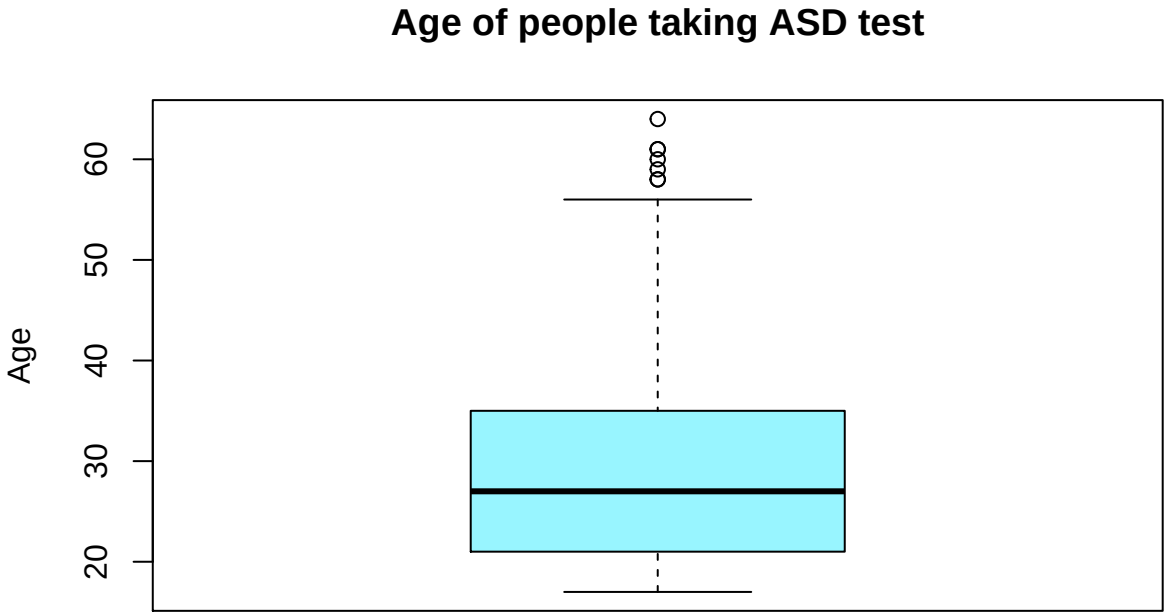
Results and discussion

Results

We will now take a look at the ages of the people taking the test are.

Figure 1A, As the boxplot shows, there's one huge outlier. One person would be 383 years old which just isn't humanly possible. The solution to this is taking out the whole row.



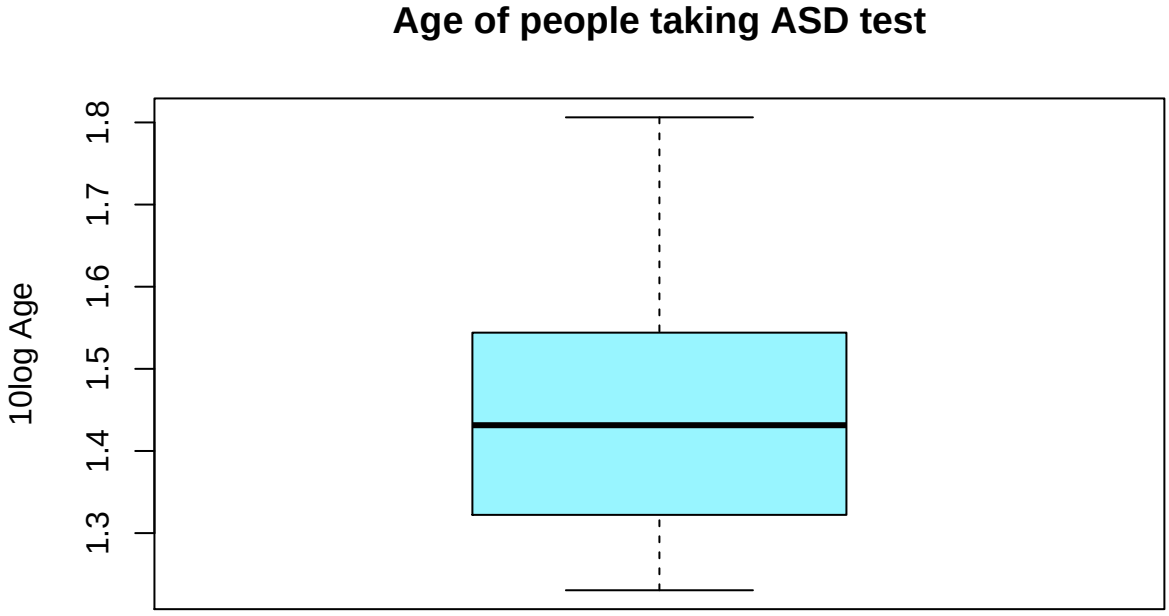


Figure 1B/C, This is the fixed plot by using the 10log function. It shows that the majority of the people taking the test is Mid age.

Correlations

This boxplot shows that there's not much old people (60+) doing the test. The people who take this test are usually mid aged.

Now we will take a look at the test, how much people had what kind of anwser. The goal of this plot is to take a look at what the people scored.

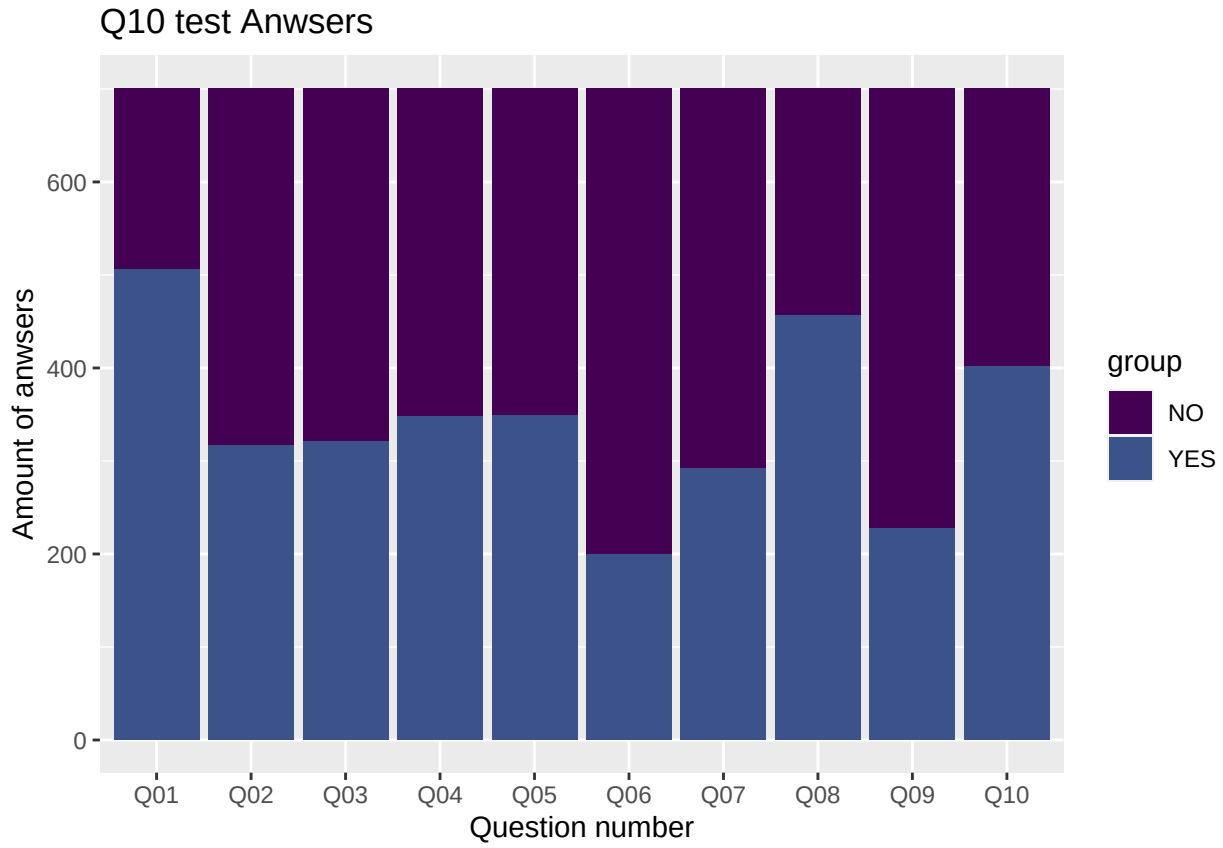


Figure 2, The conclusion of this plot is that most questions are anwsered positively (Without getting a point).

Let's take a look at the correlations now. To start off, the end score will be measured against people actually having ASD. This will give a good view of the test because the test results will directly be compared to them having ASD.

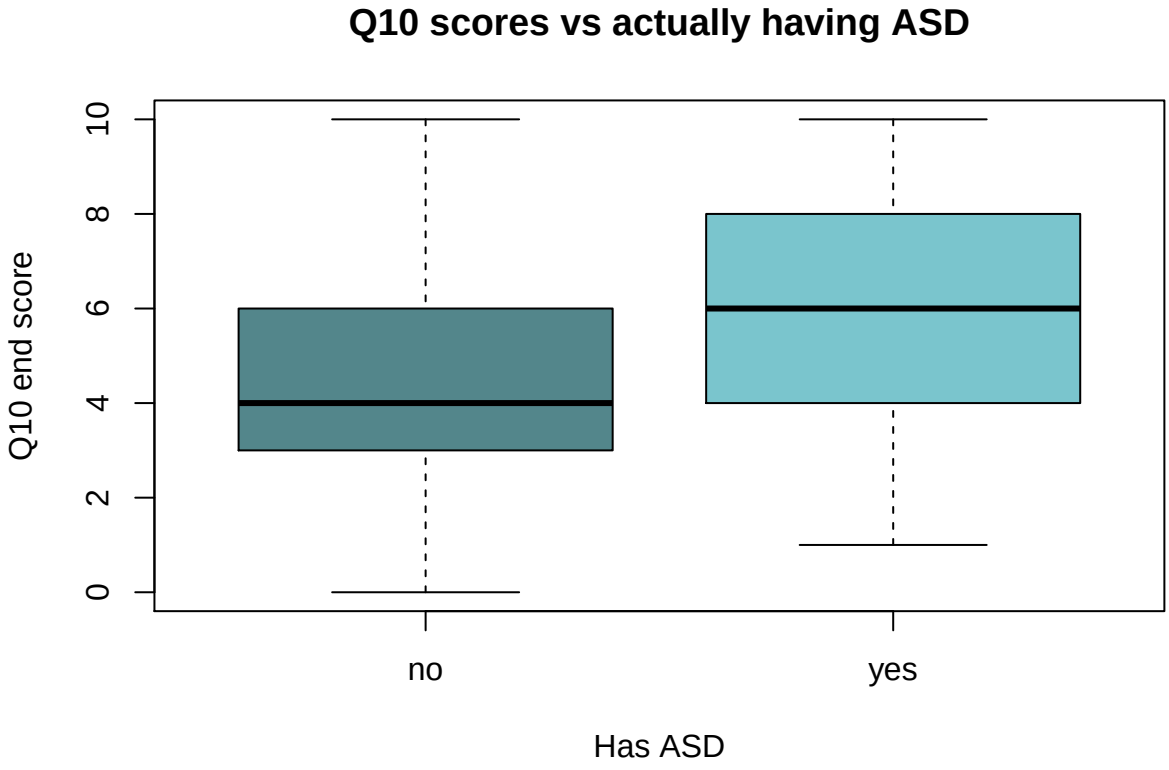


Figure 3, As seen in the plot, the scores do actually correlate with someone having ASD. This is true because the scores of the people having ASD are significantly higher than the other people. Let's confirm this by doing a t-test

```
##
##  One Sample t-test
##
## data:  data$end_score
## t = 51.867, df = 700, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  4.703675 5.073785
## sample estimates:
## mean of x
##   4.88873
```

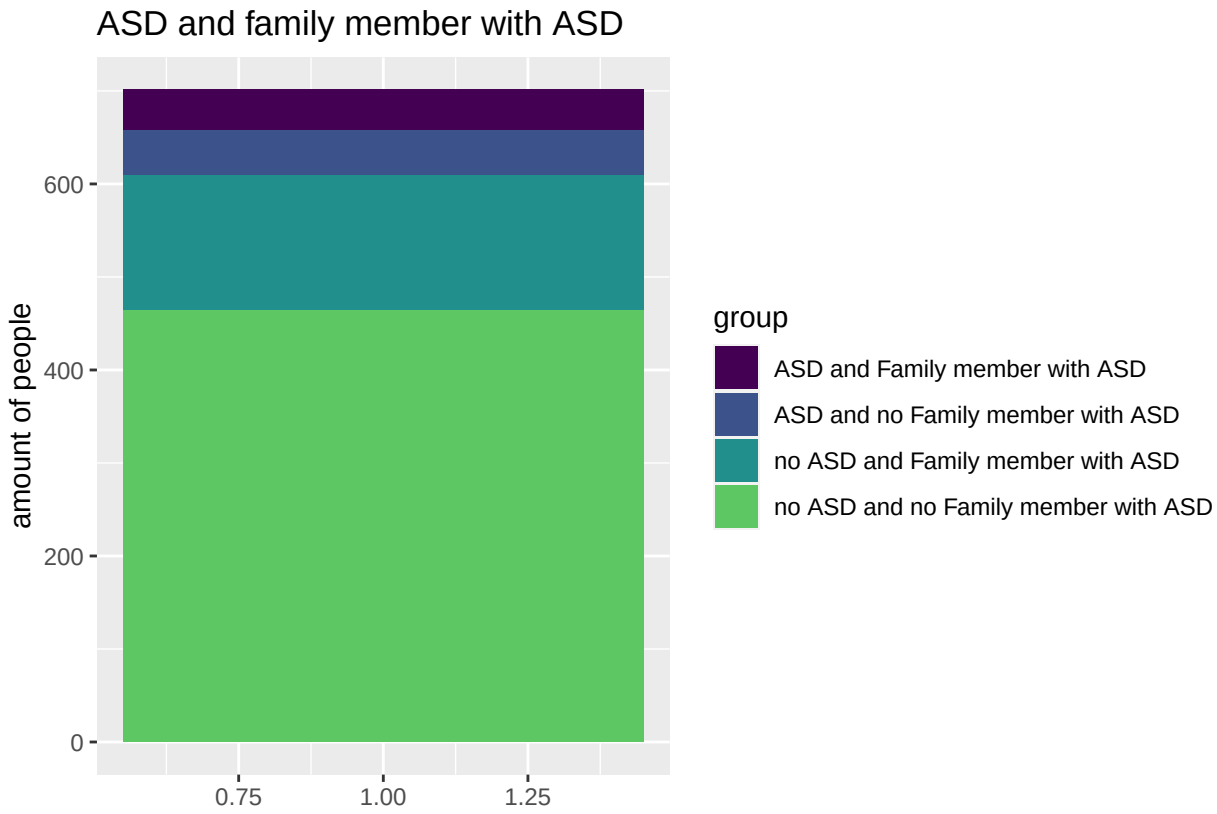


Figure 4, This figure shows that Most people that take the test do not have ASD and no family member with ASD. There's also a big part of the group that do not have ASD but have family members with ASD.

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

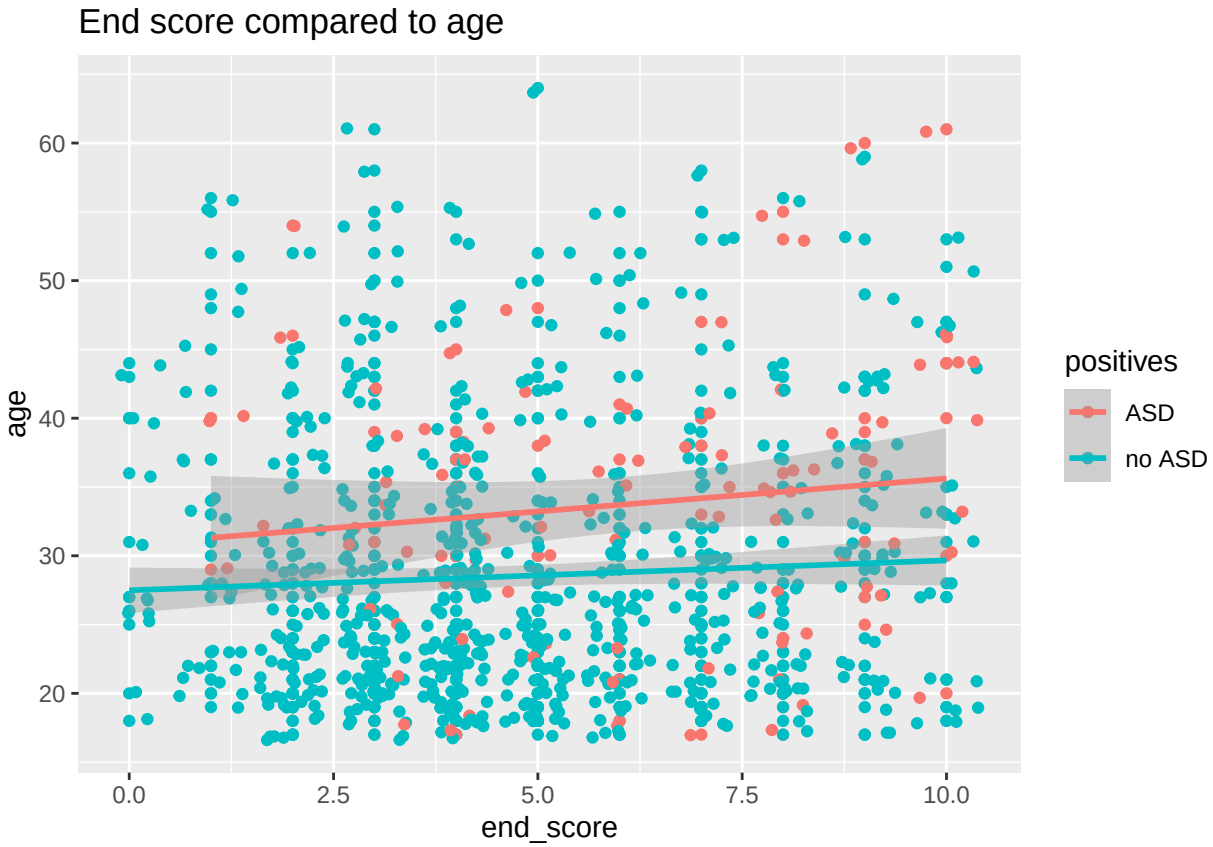


Figure 5, Some correlation that would not be good would be the correlation of age and the end score. Something that could be possible is that the older you are, the worse you make the test. This would be bad because then the AQ-10 test would be based on age and not on people actually having ASD. It does not look like this is the case though. Let's confirm this by doing a correlation test!

```
##
##  Pearson's product-moment correlation
##
## data:  data$end_score and data$age
## t = 2.5981, df = 699, p-value = 0.009571
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.02392249 0.17061353
## sample estimates:
##       cor
## 0.09779918
```

Discussion

All of the results conclude that the data is indeed usefull for a machine learning algorithm. The plots show that The important variables (Questions score and jaundices) actually have a corralation of people having ASD. The First plot shows the age people are taking this test. This is good to know to see if age has a correlation to the test. The second plot shows the anwsers of the questions. It shows that the first and eighth question actually has a much more yes than no rate. This is quite unusual. It could mean that some questions are not right and give people points without actually having ASD. This should be investigated on more since it could make the machine learning algorithm less trustable.

The third plot shows the correlation between peoples end scores and them actually having ASD. The conclusion of this figure is that the difference between their end score is actually significant and is confirmed with a t-test. The fourth plot shows the amount of people with ASD taking the test compared to the people with ASD who also have family members with ASD. The results are kind of unexpected. There is a big chunk of people who don't have ASD but still have family members with ASD. This makes sense because these people probably want to take the test because they know they have a family member with ASD and are scared they have it too.

The last figure shows the correlation between the age of the people taking the test compared to their end scores. With people with ASD marked. The conclusion of this is that there is no correlation between the two in the end which is confirmed by using a correlation test. This is a good thing because if there would be a correlation. The test could be based on age instead of on the fact that people have ASD or not.

The goal was to find out if the Q10 test actually knows if people have ASD. The conclusion of this is that it can never fully know it. But it can predict it sometimes. The weaknesses of the test are that the anwsers could be based on other things than ASD. This test together with more information about a person could be used in the future by doctors to figure out if a patient has ASD or not.