REPORT

INSIGHTS FROM THE OPEN UNIVERSITY LEARNING ANALYTICS DATASET: EXPLORING STUDENT ENGAGEMENT AND PERFORMANCE

Minh Nguyen (Brian)

I. Overview of Dataset

This report provides an in-depth analysis of the Open University Learning Analytics Dataset (OULAD). The OULAD dataset is an anonymized collection of data that includes information about courses, students, and their interactions with the Virtual Learning Environment (VLE) across seven selected courses, referred to as modules. The dataset offers insights into course presentations, which commence in February and October, identified by "B" and "J" respectively. The dataset is composed of several interconnected tables, each containing unique identifiers, and all tables are stored in CSV format.

The primary aim of this report is to explore the data contained in the OULAD dataset, with a specific focus on understanding student demographics, their engagement with the courses, and the outcomes of their assessments. By analyzing these aspects, the report seeks to uncover patterns and trends that can provide valuable insights into student behavior and performance.

Source: [Open Learning Analytics | OU Analyse | Knowledge Media Institute | The Open University]

Number of students registered for the modules and presentations

In total, there are 32,593 records in the student information and registration files. Out of these, there are 28,785 distinct students registered for the courses and assessments. Additionally, it's evident that some students took more than one module and assessment, 3,538 students are registered for 2 code modules, 245 students are registered for 3, 24 students are registered for 4, and 1 student is registered for 5.

num_students	threshold
28785	1
3538	2
245	3
24	4
1	5

The number of assessments

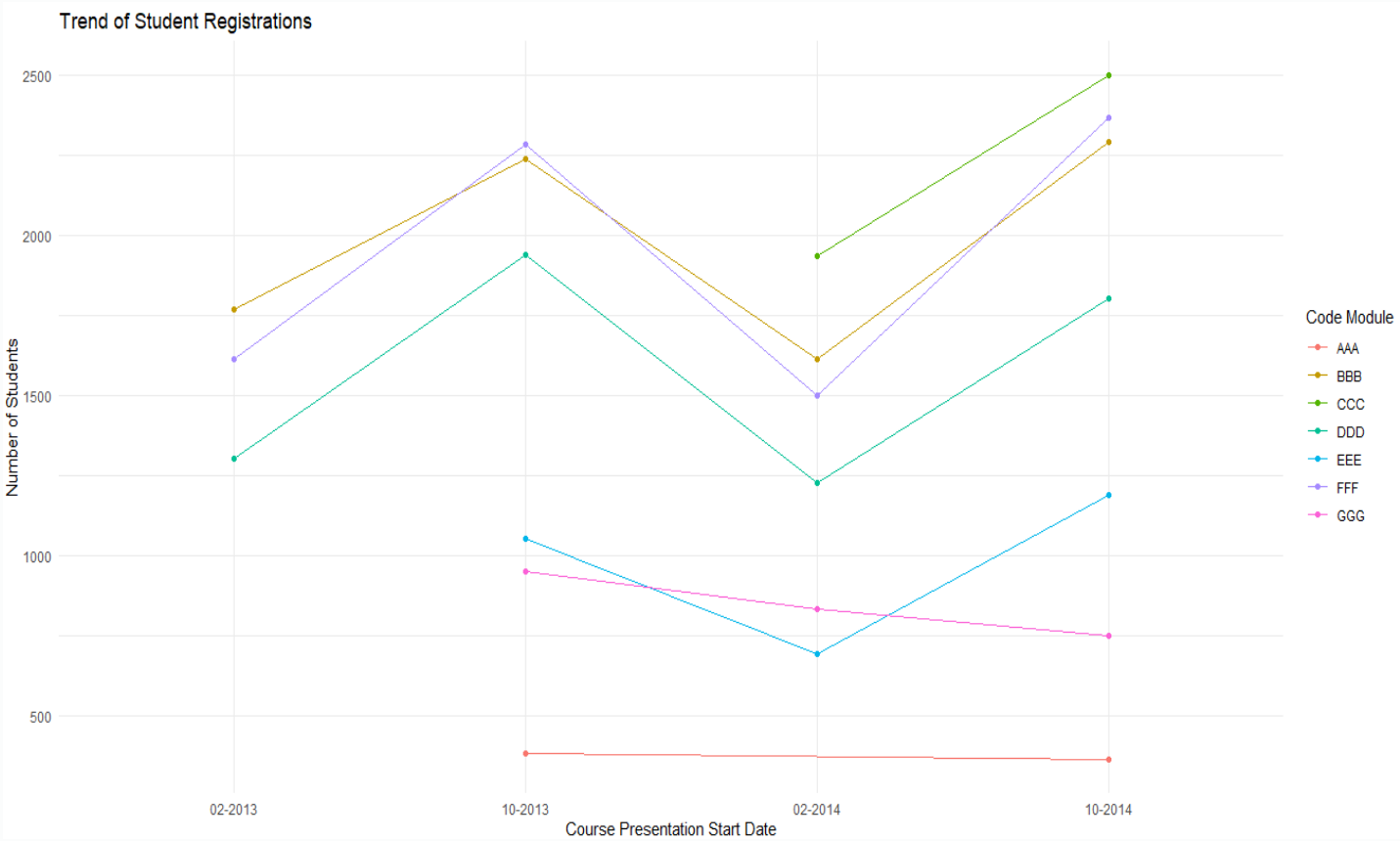
code_module	2013J	2014J	2013B	2014B
AAA	6	6	0	0
BBB	12	6	12	12
CCC	0	10	0	10
DDD	7	7	14	7
EEE	5	5	0	5
FFF	13	13	13	13
GGG	10	10	0	10

I. Overview of Dataset

Number of students registered for the modules and presentations

code_module	2013J	2014J	2013B	2014B
AAA	383	365	0	0
BBB	2237	2292	1767	1613
CCC	0	2498	0	1936
DDD	1938	1803	1303	1228
EEE	1052	1188	0	694
FFF	2283	2365	1614	1500
GGG	952	749	0	833

The number of students registered for modules and course presentations commencing in February 2013 and 2014 was lower than in October 2013 and 2014. In particular, the number of students was lowest in February 2013, as only three modules/courses were offered at that time. October 2014 had the highest number of students registered, with a full offering of seven courses.

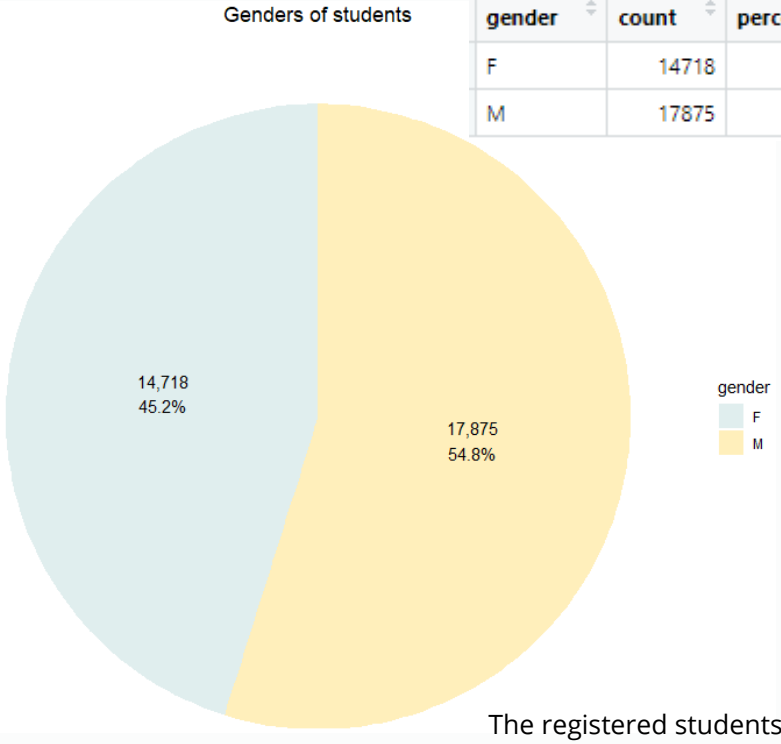


II. Demographic analysis

Genders

Registered students by genders

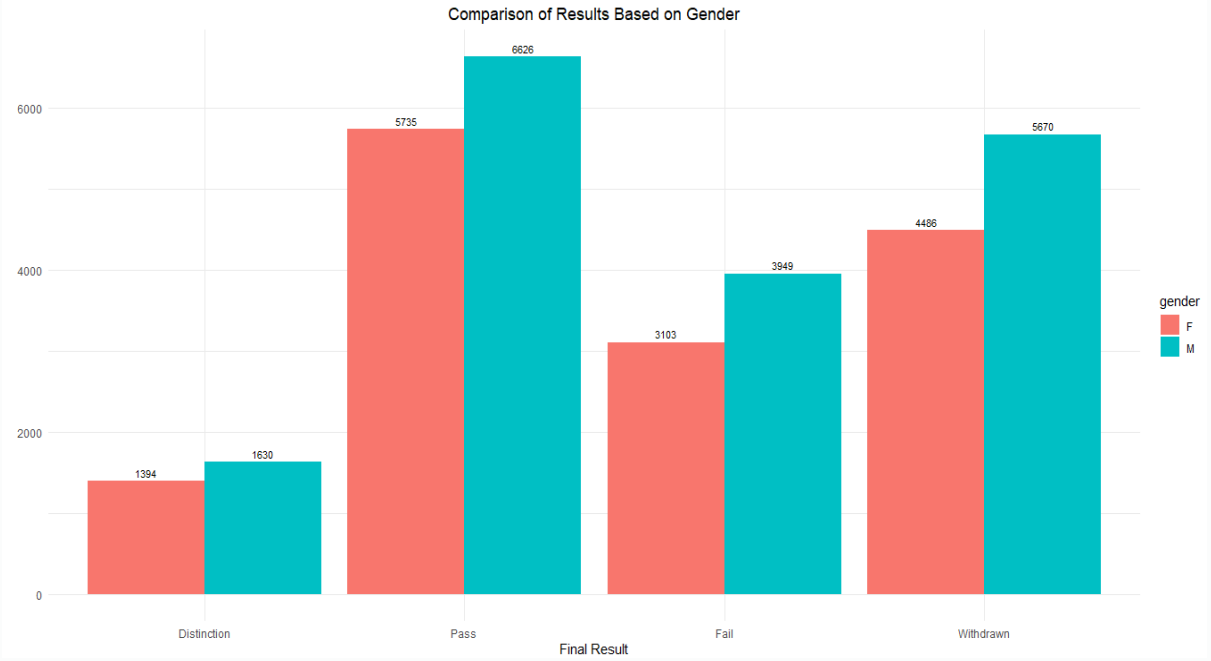
gender	count	percentage
F	14718	45.2
M	17875	54.8



The registered students include 14,718 females, making up 45.2% of the total, and 17,875 males, accounting for 54.8%.

Final results by genders

gender	Distinction	Fail	Pass	Withdrawn
F	1394	3103	5735	4486
M	1630	3949	6626	5670

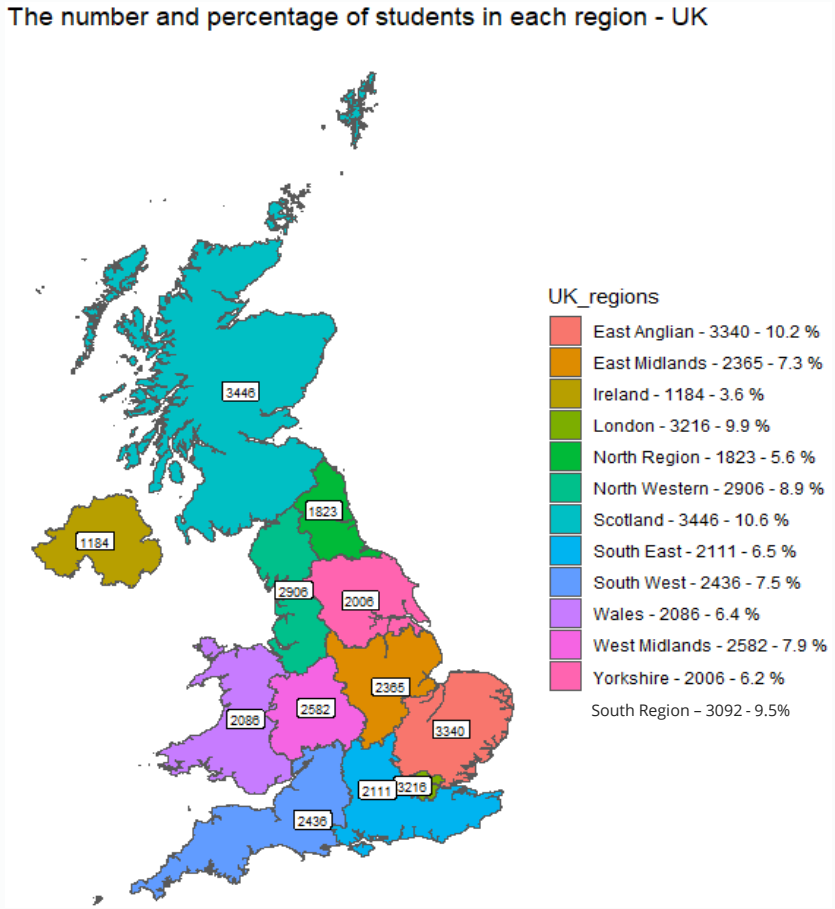


Note: The data and results are based on the number of records, not distinct students.

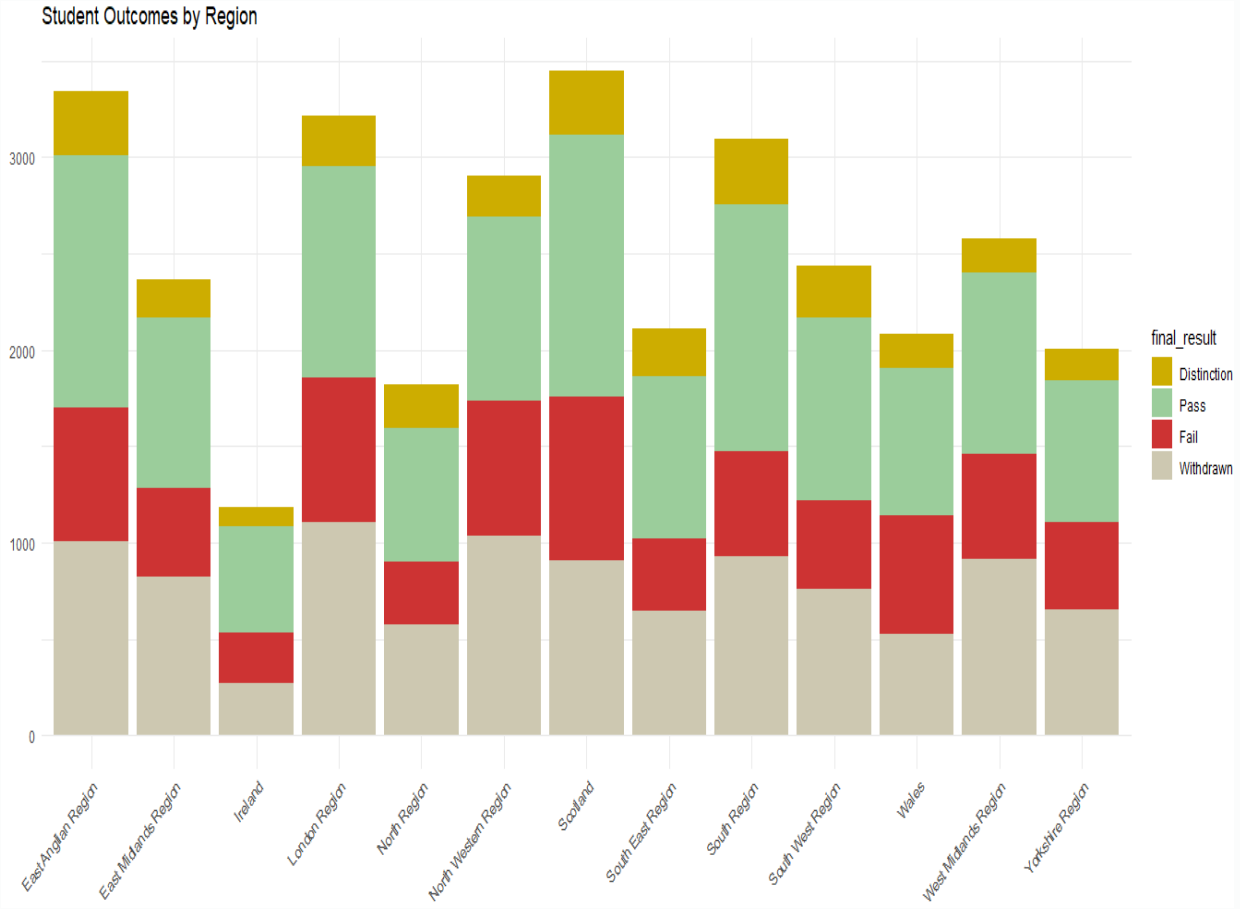
II. Demographic analysis

Regions

Registered students by regions



Comparison of results by regions

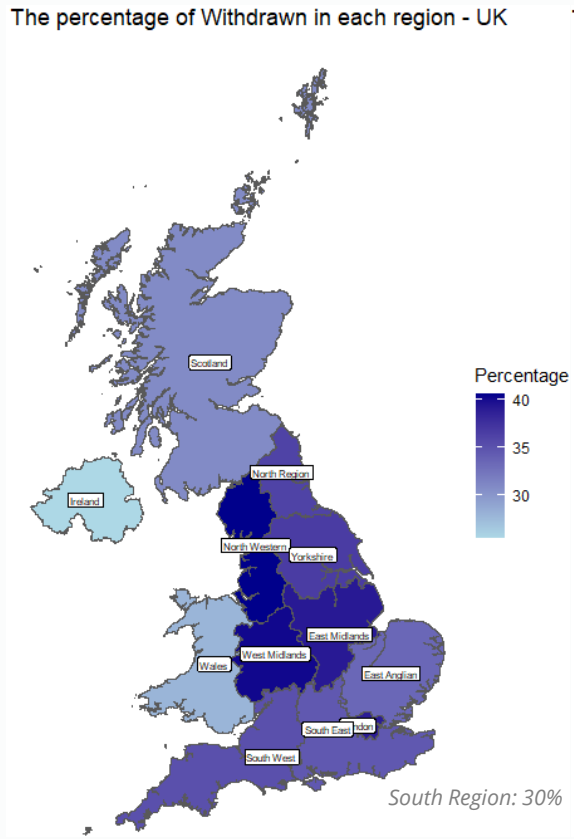


Note: The map not includes data for South Region due to lack of Geography data

II. Demographic analysis

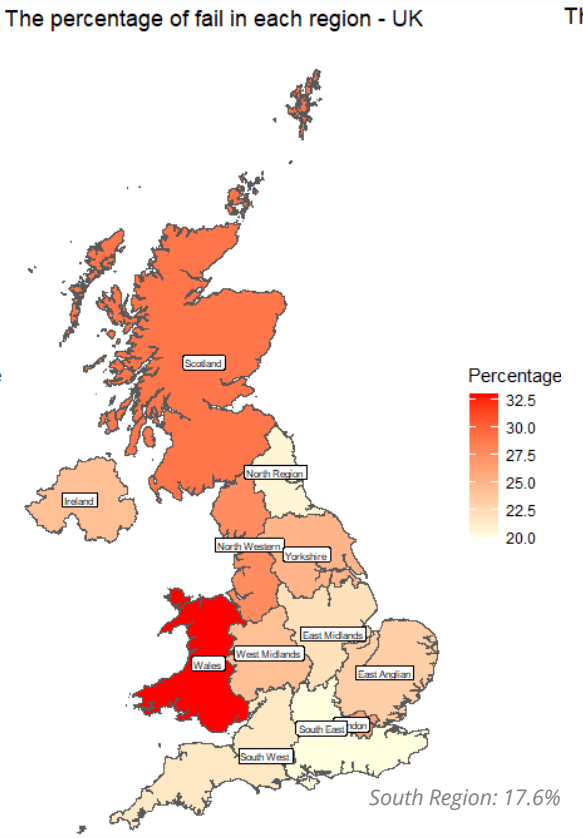
Regions

Withdrawn by regions



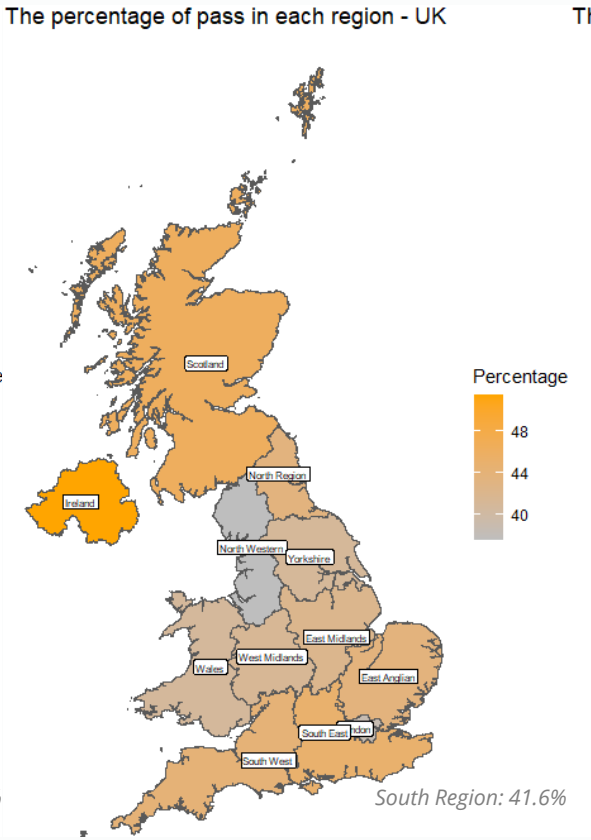
The North Western, Midlands (West and East), and London regions have a higher percentage of student withdrawals relative to the number of registered students in those regions.

Fail by regions



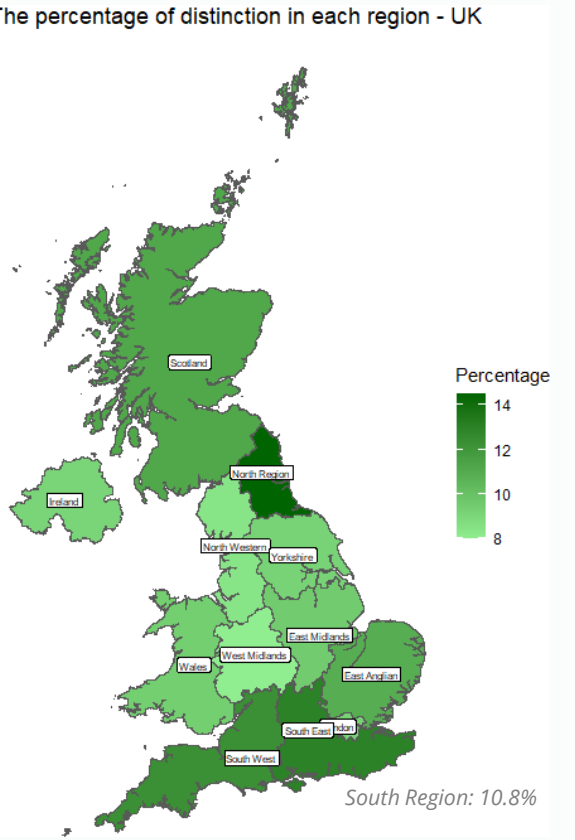
Wales and Scotland have a higher percentage of student failures, while the South East, North and South regions have a smaller percentage.

Pass by regions



Ireland and Scotland have a higher percentage of student passes, while the North Western region has the smallest percentage.

Distinction by regions



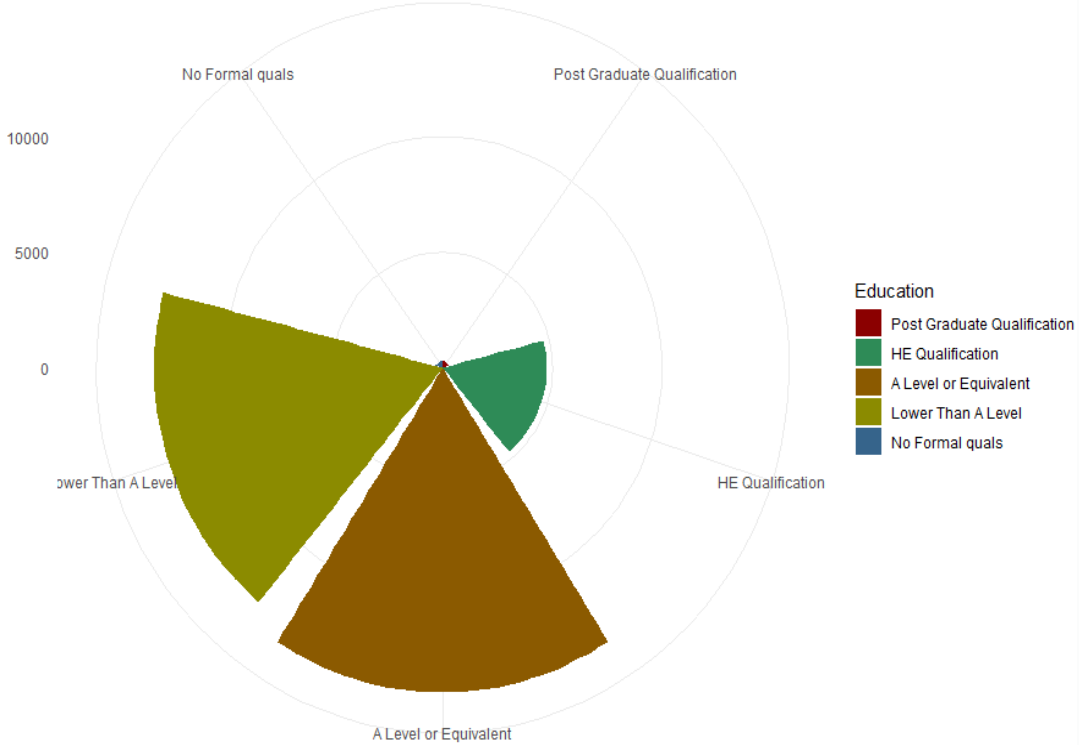
The North region and South East have a higher percentage of students achieving distinction grades, while the North Western and West Midlands have a smaller percentage.

Note: The map not includes data for South Region due to lack of Geography data

II. Demographic analysis

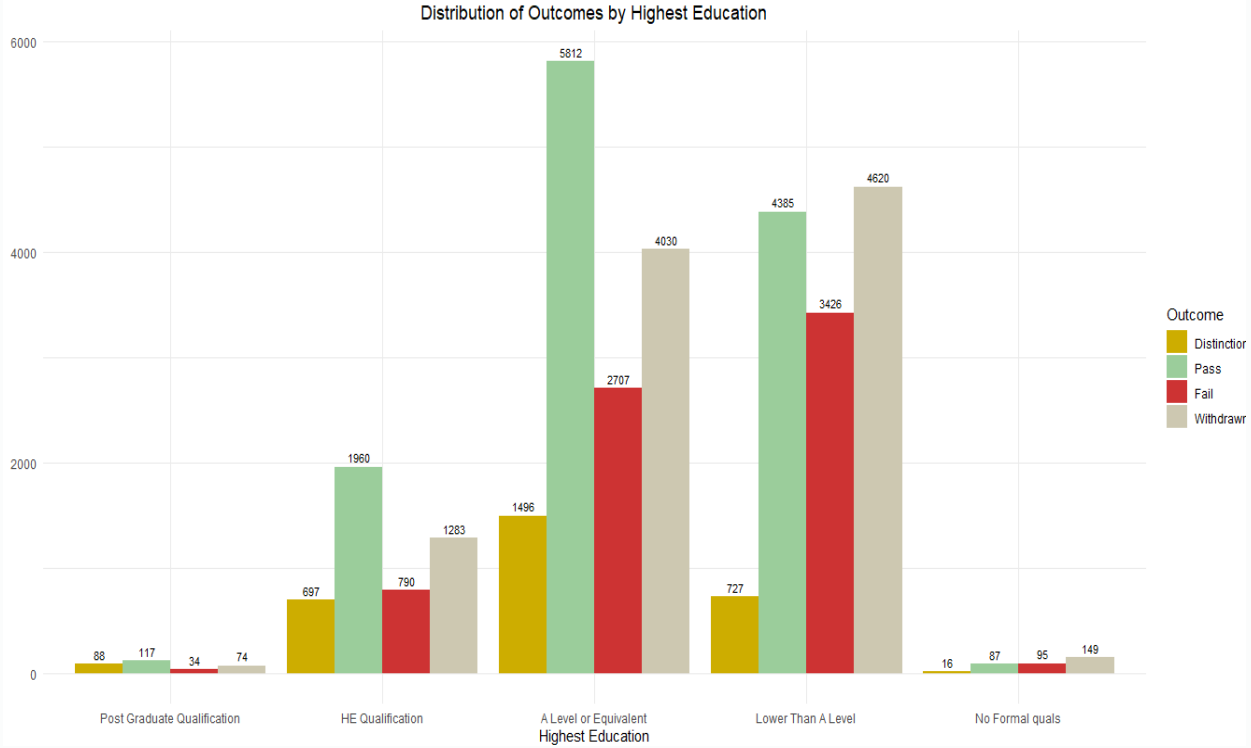
Highest Education

The Education qualification of students



Most students' highest education falls into categories such as "A Level or Equivalent" and "Lower Than A Level" whereas "Post Graduate Qualification" and "No Formal quals" have a notably smaller number of students

Distribution of Outcomes by Highest Education

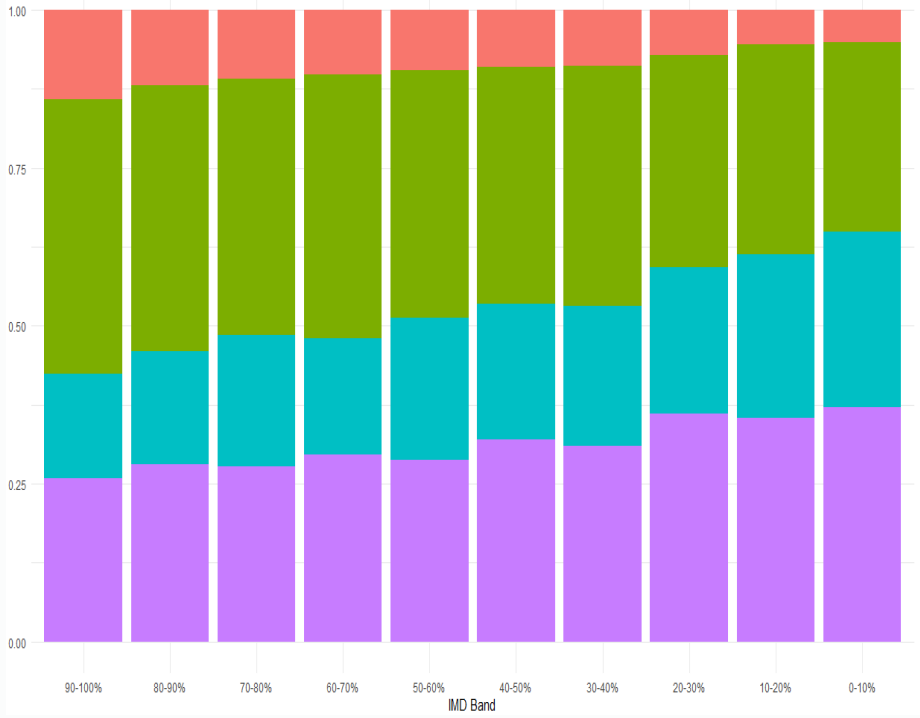


Students with the highest education levels of "No Formal quals" and "Lower Than A Level" tend to have a higher incidence of withdrawals and fails compared to other educational backgrounds. Conversely, students with "Post Graduate Qualification," "HE Qualification," and "A Level or Equivalent" education levels demonstrate higher rates of passing assessments and higher scores compared to other educational backgrounds.

II. Demographic analysis

IMD Band

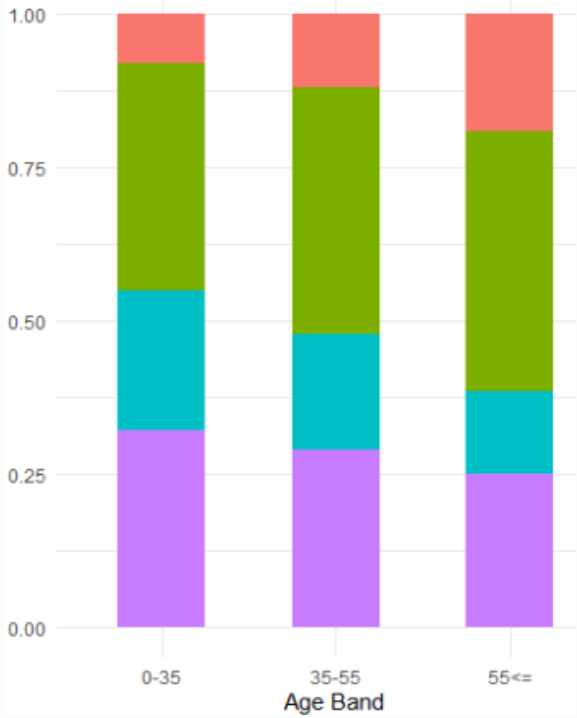
IMD band and results



Students with lower IMD bands tend to withdraw and fail more frequently compared to those with higher IMD bands. Similarly, they achieve lower numbers in pass and distinction results.*

Age Band

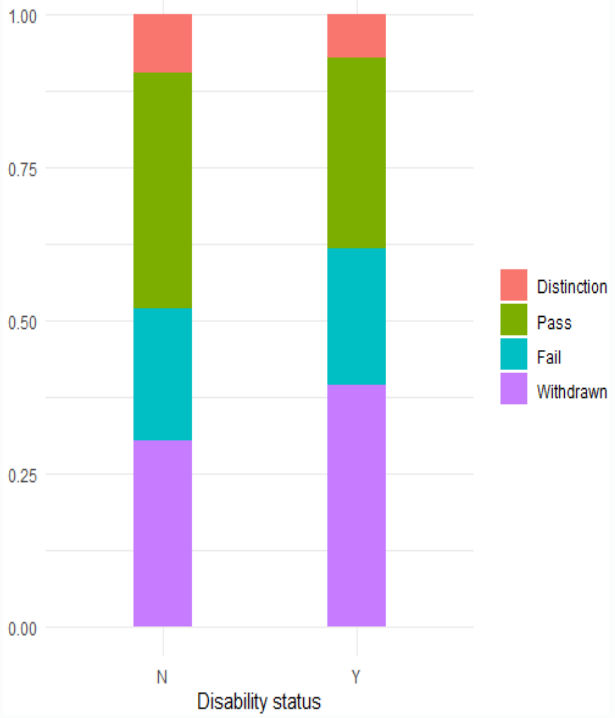
Age band and results



The higher the age band of students, the greater the number of students achieving pass and distinction results, and the fewer instances of fail and withdrawal.

Disability

Disability and results



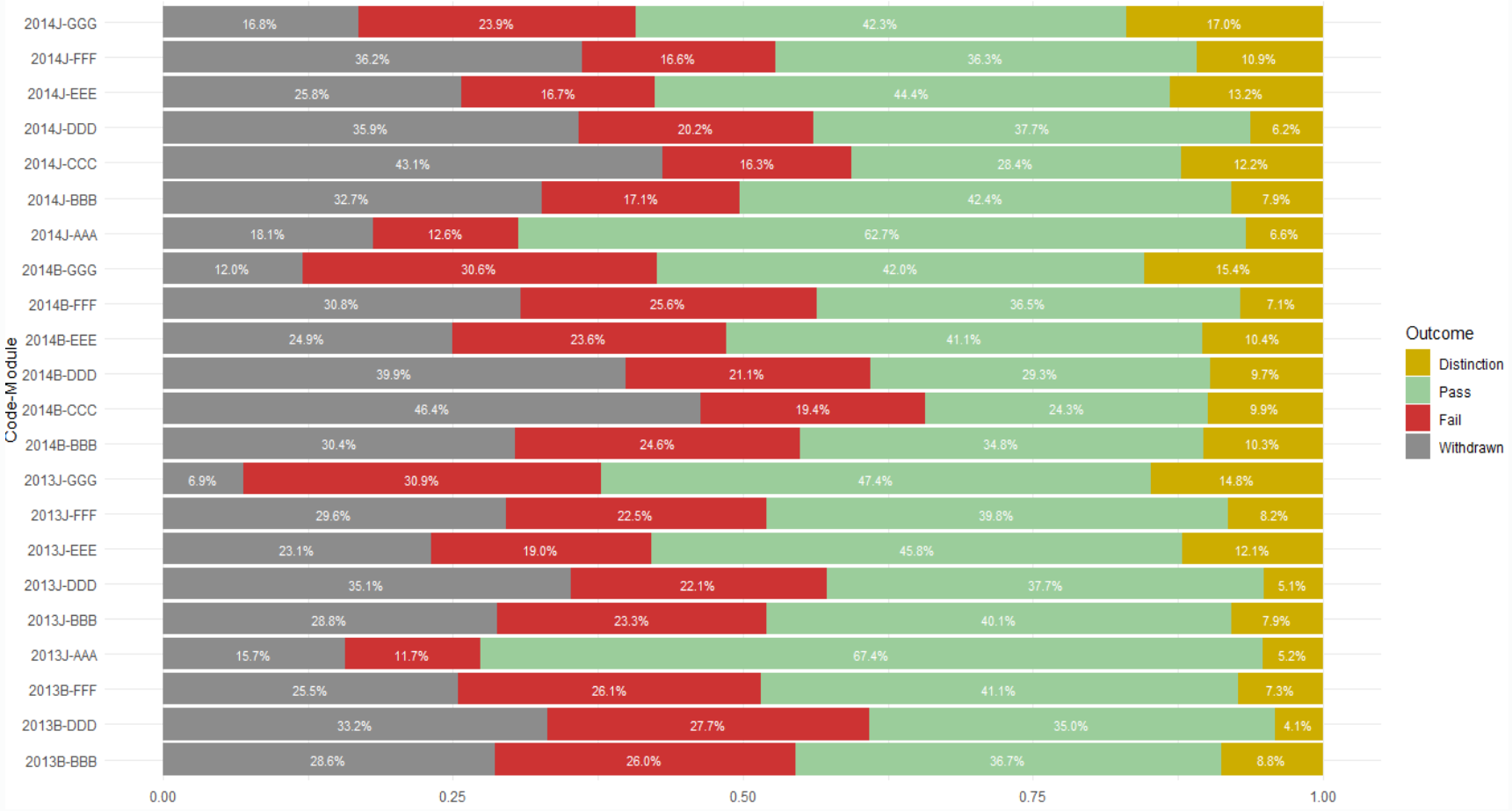
For students with disabilities, they tend to experience a bit higher rates of failure and withdrawal compared to non-disabled students, while achieving pass and distinction results less frequently.

*Note: *The plot excludes 1111 records (3.4%) without IMD band.*

III. Code module and Results analysis

Code module

Code module and results

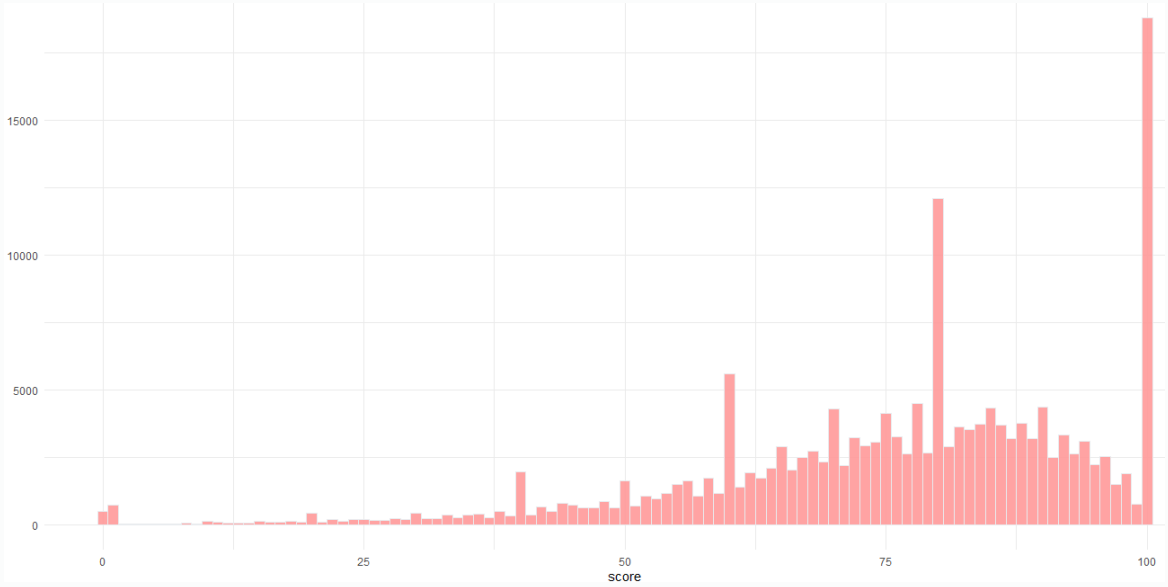


Across all code modules, the "2013J-GGG" module has the lowest percentage of student withdrawals (6.9%), whereas "2014B-CCC" and "2014J-CCC" have high withdrawal percentages, each exceeding 40%.
The percentage of students failing in the "2013J-GGG" and "2014B-GGG" modules is quite high, around 30%. In contrast, the failure rates in "2014J-AAA" and "2013J-AAA" are relatively low, at 12.6% and 11.7%, respectively.
The percentage of students passing in "2014J-AAA" and "2013J-AAA" is very high, at 62% and 67%, respectively, while "2014J-CCC" and "2014B-DDD" have significantly lower pass rates, at just 28% and 29%.
Finally, the percentage of students achieving a distinction in "2014J-GGG" is the highest at 17%, whereas "2013B-BBB" has the lowest distinction rate at 4.1%.

IV. Score and Sum of clicks - Virtual Learning Environment (VLE)

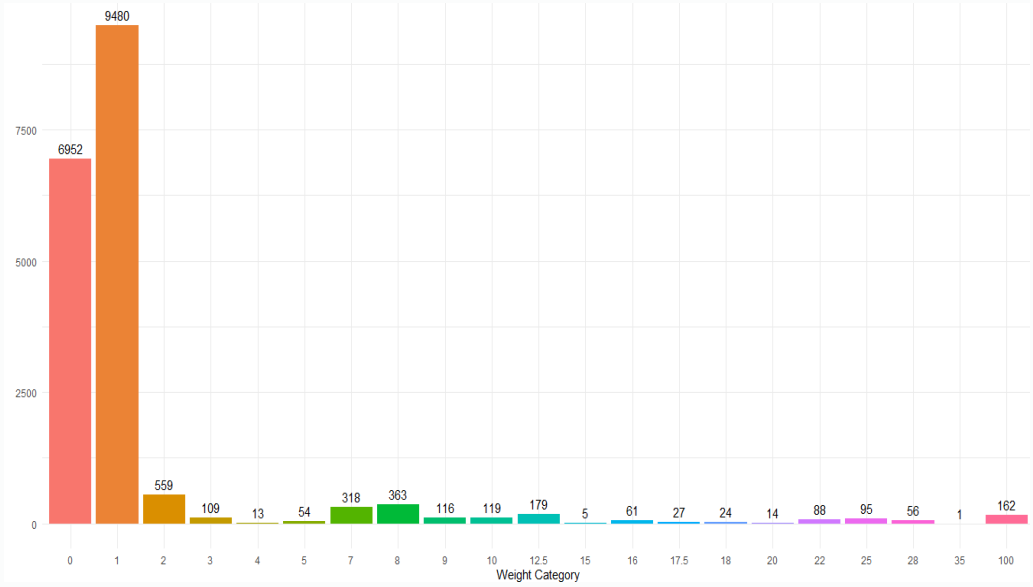
Score

Distribution of Score



In general, a large of number of students got score over 50, especially many students got score of 100.

Score of 100 by Weight Category

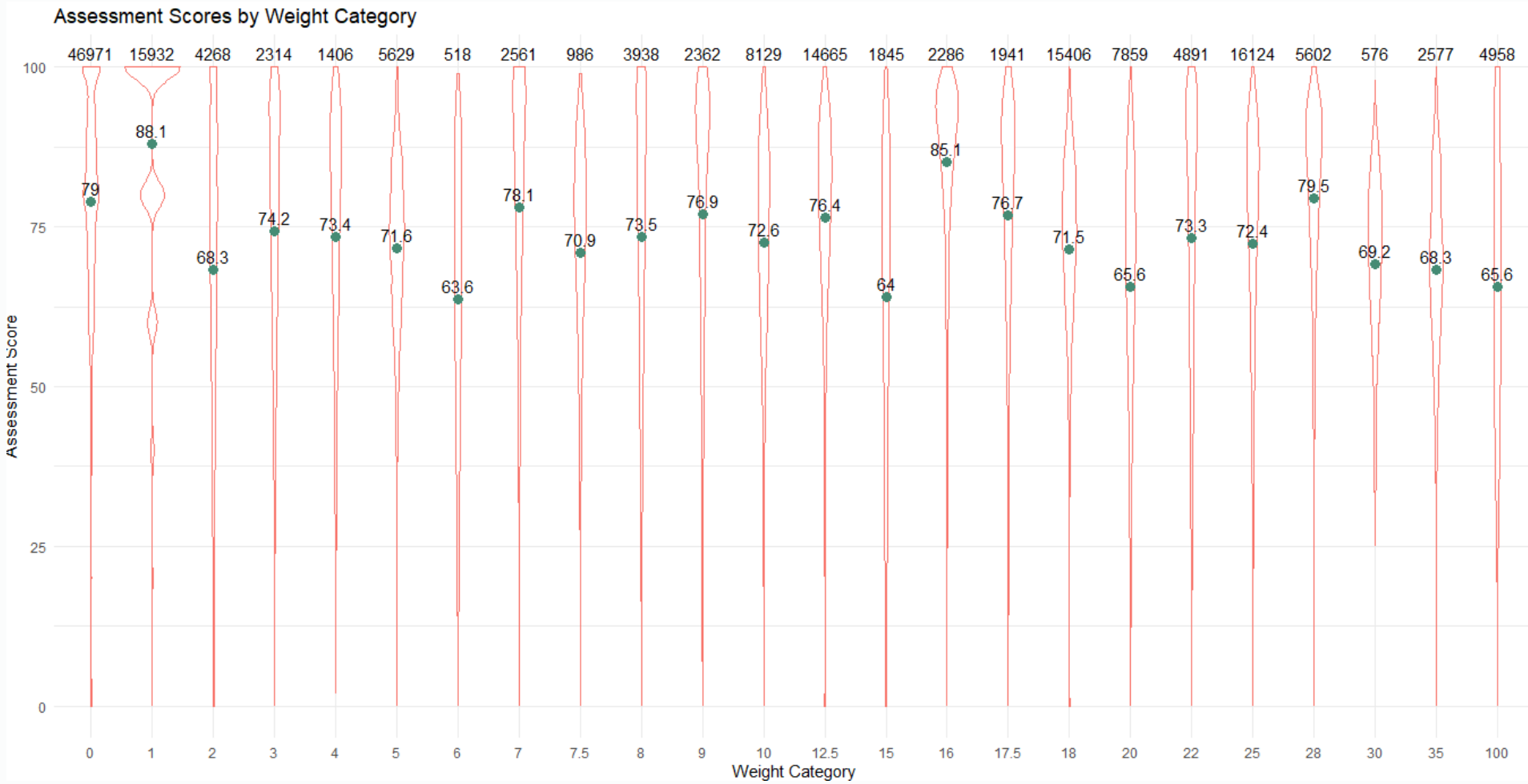


However, most scores of 100 are from assessments with a weight between 0% and 3%.

IV. Score and Sum of clicks - Virtual Learning Environment (VLE)

Score

Assessment Scores vs. Weight Categories

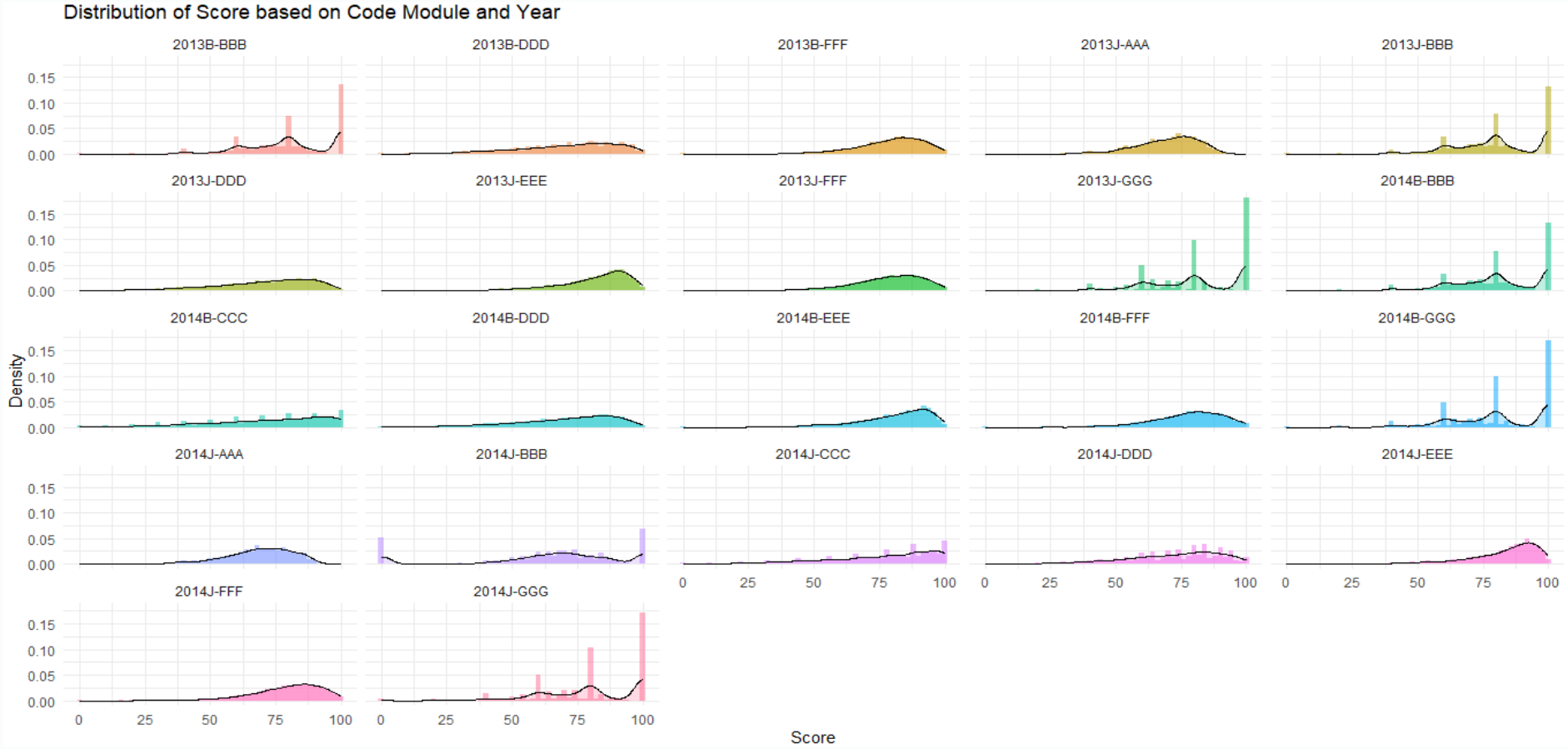


In general, students tend to achieve higher scores in assessments with a weight of 1% or 2%, which do not significantly impact the total assessment weight.

IV. Score and Sum of clicks - Virtual Learning Environment (VLE)

Score

Distribution of Score by Module & Year

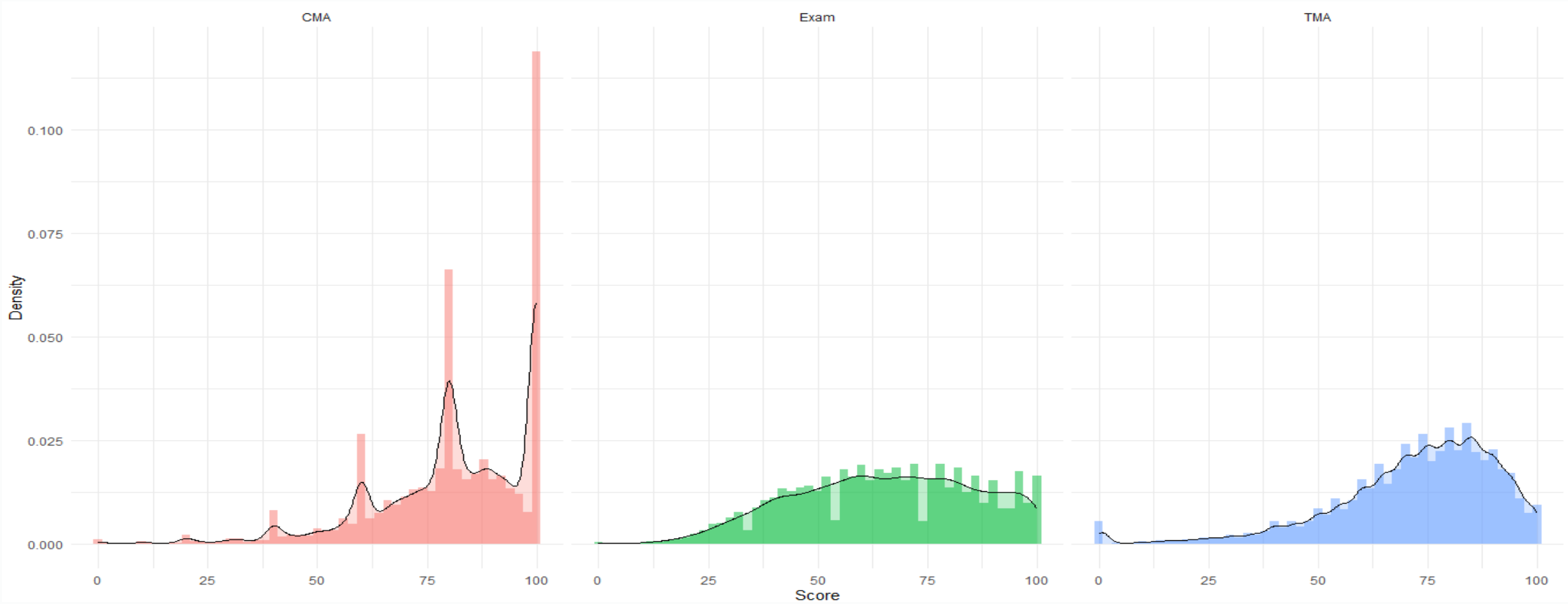


Modules BBB and GGG (Code: 2013B-BBB, 2013J-BBB, 2013J-GGG, 2014B-BBB, 2014B-GGG, 2014J-GGG) in both 2013 and 2014 have similar results, with a high volume of students achieving scores between 75 and 100.
In other modules, the scores are also concentrated from 50 upwards, but they gradually decrease as they approach 100

IV. Score and Sum of clicks - Virtual Learning Environment (VLE)

Score

Distribution of Score and type of assessments

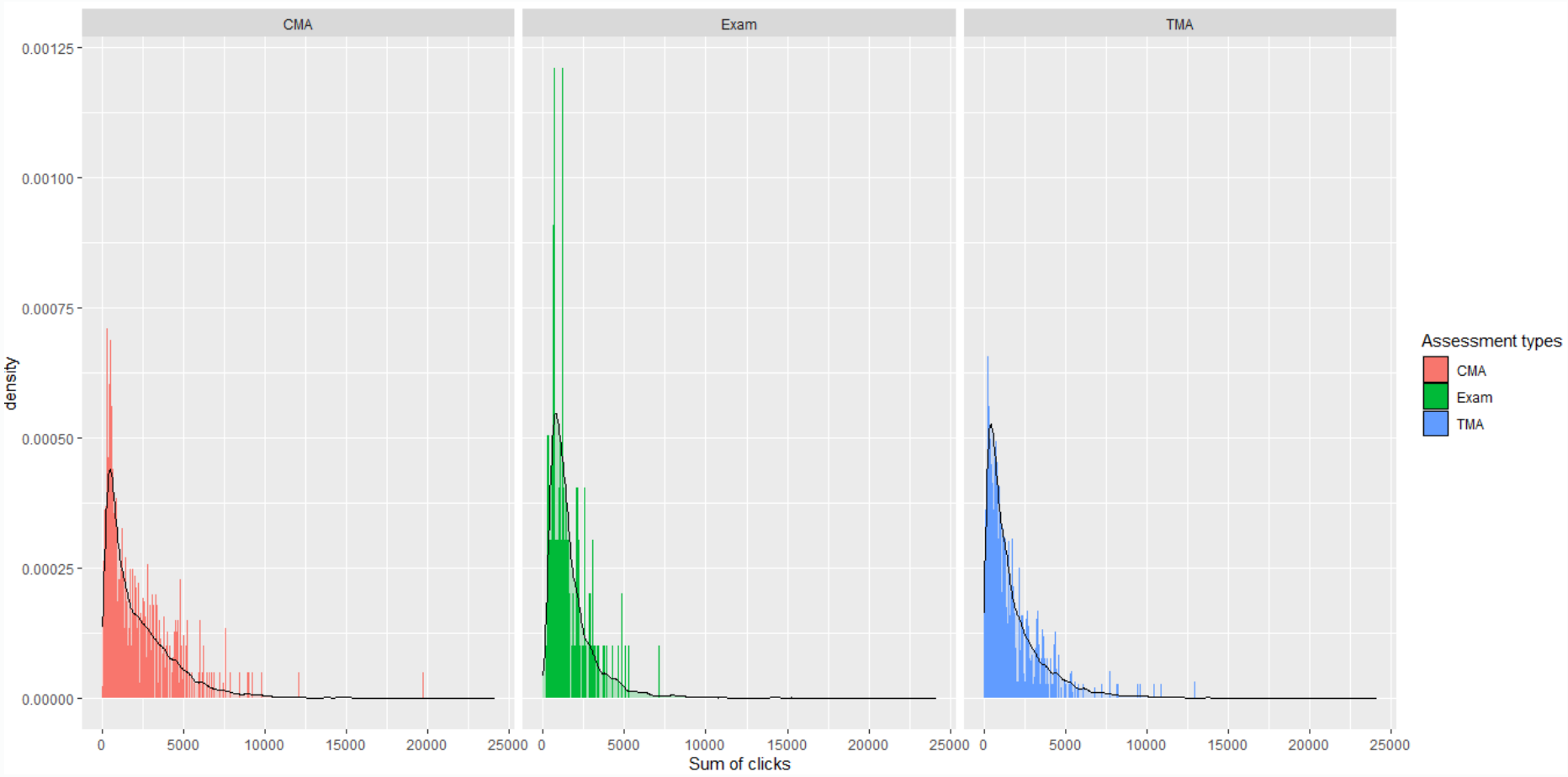


Students tend to achieve higher scores in assessments of type CMA compared to Exam and TMA assessments

IV. Score and Sum of clicks - Virtual Learning Environment (VLE)

Sum of clicks - Virtual Learning Environment (VLE)

Sum of clicks and type of assessments

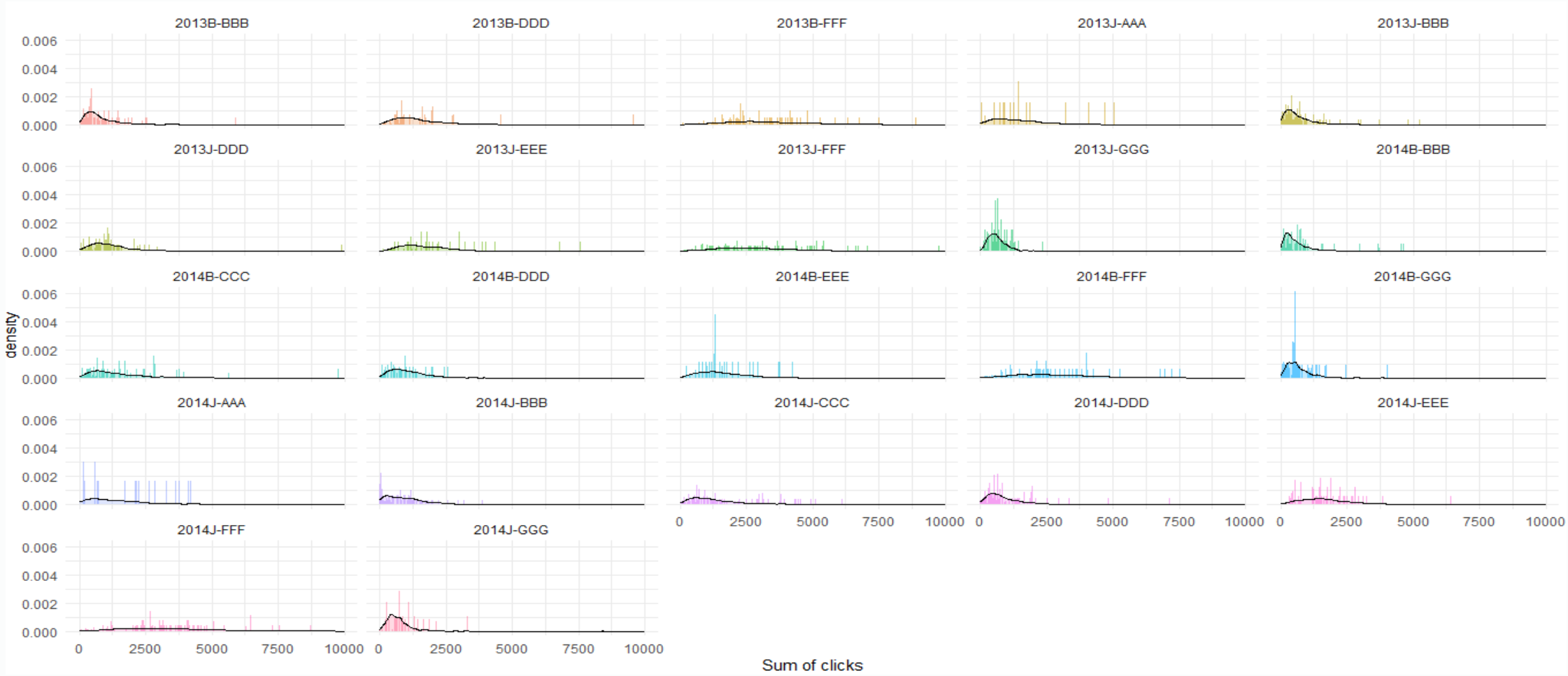


Students tended to spend more time on the Virtual Learning Environment (VLE) accessing materials prepared for the Exam compared to materials for CMA and TMA assessments, as indicated by the higher sum of clicks for the Exam assessment type.

IV. Score and Sum of clicks - Virtual Learning Environment (VLE)

Sum of clicks - Virtual Learning Environment (VLE)

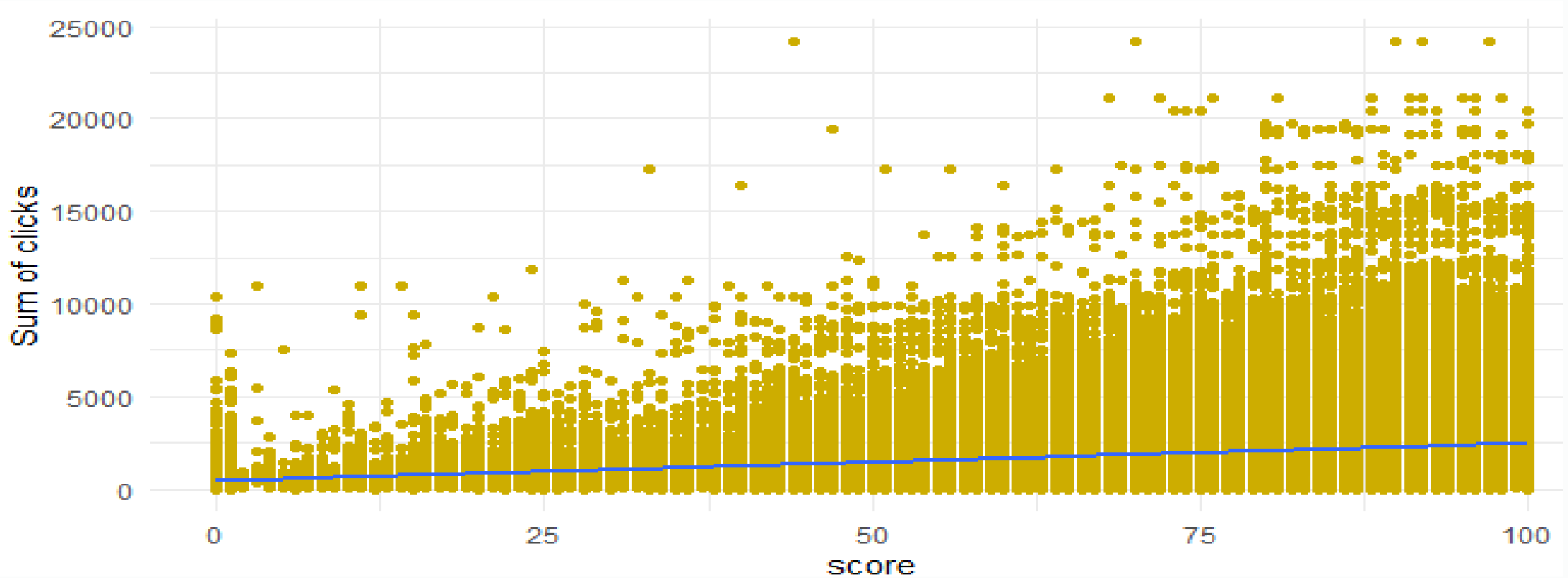
Sum of clicks and modules



IV. Score and Sum of clicks - Virtual Learning Environment (VLE)

Sum of clicks - Virtual Learning Environment (VLE)

Sum of clicks and score



Students who spent more time on the Virtual Learning Environment (VLE) with higher sum of clicks appeared to achieve higher scores.

Thank you