



# LEARNING ANALYTICS

Exploratory Data Analysis on the cyber security course  
from FutureLearn





Initially, I considered each run to be a separate dataset, with the goal of comparing multiple runs of the course and producing conclusions based on the comparison. However, I realized that if we examine such analysis, it will not provide me with a clear and complete image of the study and will prevent me from performing complicated analysis with the dataset.

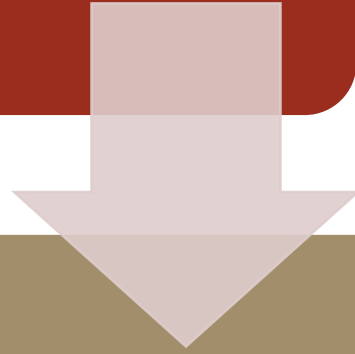


Then, from enrolment to leaving response, I wanted to examine every aspect of the dataset. But, once again, analyzing each detail of the dataset will be awkward and disorganized.

## MOTIVE OF THE ANALYSIS



I wanted to do a more in-depth investigation rather than a broad one. So, to undertake data analysis, I finalized three sets: enrolment, question response, and leaving response. I can find learner behavior from the moment they registered to the time they left using these data sets.



The language I have chosen to work on the project is R which will give me multiple libraries to work on the statistical side of the data and I have chosen RStudio as a tool to work with R language.

## CLEAR STATEMENT OF THE DATA



# CLEAR DESCRIPTION OF THE ANALYSIS WORK

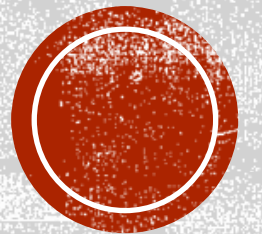
- I began my investigation with the enrolment data set, and as a first step, I consolidated all the data sets runs into a single frame.
- I compared the enrolled with their age, gender, highest level of education, and employment status using the entire data set. I compare the number of people enrolled in each run of the course with different runs of enrolment.
- Then I worked on question response using the combined data frame's number of correct and incorrect responses, and then with the number of correct and incorrect responses for each quiz question in the course.
- Then, using the leaving response, I can group learners who departed based on their reason for leaving and the number of weeks they spent in the course. These methods can be supplemented with even more thorough cross-examination of enrolment, question response, and leaving response.

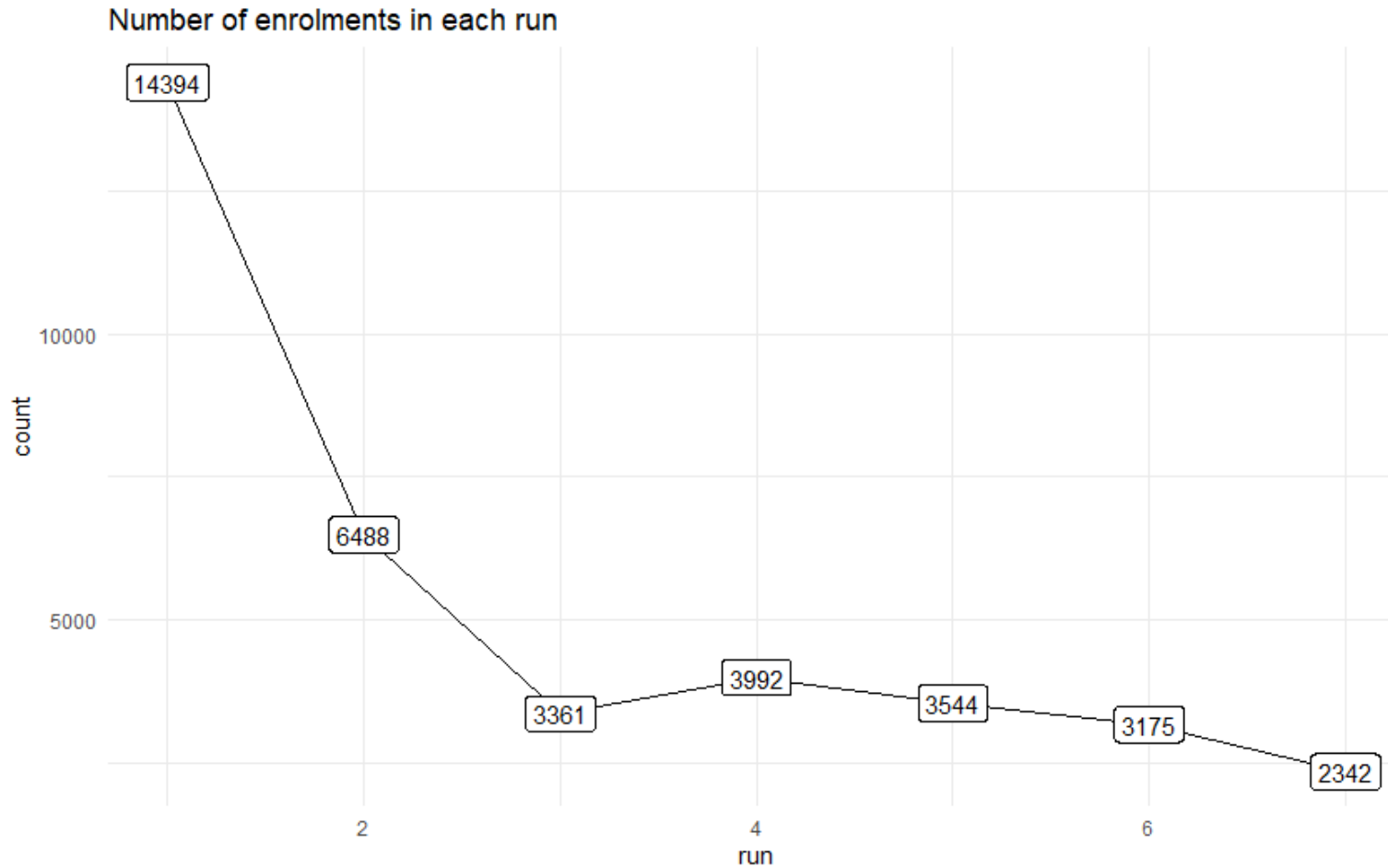
# CLEAR DESCRIPTION OF THE ANALYSIS WORK

- Enrolled persons of various ages, greatest education levels, and work position approaches quiz and why they left in between using a combination of enrolment, question response, and leaving response. It was a little difficult to locate and plot answers to questions like "Will they take the quiz and then leave the course, or will they leave the course without taking the quiz?"
- I had more questions in mind that required a lot more complex analysis, but the background information on the persons that were enrolled was incomplete, and there were numerous unknowns that disrupted the analysis flow.



# CLEAR DESCRIPTION OF KEY FINDINGS

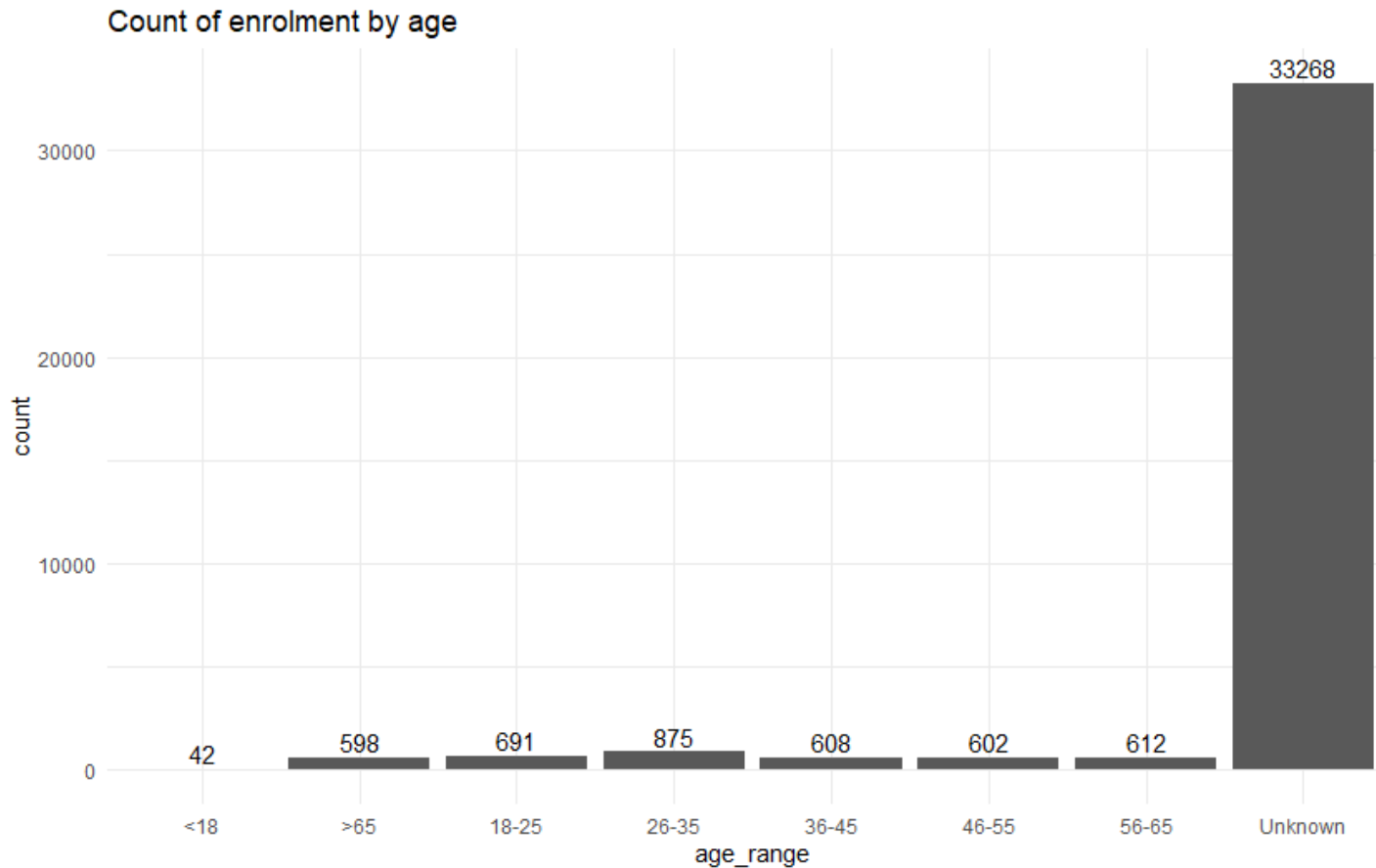




## COMPARING NUMBER OF PEOPLE ENROLLED IN EACH RUN OF THE COURSE

- From the plot, we can say that the enrolment numbers decreasing with each run. The number of enrolled in final run is 1/7th of the first one.
- This shows that the people lost interest in the course over the period. This is very alarming as the team at Future learn should look at it carefully and increase the number of courses relevant to the current market.





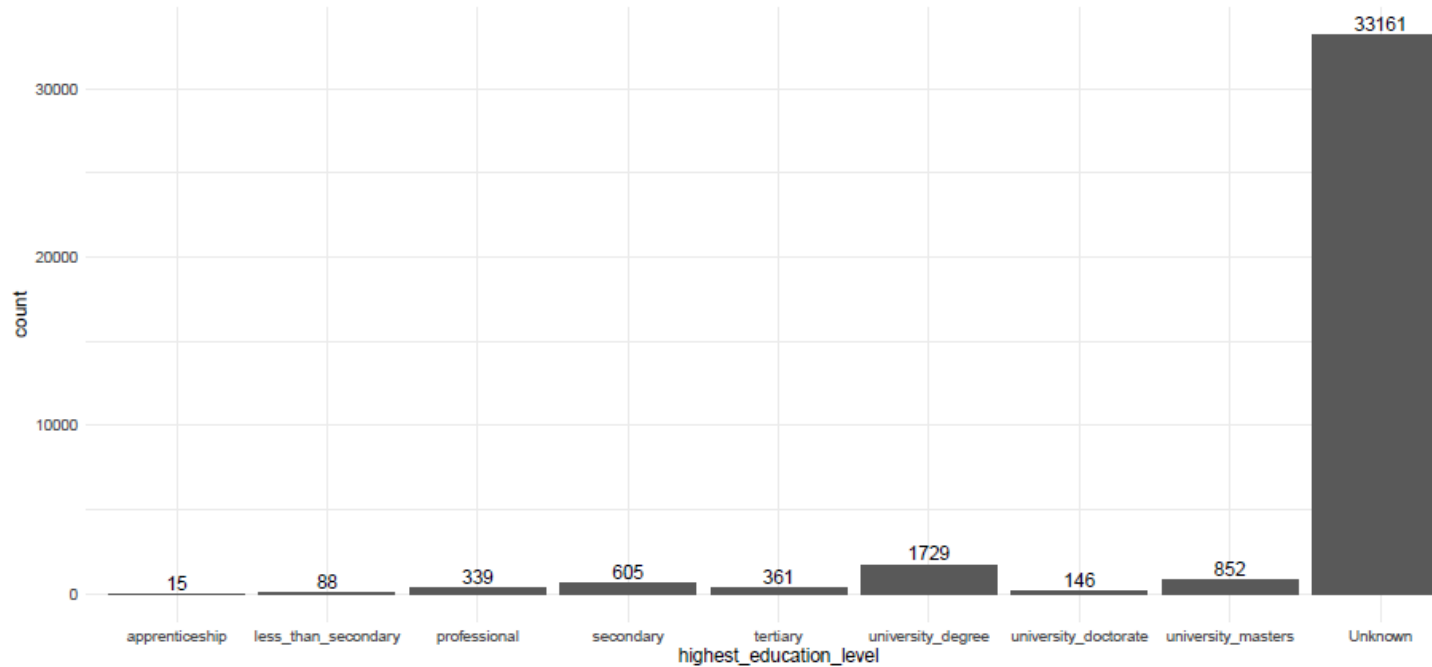
## COMPARING LEARNERS FROM DIFFERENT AGE GROUP

- From the plot, we cannot see any discrete differences in the block of age range across ages. Leave “Unknown” group aside.
- If they set the personal details entry as mandatory, we can fetch even more clear picture about the age range.
- They should create even more courses or tweak cyber security course according to the learners who represent majority of the age range.





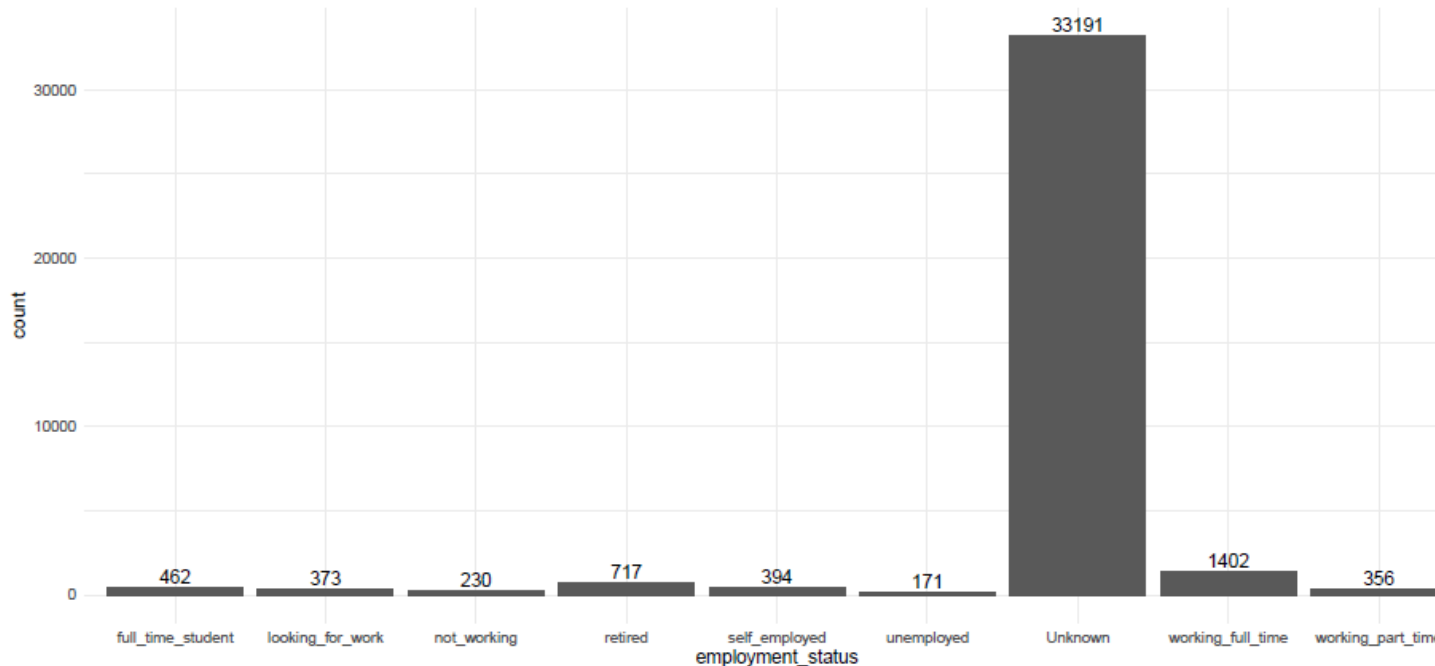
# COMPARING LEARNERS WITH THEIR HIGHEST EDUCATION LEVEL



- From the plot, we cannot see any obvious differences between the gender - male and female.
- It is good since the team don't need any promotional strategy to attract any set of people. On the other hand, non-binary and other gender people had enrolled in lesser number compared to other genders.
- The team need to work on promotional strategy to cover non-binary people since we as an organizer cannot neglect people of any population.
- But again 'Unknown' played major role in disturbed our analysis to some extent.



# COMPARING LEARNERS FROM THEIR EMPLOYMENT STATUS



- From the plot, we can see that it is obvious that the highest numbers of learners who are in their full-time job. But it is very strange that the second highest number of learners are retired people.
- As usual, the students had enrolled in the course in high number. But they need to concentrate on people who are not working or looking for job since the number is very low.
- We can explore that market considering the people are unemployed usually needs something to study or learn new to enhance their careers.
- The number of self-employed is comparatively less and we can pull people who are self-employed by including the course dedicated to them.

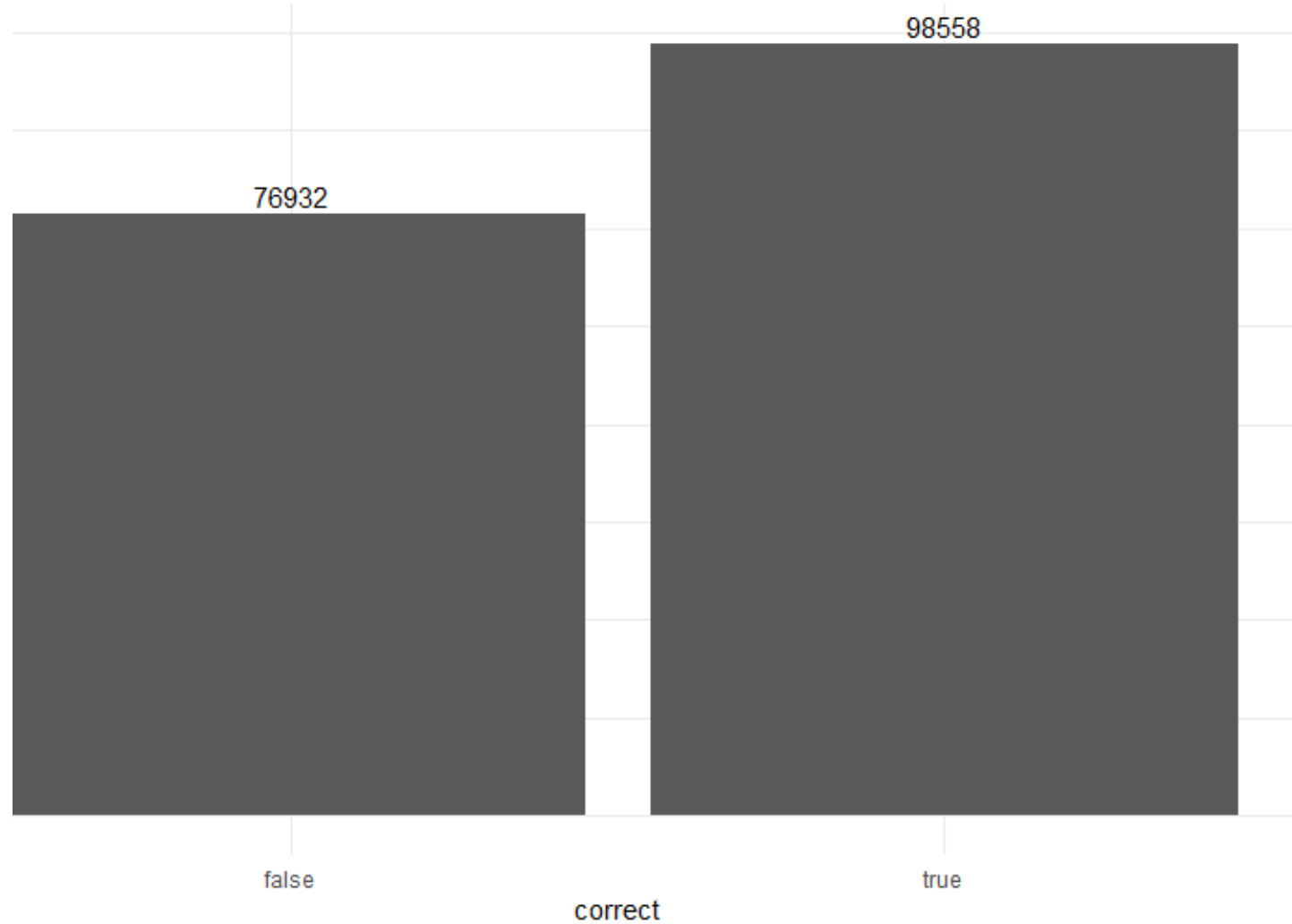


**NOW, WE ARE GOING LITTLE  
DEEP INTO THE ANALYSIS BY  
CONNECTING ENROLMENT DATA  
SET WITH QUESTION RESPONSE.  
HERE, WE CAN CALCULATE  
DIFFERENT ASPECTS OF THE  
CONNECTION BETWEEN THE  
LEARNERS AND HOW THEY  
RESPOND TO THE QUESTION  
FROM THE COURSE.**





s by responses



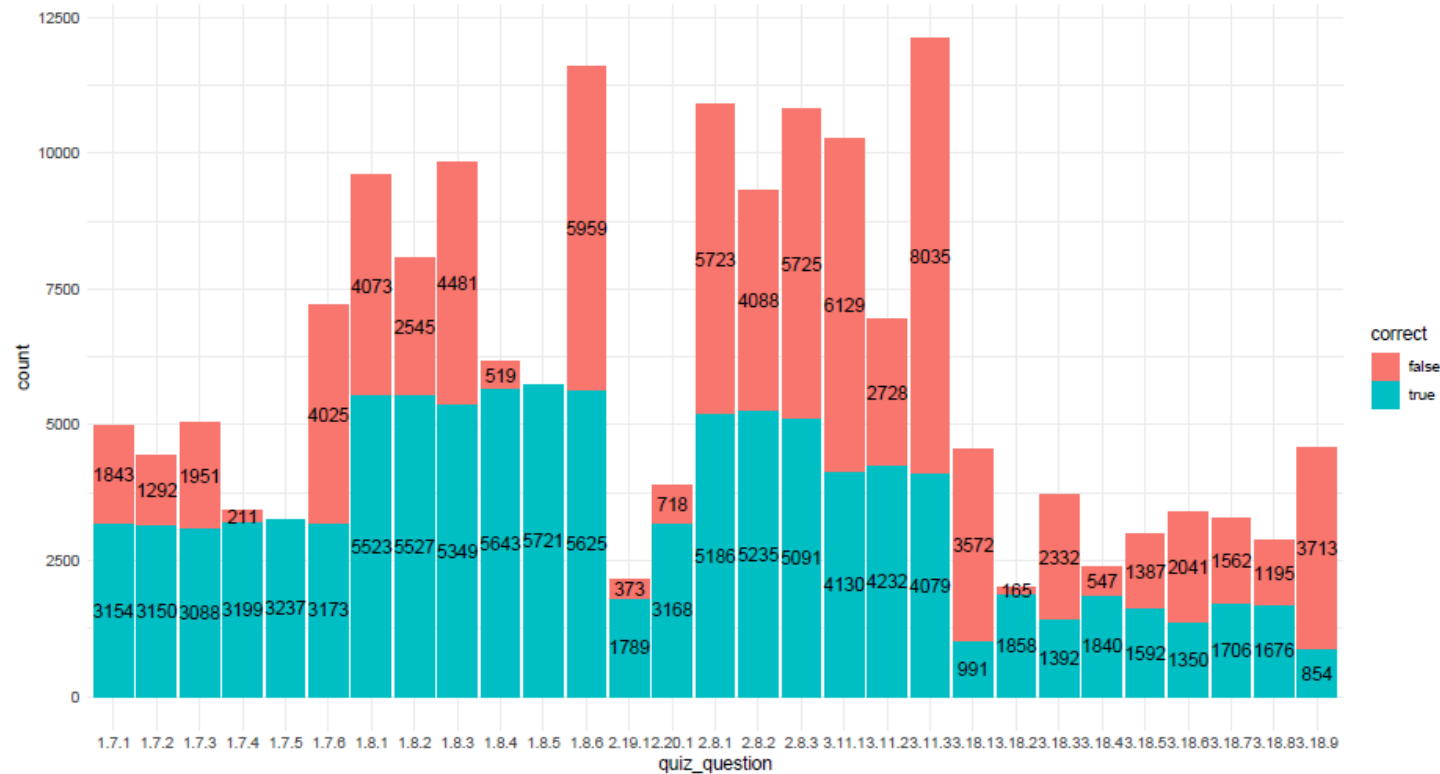
## COMPARING NUMBER OF QUIZ RESPONSES WITH THE RESPONSE

- From the plot, it is positive to some extent as the number of correct responses is higher than the number of wrong responses.
- But the negative point here is the difference between them. since the difference between them is non negligible, the team must take this very seriously.
- The number of wrong responses shows that many learners do not understand the course well enough to give the right answer. Since the quizzes were multiple choice questions, we cannot even that as a credit.
- There is a chance that the right answer could be a luck. It is understandable that the correct and wrong responses will not affect the business at Future learn in day-to-day basis.
- But if you look at the larger picture, it is not good in terms of knowledge sharing.





# COMPARING NUMBER OF QUIZ RESPONSES FOR EACH QUESTION

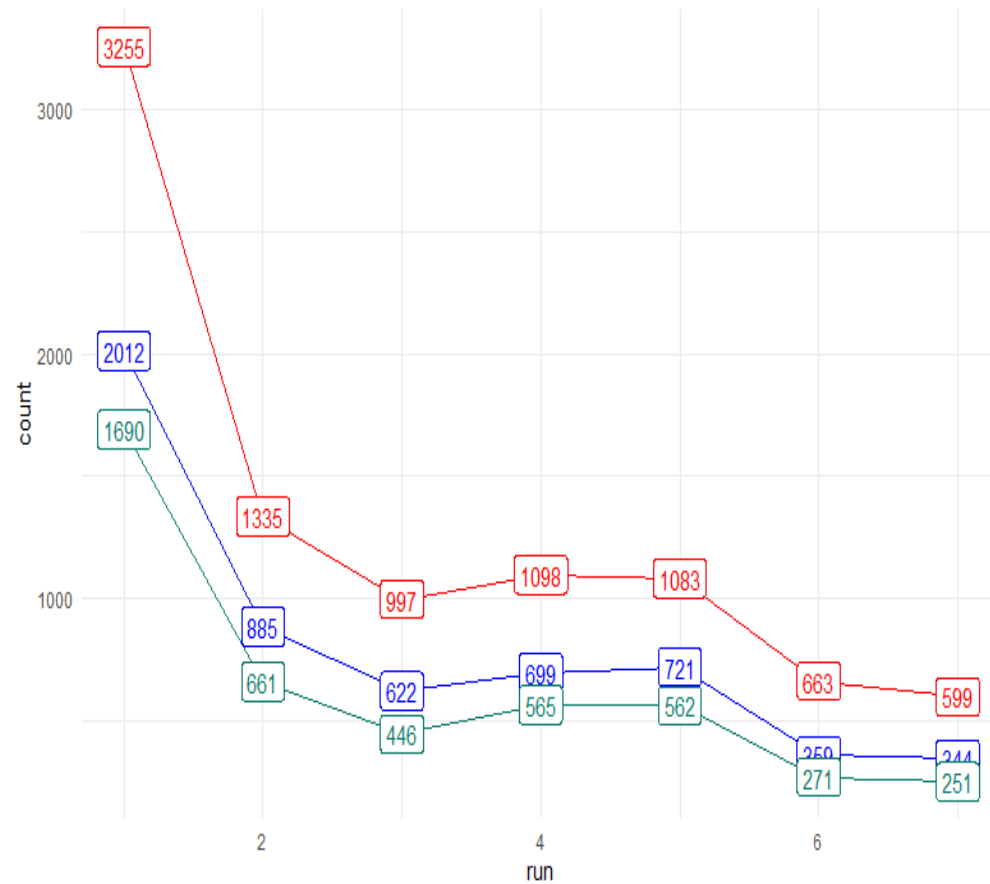


- From the plot, we will compare the same correct and wrong responses but with each question.
- We can see that number of wrong responses is higher in the later part of the course. It is very evident that the learners are losing interest in the course as they travel with the course.
- The team should improve the course structure in the “Security in the future home” part of the module.

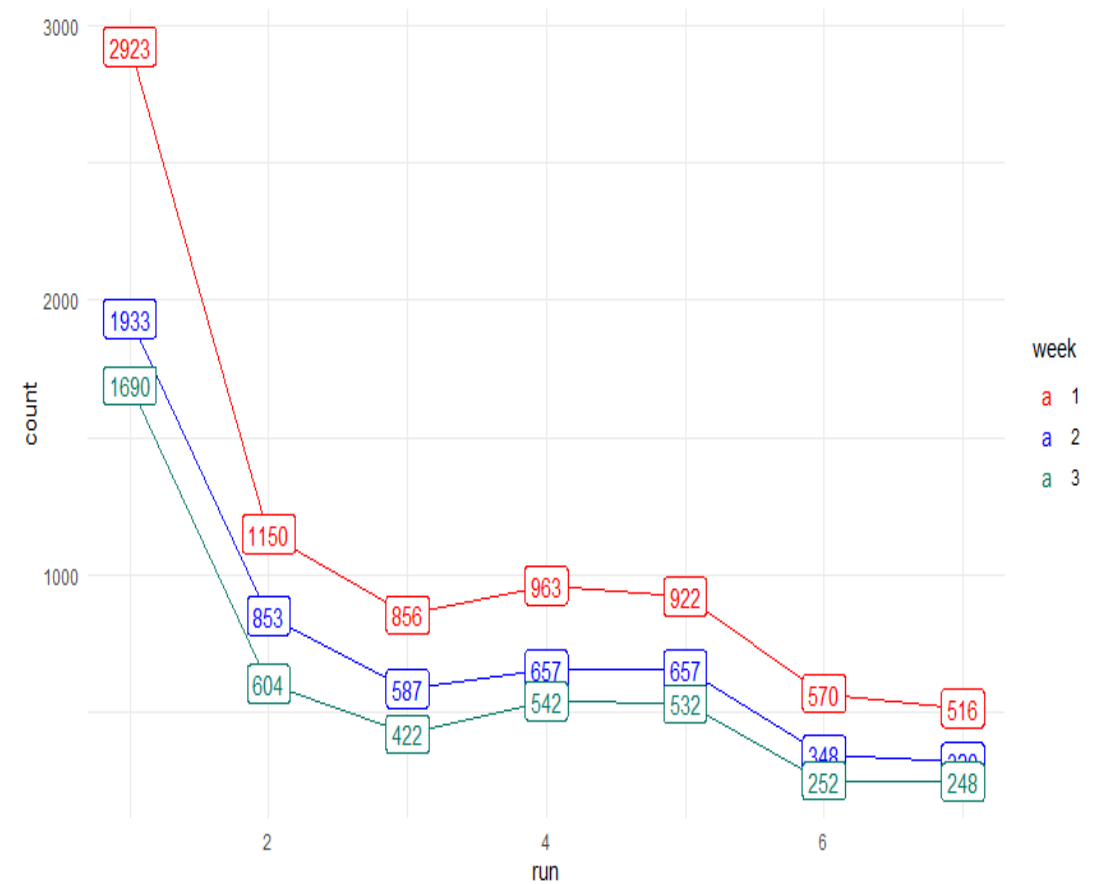


# COMPARING NUMBER OF CORRECT AND WRONG RESPONSES IN EACH RUN OF THE COURSE

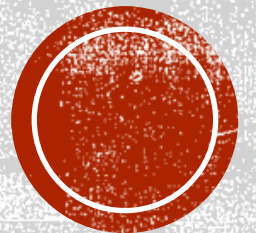
Number of correct responses to quiz by week in each run



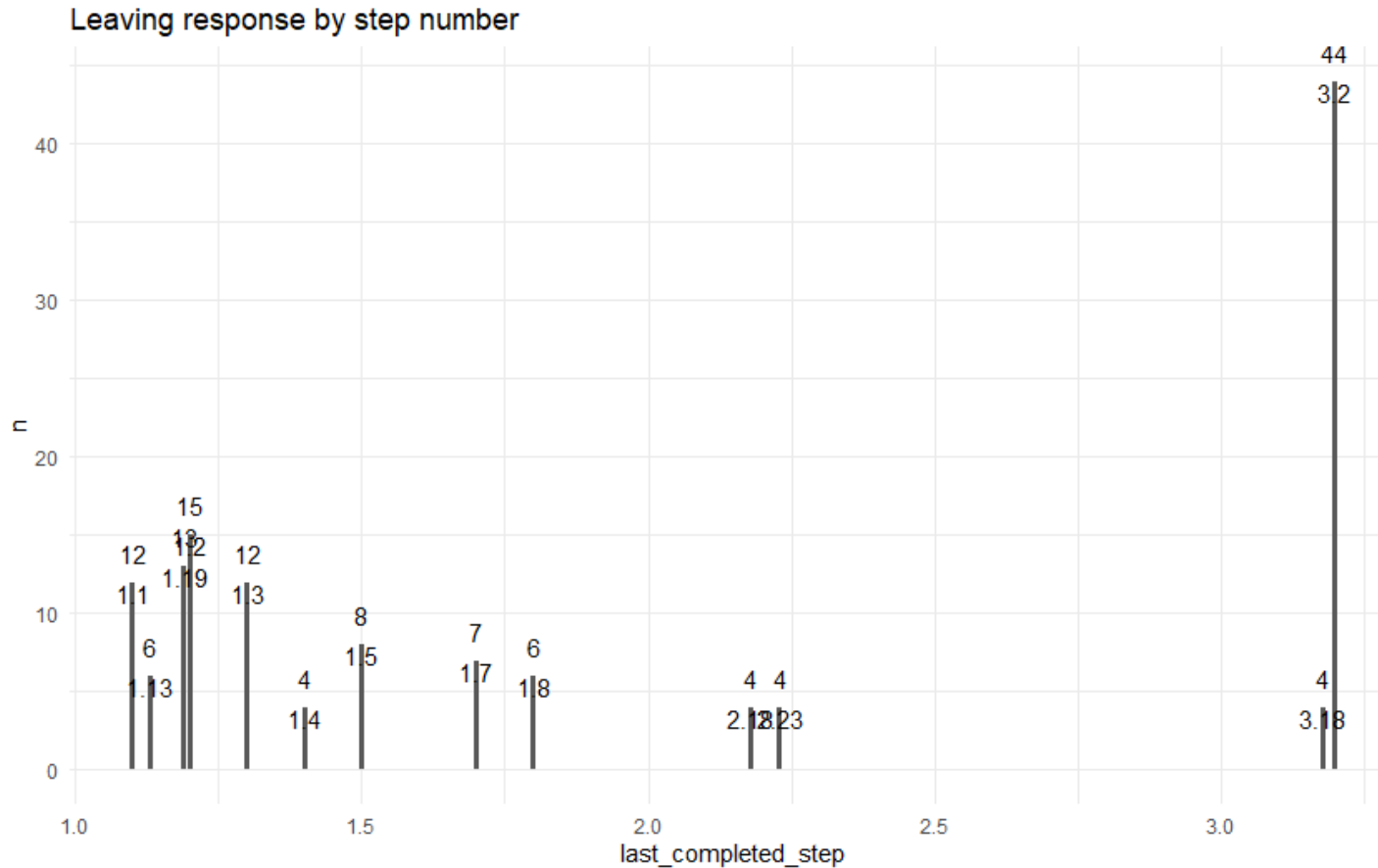
Number of wrong responses to quiz by week in each run



**NOW, WE ARE GOING TO CONNECT ENROLMENTS  
AND QUESTION RESPONSES WITH LEAVING  
RESPONSES. IT WILL BROADEN OUR  
ANALYSIS TO CERTAIN EXTENT.**



## COMPARING NUMBER OF PEOPLE LEFT THE COURSE IN EACH STEP OF THE COURSE - TOP 10

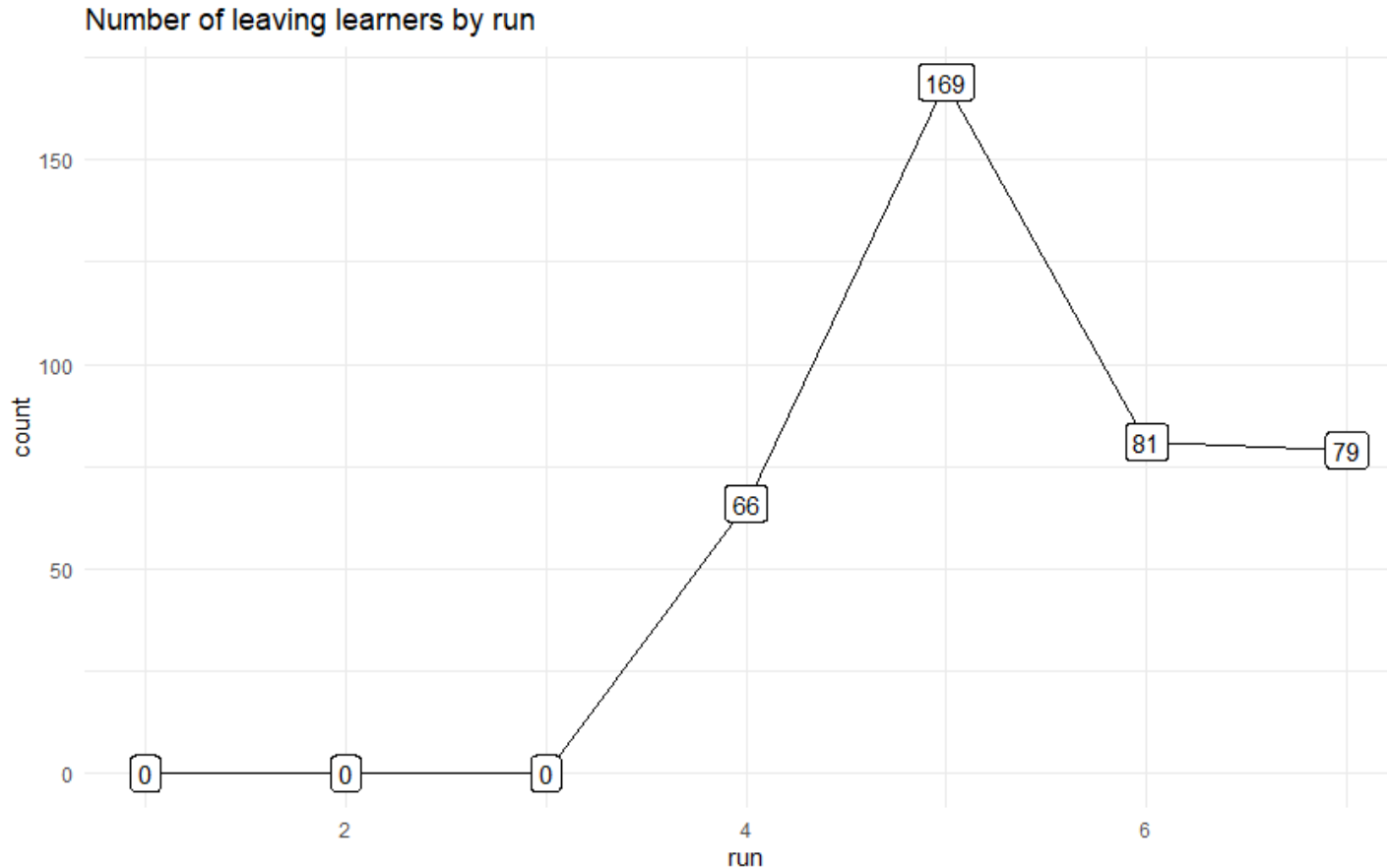


- From the plot, we can say that most of the learners who left from the course were left at the step 3.2.
- Around 44 people left the course at that step. It must not be a coincidence. When we look at the step-in depth, it has a video explains “Devices in the future home”.
- There is a chance that the video might be boring, or the course might hit a low point in terms of interaction and engagement at that point.
- The team at Future learn should take this seriously and work on this step a little more compared to another step.



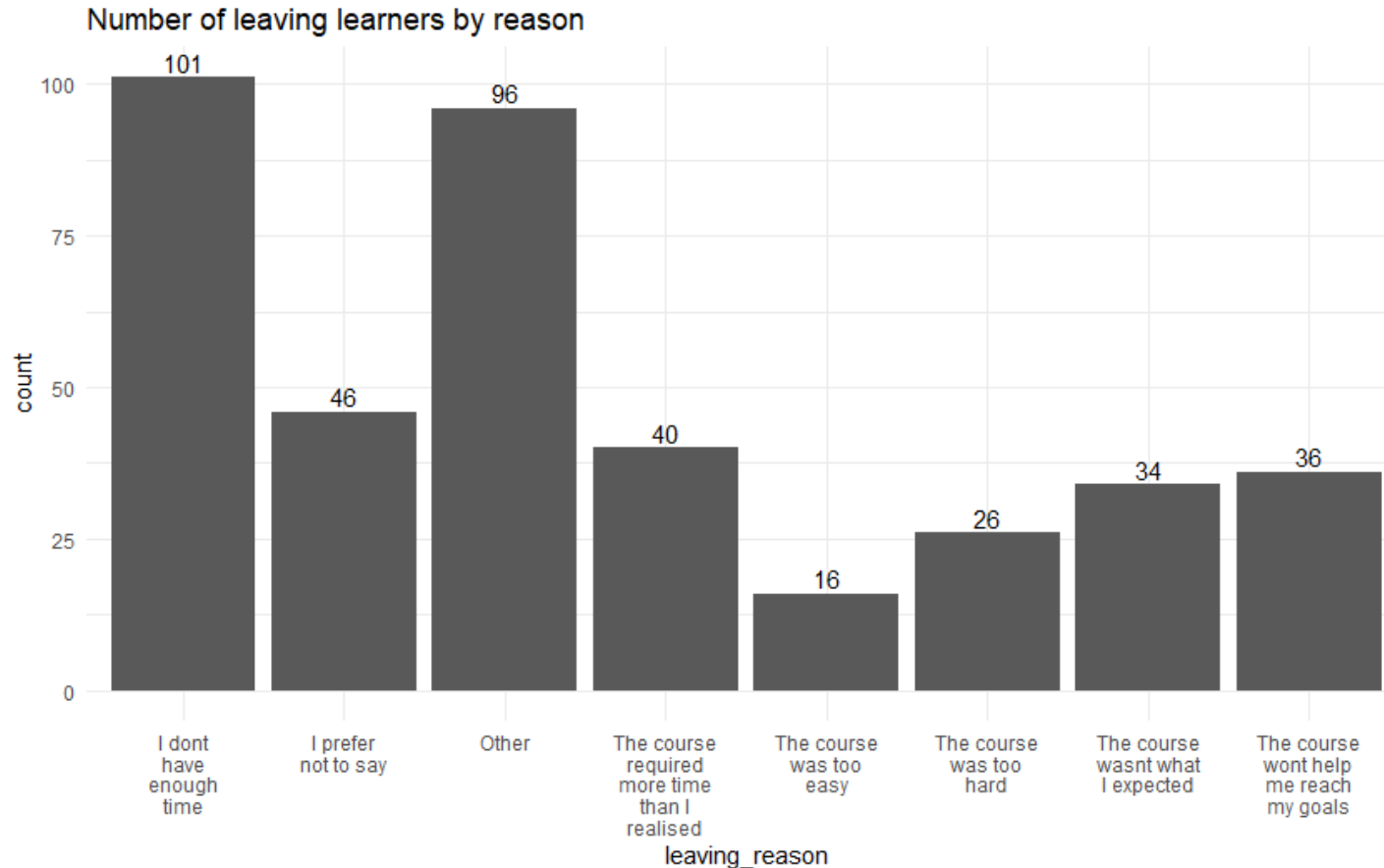


## COMPARING NUMBER OF PEOPLE LEFT THE COURSE IN EACH RUN



- The plot explains about the number of learners left the course in each run. Since we do not have any data about the people left from first run to third run.
- We will consider the data only from fourth run to seventh run. The number of people left in fifth run of the course is higher than other runs even though the enrolled are comparatively lesser than the fourth run.
- The difference between people left in fourth and fifth run is more than 100. We can confirm that the fifth run had not performed well and produced good results compared to other runs.

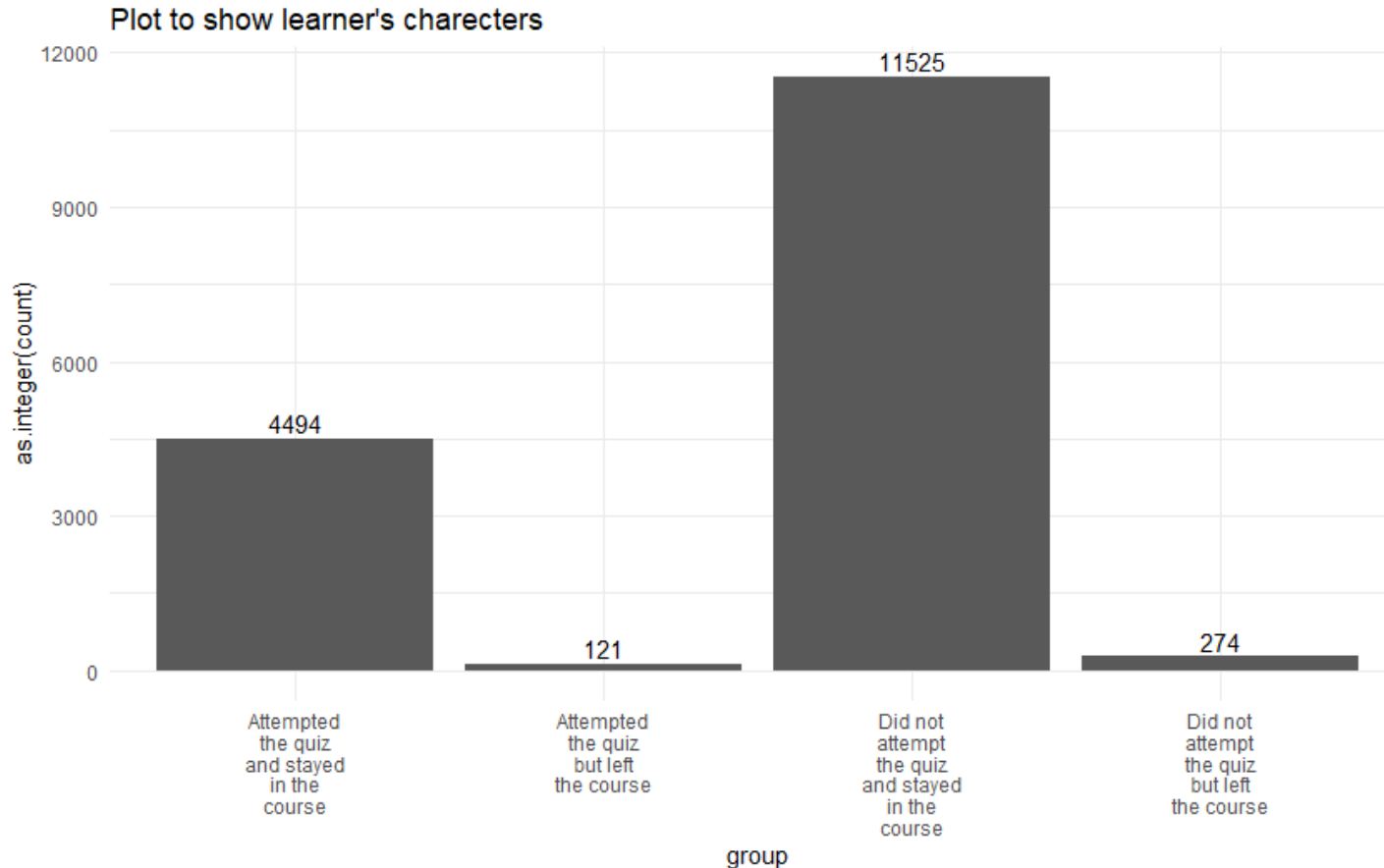




## COMPARING NUMBER OF PEOPLE LEFT THE COURSE WITH THEIR REASON FOR LEAVING

- From the plot, we can identify the number of learners left the course with the reason of leaving they mentioned.
- As we can see, the highest number of learners left the course because they do not have enough time. The team need to make the course flexible according to the time the learner has.
- The fast-track batches are one such recommendation they can consider. It is closely followed by the reason 'Other'. Each learner would have different answers to choose the 'Other'. We cannot neglect that, but we do not have enough data to analyze it.
- The reasons 'I prefer not to say' and 'The course required more time than I realized' was selected by 46 and 40 learners respectively. This denotes that many people consider the course as too long to complete.





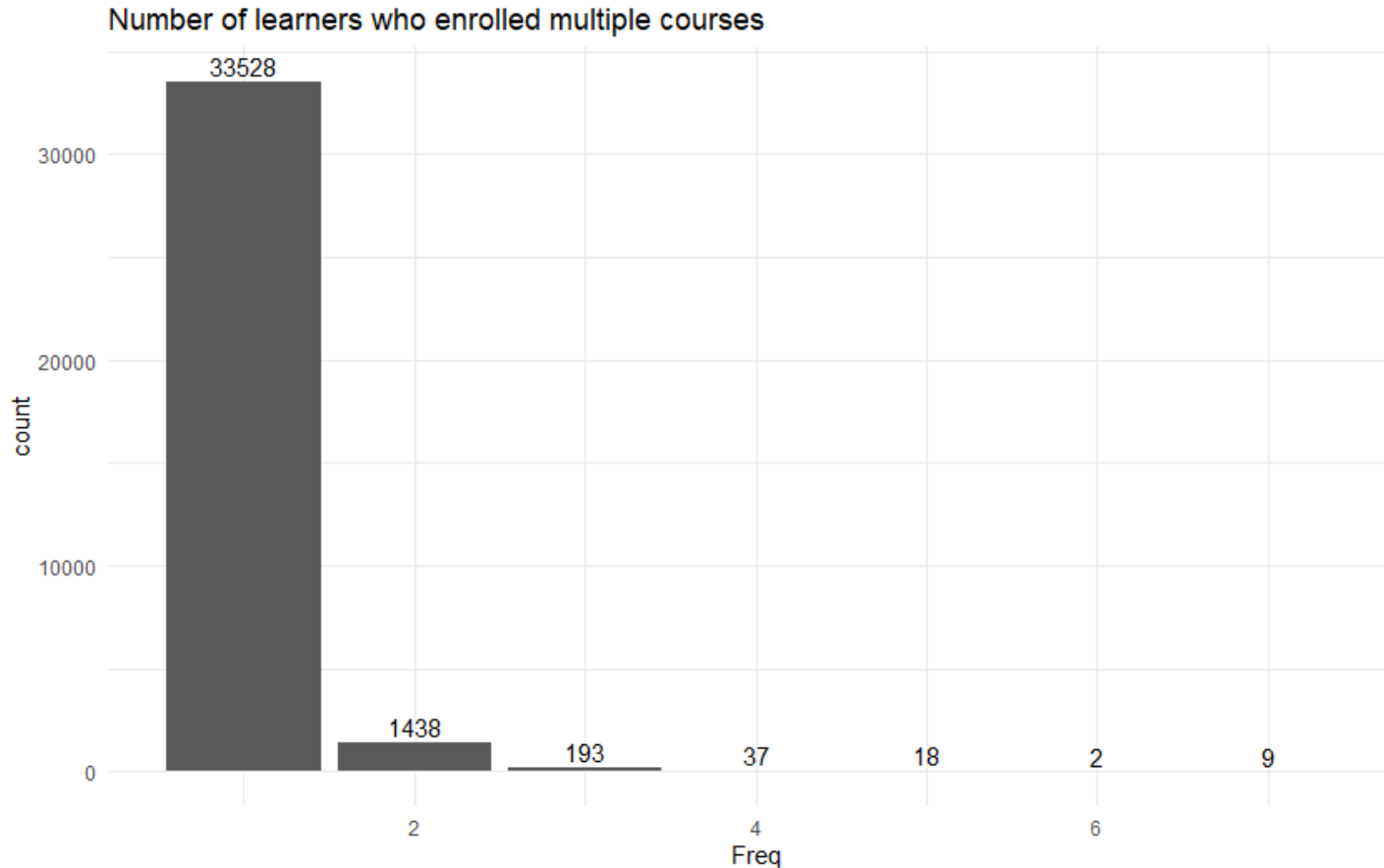
## UNDERSTANDING LEARNER'S BEHAVIOUR

- Since the highest number of people who did not attempt the quiz and stayed in the course, we can say that there are many learners who are neither actively attempting the quiz nor they left the course.
- When we compare the people who left the course, we can say that the numbers are high in people who did not attempt the quiz but left the course.
- This is obvious that they do not understand the course or do not have interest in the course which probably be the reason behind the highest number in this group.





## COMPARING THE NUMBER OF PEOPLE ENROLLED IN ALL RUNS OF THE COURSE

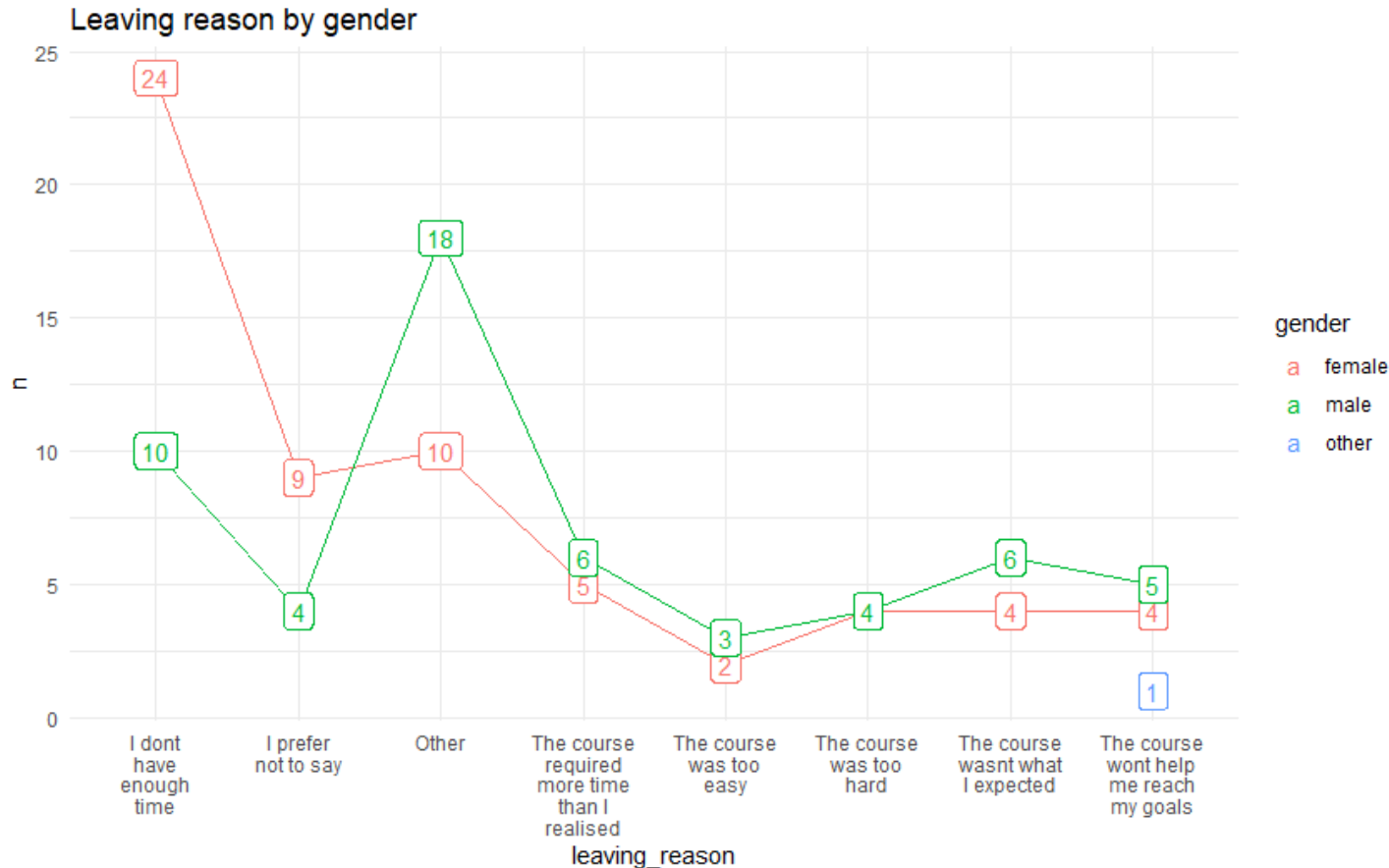


- From the plot, we can compare the number of people enrolled in all runs of the courses.
- As we can the number of people enrolled only in one run of the course is higher compared to other numbers.
- This is the expected one as this is the good sign since the number of learners who enrolled in one run of the course is more than 90% percent of total enrollment.



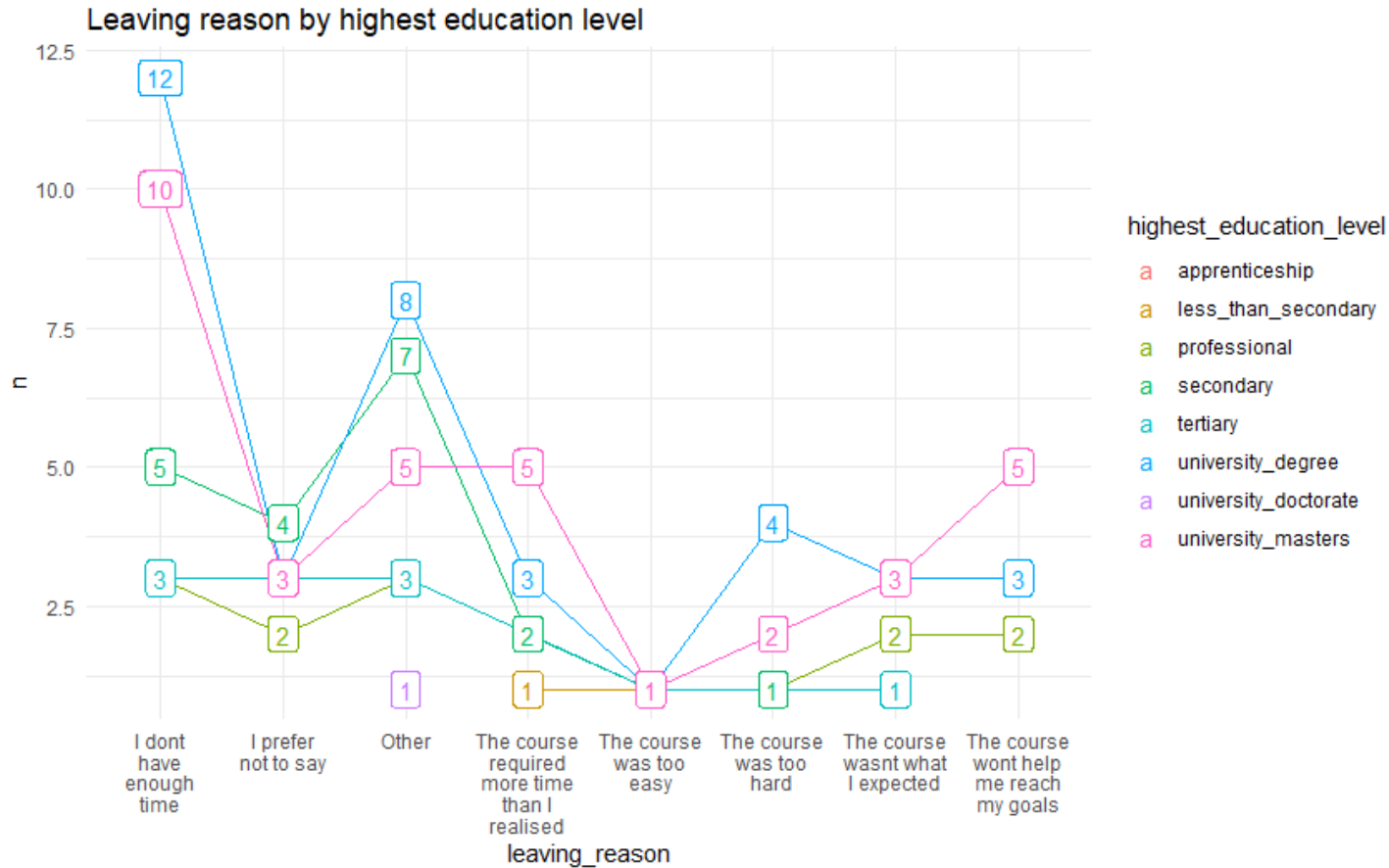


# CORRELATING LEARNER'S REASON FOR LEAVING WITH THEIR GENDER



- Gender comparison shows that the female left for the reason 'I do not have enough time' more than any other reason. But the reason 'Other' is more for male learners.
- They had different sub reasons for choosing 'Other'. The team should investigate this and can give a crisp version of full course according to their timetable.
- This will engage many female learners to actively enroll in their course with their preferred span.





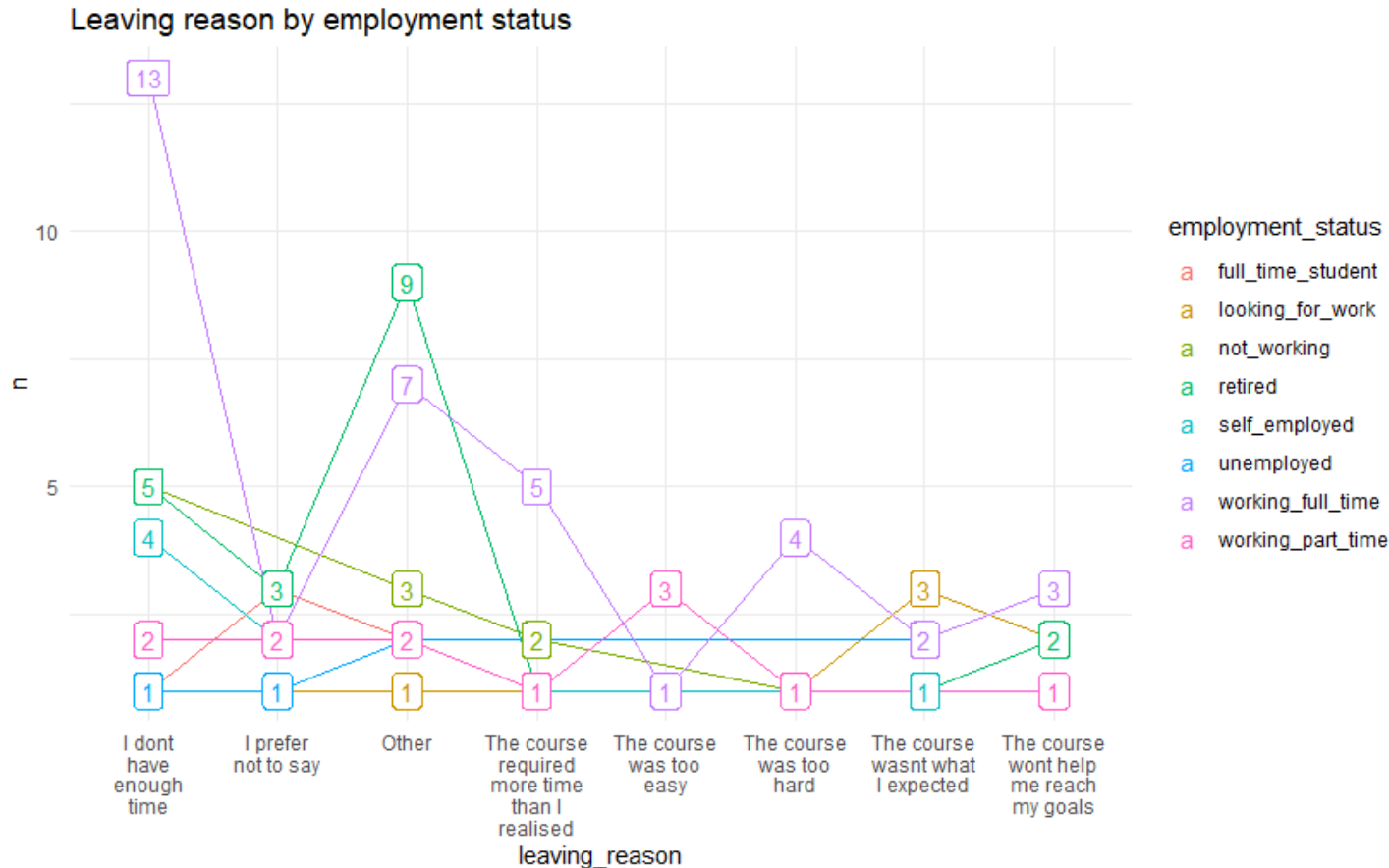
## CORRELATING LEARNER'S REASON FOR LEAVING WITH THEIR HIGHEST EDUCATION LEVEL

- The comparison between learners with their highest education level shows that the students at their university degrees did not have enough time to complete the course. It is acceptable since they already have their own curriculum to complete, and they are not considering this course as an extra knowledge gain.
- Like for female learners, it would be good if the team design a crisper version of course to fit within the timetable of learners who do not have enough time to complete the course.
- In addition to that, we can see around five people left the course with the masters education level with the reason that the course will not help them to reach their goals.
- It means that the course does not have a mass appeal to all possible learners.



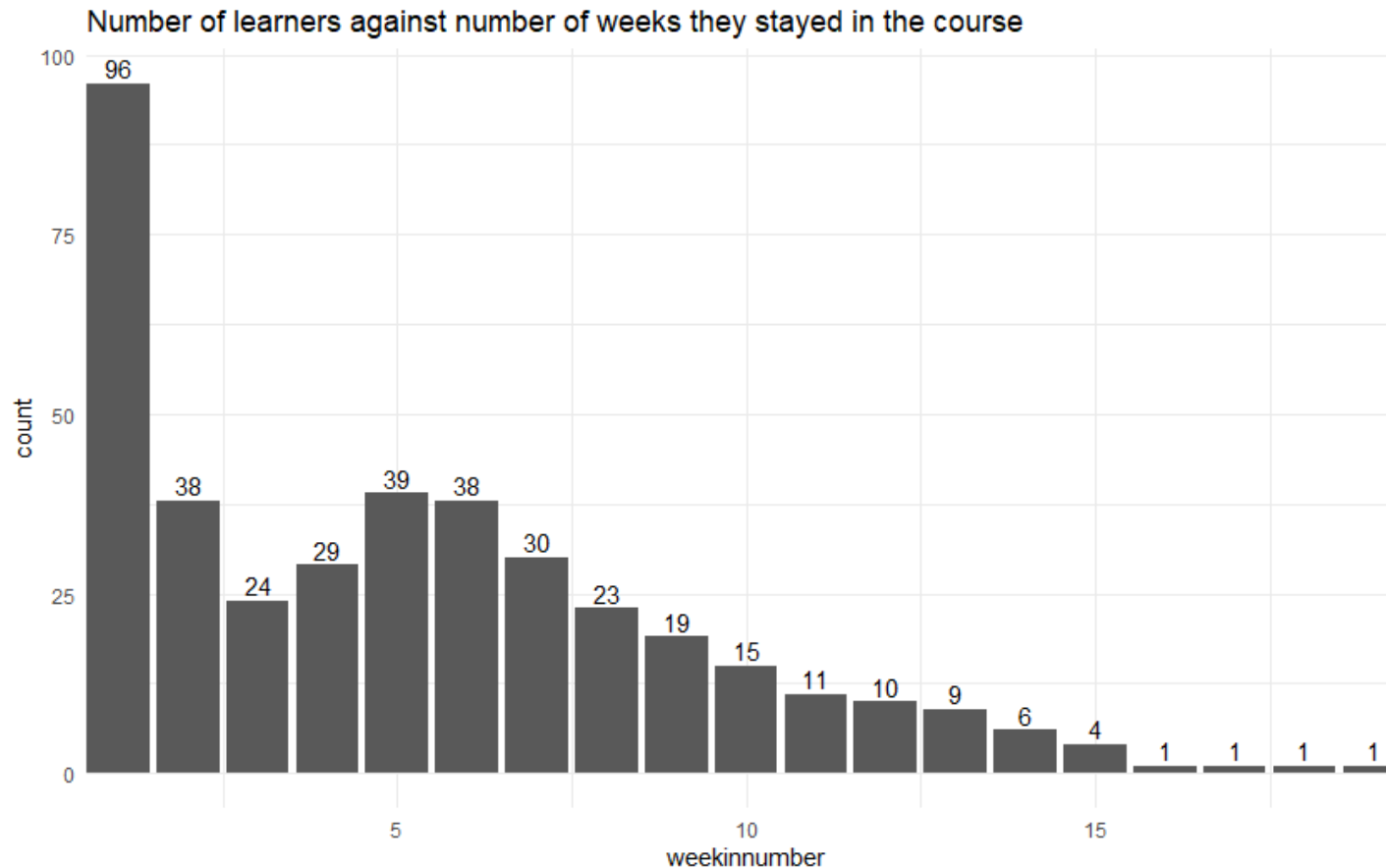


# CORRELATING LEARNER'S REASON FOR LEAVING WITH THEIR EMPLOYMENT STATUS



- The comparison between learners from different employment status shows that the people who are working full time do not have enough time to complete the course. For all these problems, the short and crisper version of course is the only solution.
- The team should really focus on the length of the course since it is the major concern among different sets of learners. The team should know that most of the learners will enroll in this course if they need any additional skill, if they want any promotion in their current workplace, if they want to spend their leisure time effectively by gaining some knowledge, if they need to crack any interview and so on.
- This will not be the primary focus in their daily routine. So, it would be ideal if the team concentrates on the duration of the course and include some advanced topics to cover users who already have basic knowledge about the topics.





## COMPARING NUMBER OF LEARNERS AGAINST NUMBER OF WEEKS THEY STAYED IN THE COURSE

- We will compare the learners who left the course with the number of weeks they stayed in the course. We can achieve the result with the difference of the time stamp they left at, and the time stamp they enrolled at. With the plot, we can see that the number of people left at the first week is higher than the rest of the plot.
- It suggests that the learners who enrolled the course and left the course within a week when they see the length of the course and thinks that they could not complete the course or fit in the course within their timetable.
- This might be the reason of hike in the first week of the course. The course is designed to complete within three weeks but if we look at the plot, we can see that majority of the learners who left the course tried to finish the course by looking at the number who left the course after three weeks. They stayed in the course with a hope to complete the course. Since it was not possible for them, they left the course after a try.





**AS A RESULT, I APPROACHED THE PROJECT  
USING THE CRISP-DM METHODOLOGY,  
DESIGNED WITH R USING RSTUDIO, AND  
DUG A LITTLE DEEPER INTO THE DATA  
SET WITH TWO DIMENSIONS, PRODUCING  
DIFFERENT PLOTS FOR STAKEHOLDERS TO  
UNDERSTAND THE ANALYSIS AND MAKING  
RECOMMENDATIONS TO IMPROVE THEIR  
COURSE AND INCREASE THE IMAGE OF  
THE BRAND THEY CREATED AROUND THE  
WORLD.**

