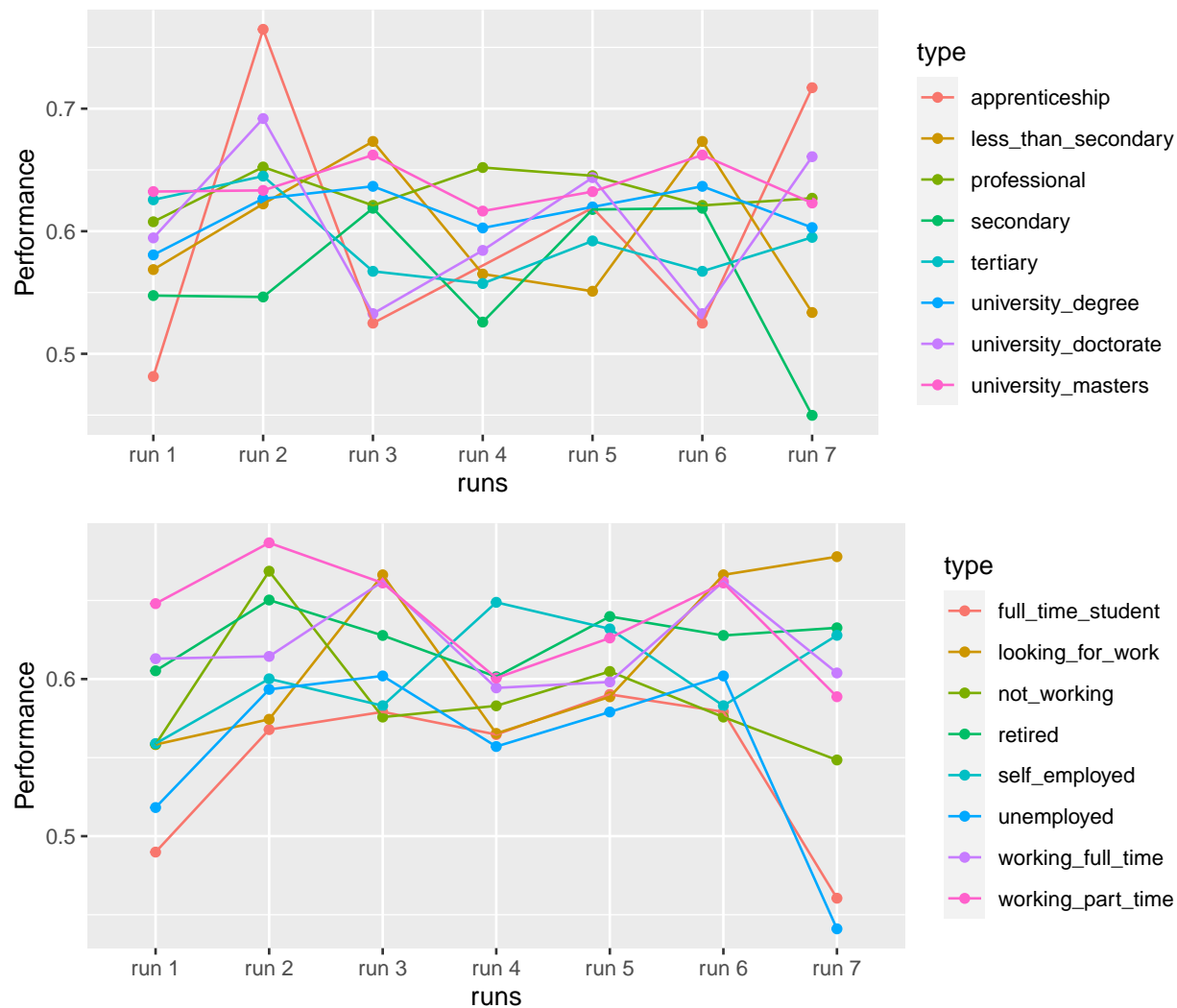


Figure 3: The average performance across different runs of the module compared between different demographics

## 2.2.2 Illustration from the total run

This summarizes data from the different runs of the online course, to compare the average and distribution of different demographics performance.

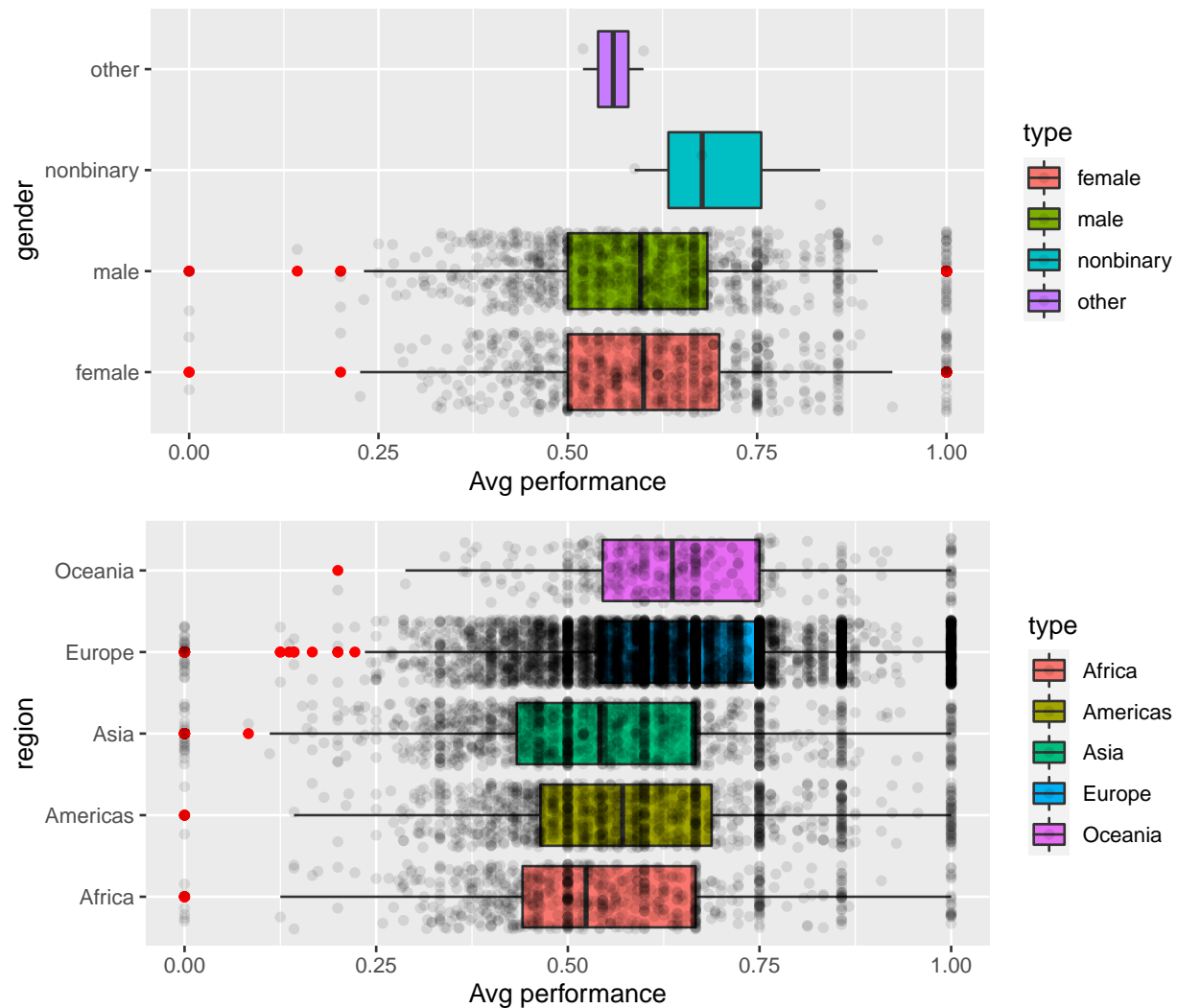


Figure 4: Boxplots illustrating the distribution of the average performance between different demographics

There is more variance in the gender of male students than they are for female students. Furthermore, The mean of male and females students are approximately the same.

Pupils from Europe have a higher mean than pupils from other regions. Additionally the data suggest that pupils from Europe are suited at the higher end of performance.

As mentioned before, there is not enough pupils that are less than 18 to precisely define them. Therefore, this information is disregarded. However, from the data it shows that pupils who are more than 65 have a average higher performance. The second highest

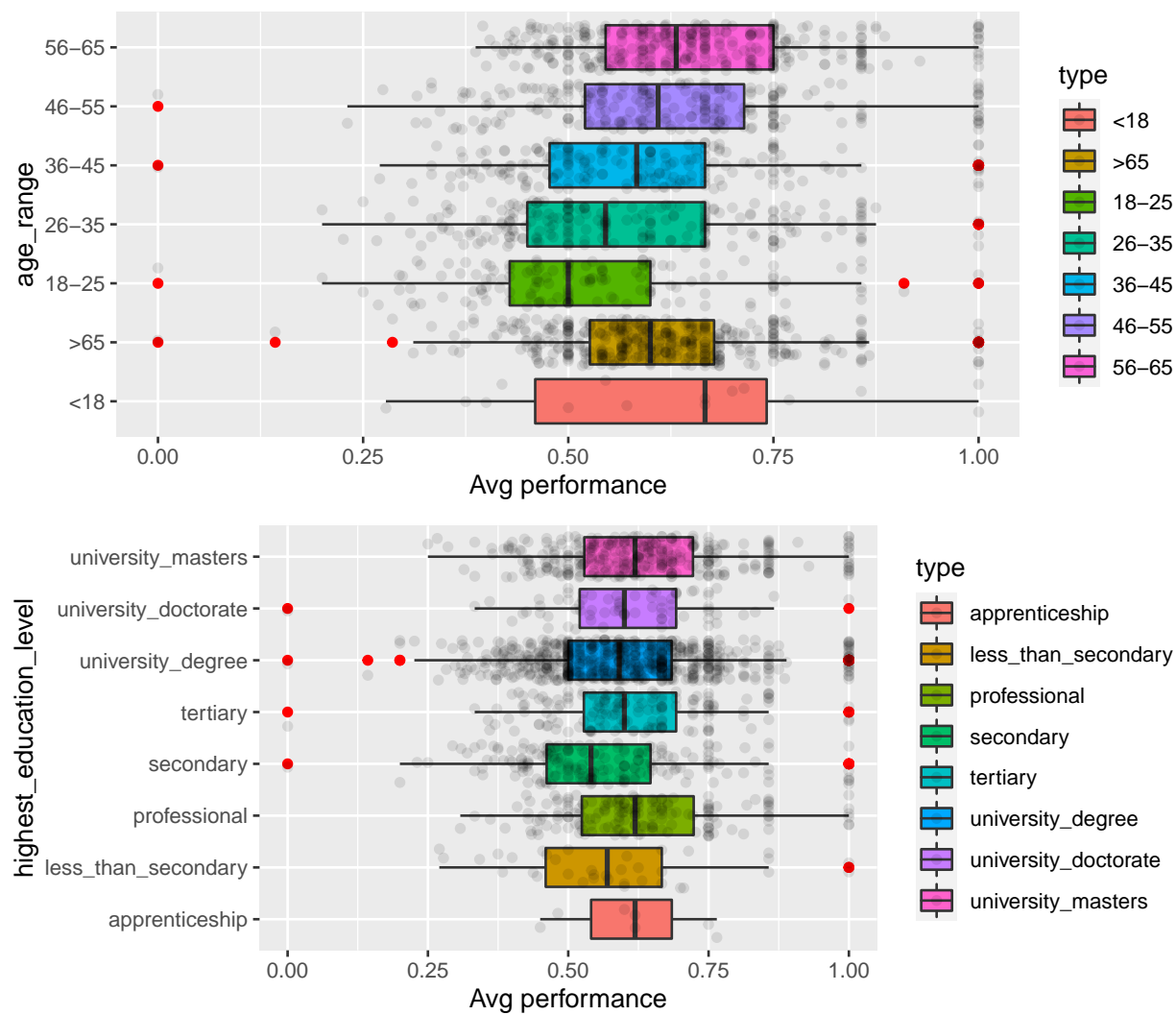


Figure 5: Boxplots illustrating the distribution of the average performance between different demographics

average performance are those who are between ages (56-65). Furthermore, pupils who are more than 65 have a low IQR, suggesting that the scores between each individual is similar/consistent. Pupils who are ages between 18 and 25 have a lower performance average. The graph, illustrates that the distribution of pupils aged 18-25 are situated at the lower end the average performance.

The pupils with the highest education level being an apprenticeship have the highest average performance. However, as the data from those who have an apprenticeship is not enough to accurately represent the population, the data is disregarded. Pupils with a university have the highest average performance, while the second highest average performance comes from pupils with a tertiary background. Pupils who have obtained a secondary schooling at a highest level education have an low average performance.

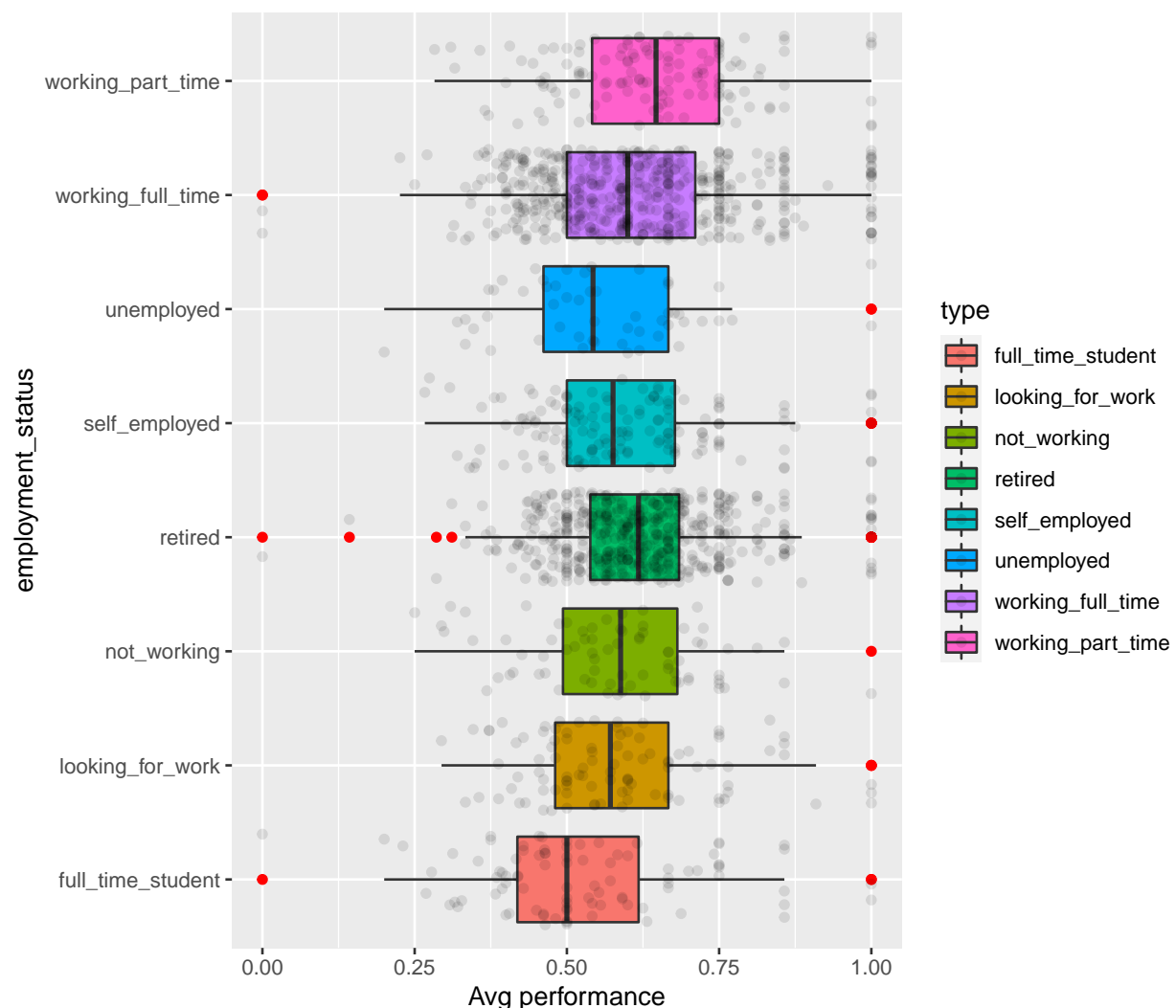


Figure 6: Boxplots illustrating the distribution of the average performance between different demographics



Pupils working full time have the highest average performance. This is followed by pupils who have retired. The distribution of people who have retired have a small IQR. This suggest that the performance obtained by each individual is fairly similar. Pupils who are full time students have a low average performance. Additionally the distribution of full time students is situated at the lower end of the average performance.

## 2.3 Student engagement

There is no clear strong pattern between the number of activities completed. However there it shows a slight negative correlation between the number of activities and the performance of each student.

This is looking at student engagement across different demographics. Student engagement is measured by the average number of quizzes completed in a week including the average number of activities completed in a week.

It appears that pupils ages 65 have the highest number of activities completed. The second highest is those who are aged 56-65. The lowest number of activities is completed is by pupils aged 18-25. Second lowest being pupils from age 26-35. Those who perform the highest number of quizzes per week are those aged 26-35. Those who perform the lowest number of quizzes per week are those aged 56-65.

Those who have retired have a higher number of number of activities completed on average. The second highest is those who are not working. Those who are retired complete the lowest amount of quizzes per week, while full time student complete the highest number of quizzes per week.

Pupils who are either non-binary complete the highest number of activities and lowest number of activities. However since the sample size from this population is small, the results are inconclusive. This hold true for pupils who identify as "other". Female pupils tend to have complete more activities in a week than Male pupils. The data further illustrates the female pupils complete less quizzes in a week than male pupils.

Pupils who have an apprenticeship tend to complete more activities and less quizzes than all the different demographics of higher education. However, given that the sample of the population collected for pupils in apprenticeships is small. This data is disregarded. Therefore, pupils who have a university doctorate degree have the highest number of activities completed in a week The second highest are Pupils who are professionals. However, professional pupils complete a lowest number of quizzes in a week.

Pupils from countries in Oceania complete the highest number of activities a week. The lowest number of activities completed are pupils from countries in Africa. Pupils with the lowest number of quizzes completed are from countries in Europe. Pupils with the highest number of quizzes are from countries in Africa

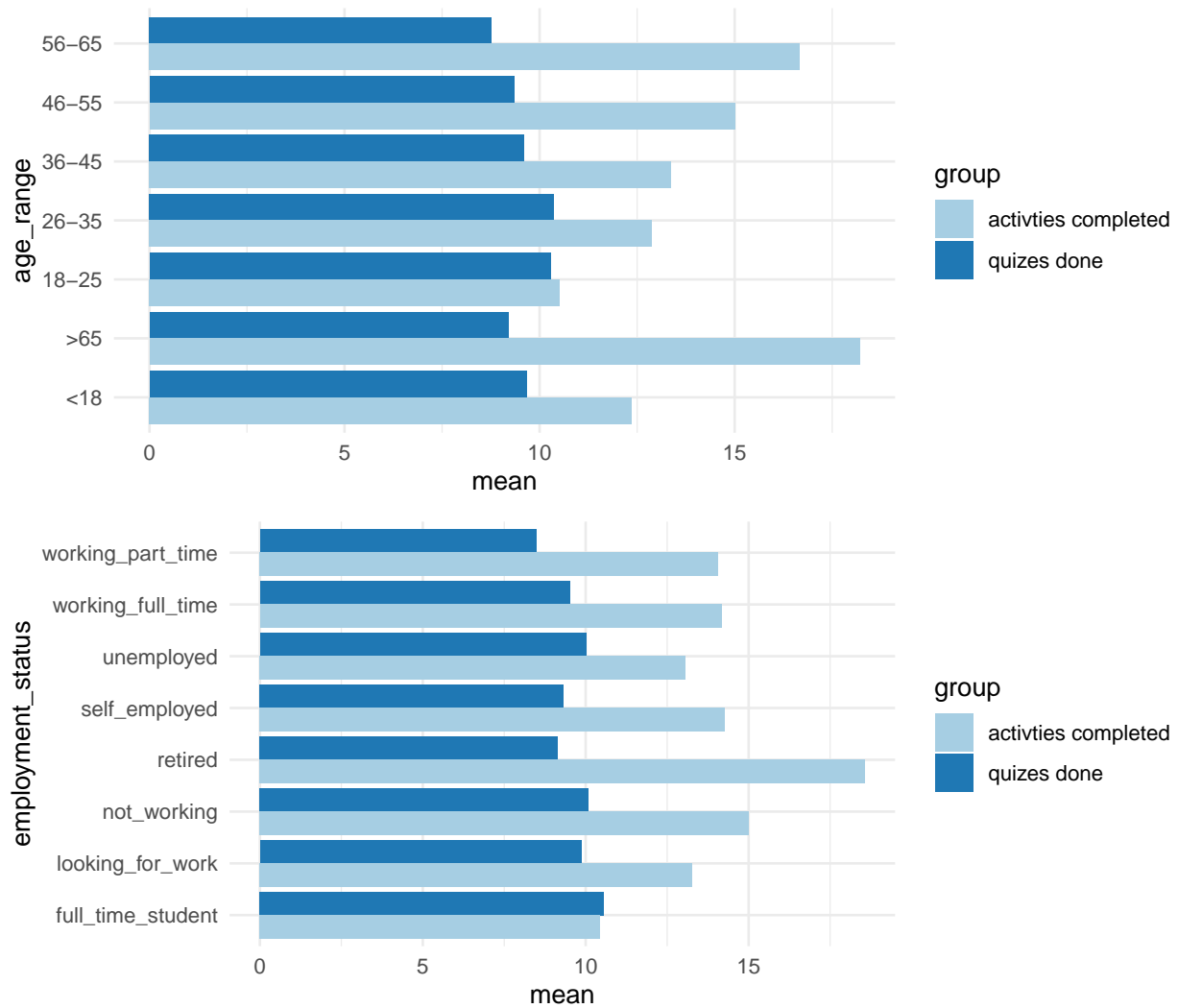


Figure 7: Bar graphs illustrating student engagement measured using number of quizzes and number of activities completed.

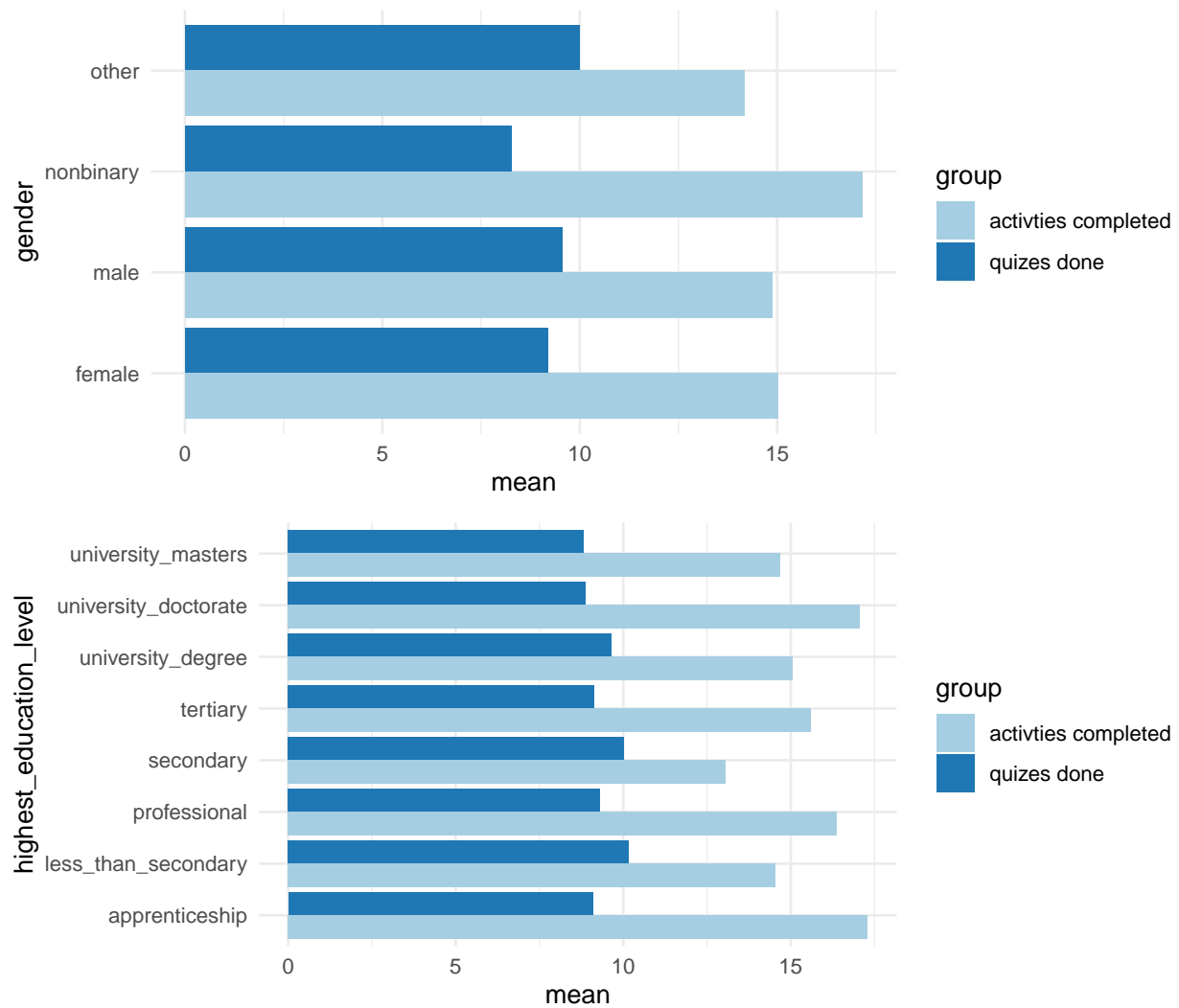


Figure 8: Bar graphs illustrating student engagement measured using number of quizzes and number of activities completed.

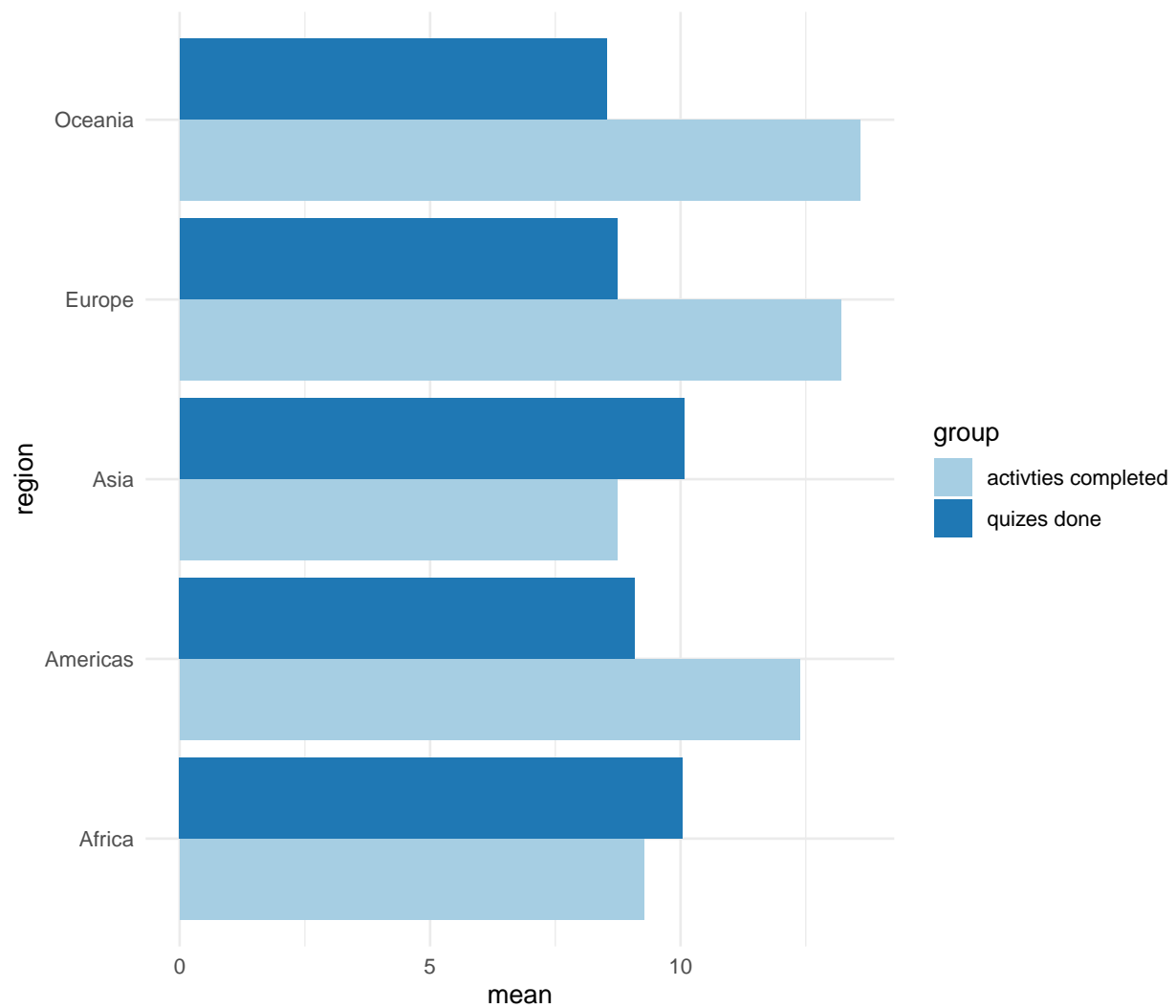


Figure 9: Bar graphs illustrating student engagement measured using number of quizzes and number of activities completed.

## 2.4 Conclusions and Reasoning

The people who had a better higher performance included:

- Pupils aged 56-65 and 65 tend to perform better.
- Those who are retired and or working full time perform better
- Full time students perform the lowest
- Pupils with university degree perform better
- Pupils from oceania

Generally it was observed that a pupil will complete more quizzes per week have a lower performance. This is because pupil with a higher number of quizzes tend to make attempts in multiple combination until an answer is correct. This suggest that they may have not gained the right understanding and are most lickley guessing. Additionally, it's observed that the number of activities completed and performance has less of a negative relationship than number of quizzes completed per week and performance. Perhaps this is indication why some demographics (>65) had a higher performing (as they took much less attempts to answer questions) and activities. Activities take more effort than quizzes, as involves self research to find an answer. That indicates those who are more engaged take more activities, thus gaining a better understand and a higher performance in quizzes (less reattempts).

Full time students had a lower performance and completed more quizzes than any pupils with different employment status. This holds true for pupils with the age range 18-25. Retired people tend to perform better among est those with different employment status. The data also suggest that pupils >65 perform the best. Therefore, it is reasonable to suggest that full time students tends to be those 18-25 are less engaged because as they are completing other studies.

```
## 'geom_smooth()' using formula 'y ~ x'  
## 'geom_smooth()' using formula 'y ~ x'
```

## 2.5 Models

We are going to be looking at implementing two models 1. Predicts the number performance by each student 2. Uses data to estimate the performance of a student

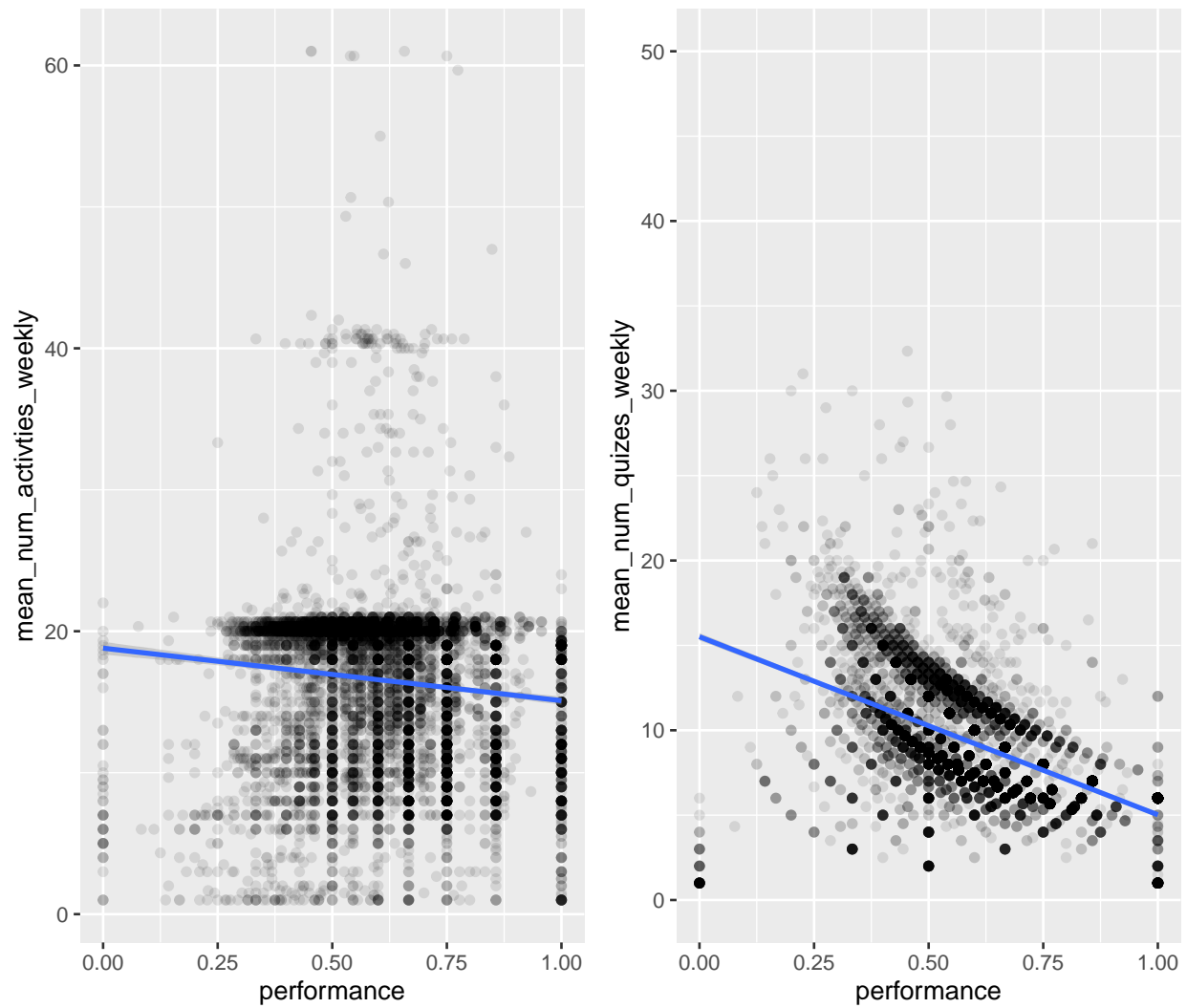


Figure 10: Scatter graphs illustrating the relationships between the performance and the average number of quizzes completed and the relationship between the performance and the average number of activities completed weekly

## Model 1:

This model takes in two main datasets - demographic data, including the gender, age, highest level of education, employment status, region and employment area. Additionally makes use of the data specifying the number of engagements. This is measured by the number of quizzes and the number of activities completed by a pupil week by week.

The demographic and engagement data is put into a linear regression model. The output is the prediction of a pupil performance.

This model can be used to provide more insight on how a student is performing, and identify any students not gaining the correct understanding. This information can then be used to decide on the suitable intervention for a pupil.

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	0.6765835738	0.0347316152	19.4803371	1.735459e-70
## gender	-0.0015965316	0.0087297626	-0.1828837	8.549316e-01
## age_range	0.0072371936	0.0025583080	2.8288984	4.777634e-03
## highest_education_level	0.0023834214	0.0027083278	0.8800343	3.790824e-01
## employment_status	0.0058465793	0.0022726461	2.5725868	1.025799e-02
## employment_area	-0.0017601316	0.0007667036	-2.2957133	2.192798e-02
## region	0.0215858825	0.0042068916	5.1310765	3.549714e-07
## avg_quizzes_weekly	-0.0227612559	0.0013258477	-17.1673228	3.731376e-57
## avg_activities_weekly	0.0006901234	0.0007445473	0.9269033	3.542323e-01

## [1] 0.01427169

## [1] 0.01653045

## Model 2

The idea of this model is a predictive analytic model that makes use of the average performance of each specific demographic from each run of the module. The example on figure 12, is predicting average performance the next run (8) by different regions. This predictive model can be also used for other demographics (age, gender etc).

This model can only predict 1 future run of the module. This is because there is only 7 observations for all runs. Therefore to predict more, more data (more runs of the course) would need to be done. There is a low signal-to-noise ratio, therefore it was difficult for the predictive model to extract any possible trend, persistence, lagged errors etc.

## Evaluation

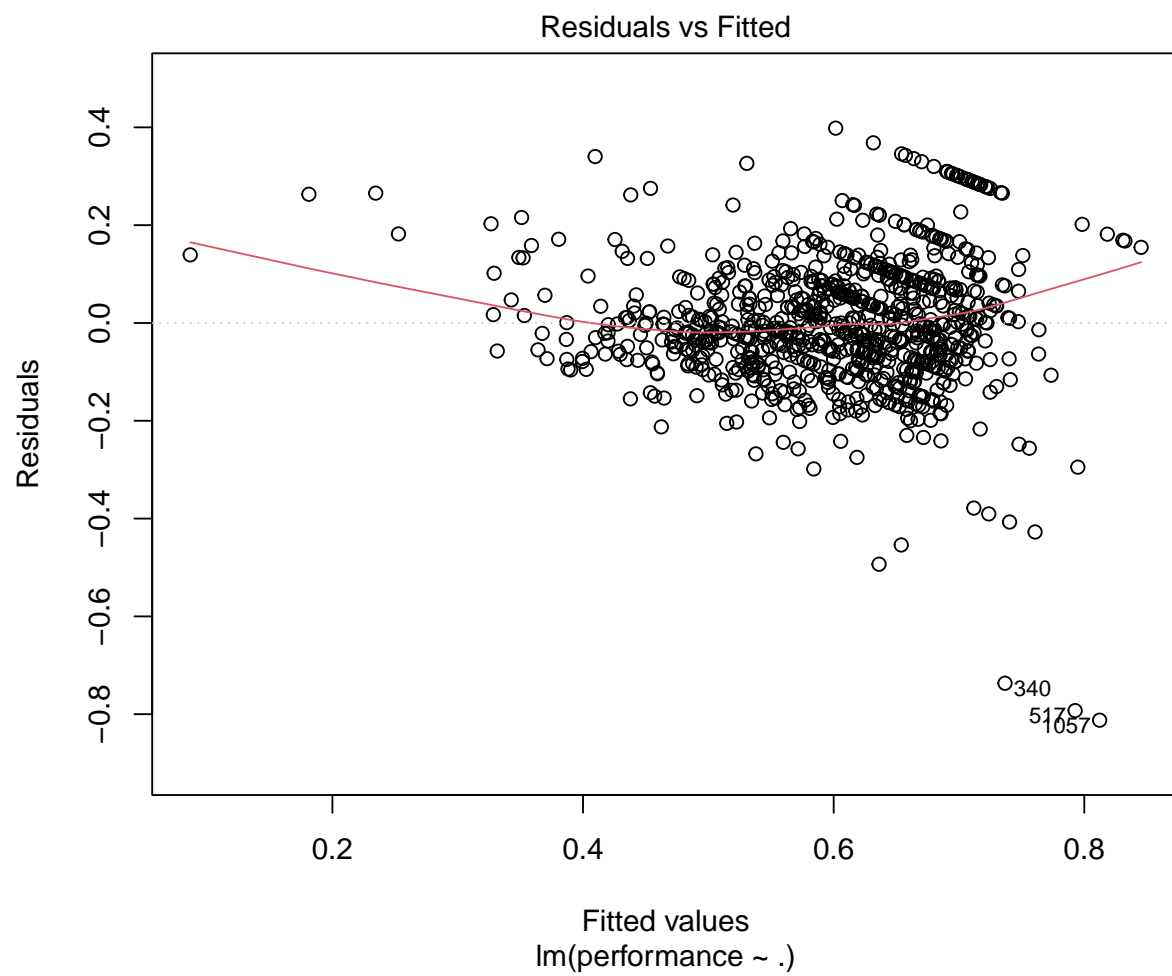


Figure 11: Model 1: predicting the performance of a pupils using demograp



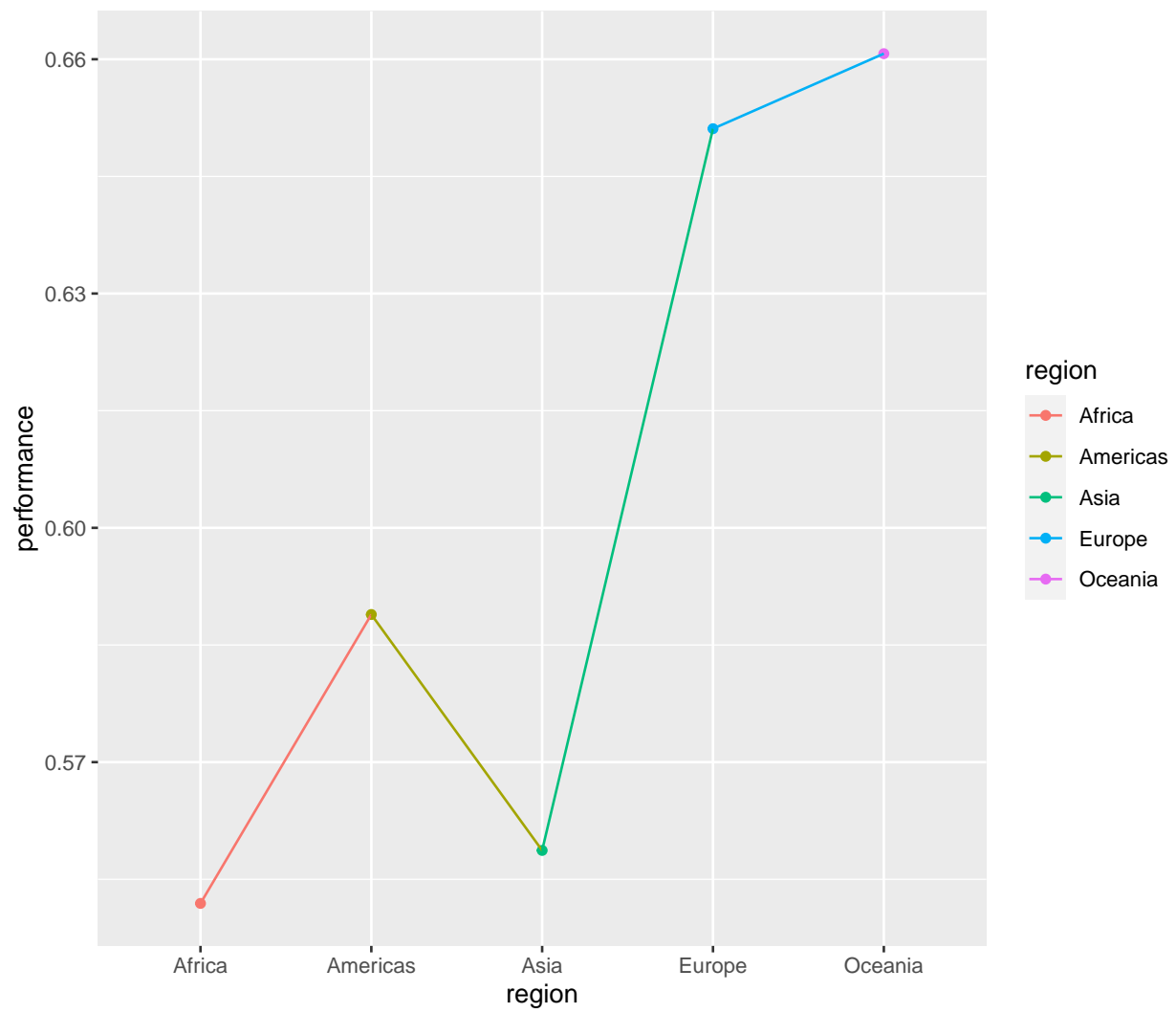


Figure 12: Model 2: Forecasting performance