

# Written Report on Factors that Help Diagnose Autism

The Repos - Jeffrey Bohrer, Alexandra Green, Anna Zhang, Kevin Lee

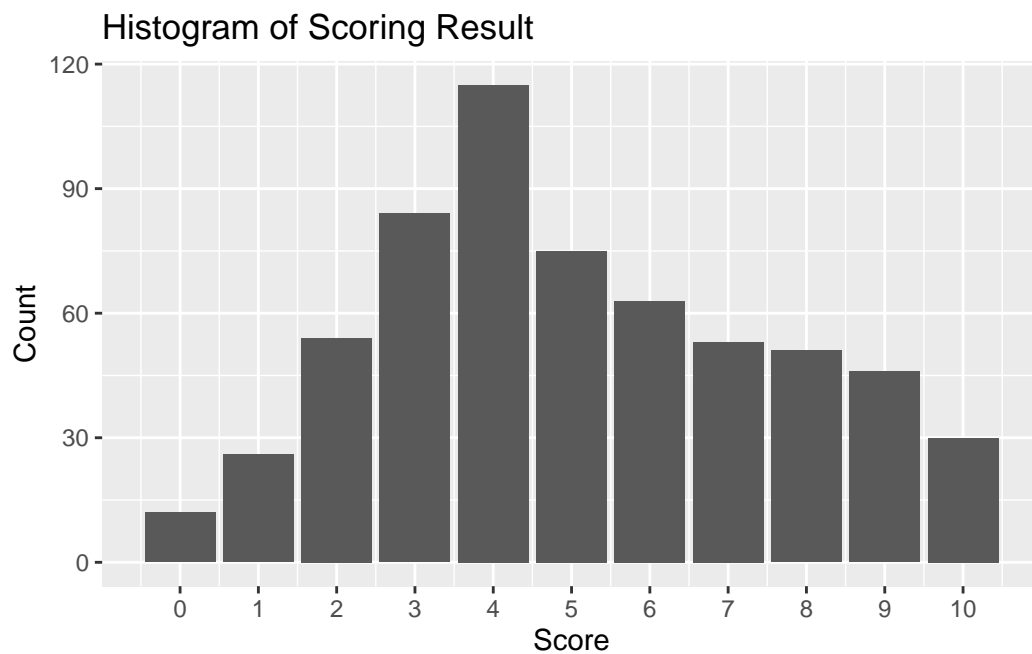
2025-03-20

## Exploratory data analysis

### Introduction

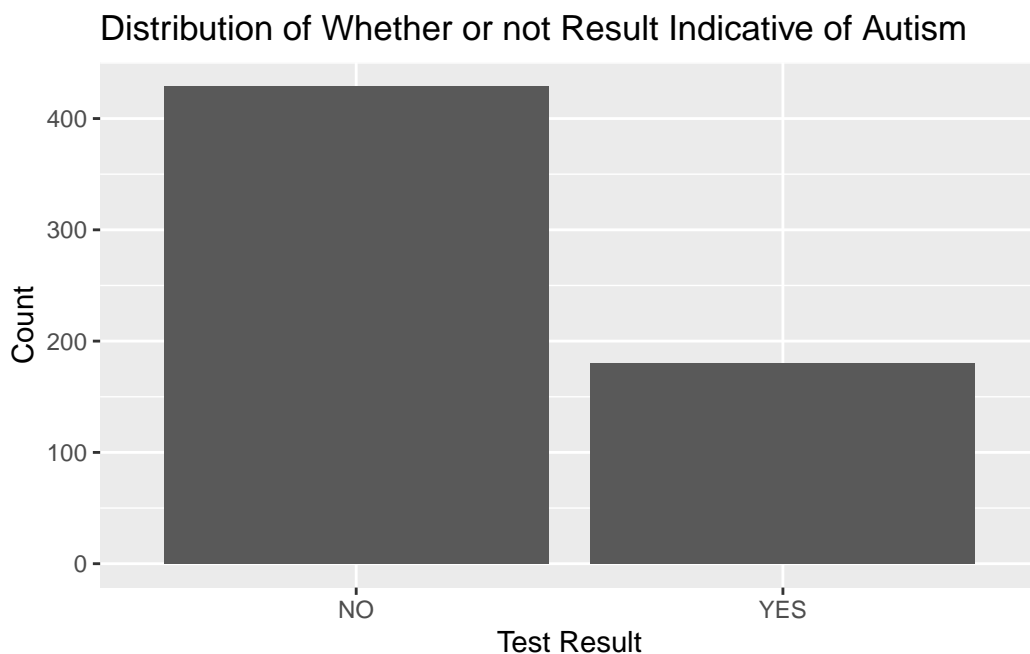
#### Univariate EDA

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	3.000	5.000	5.077	7.000	10.000

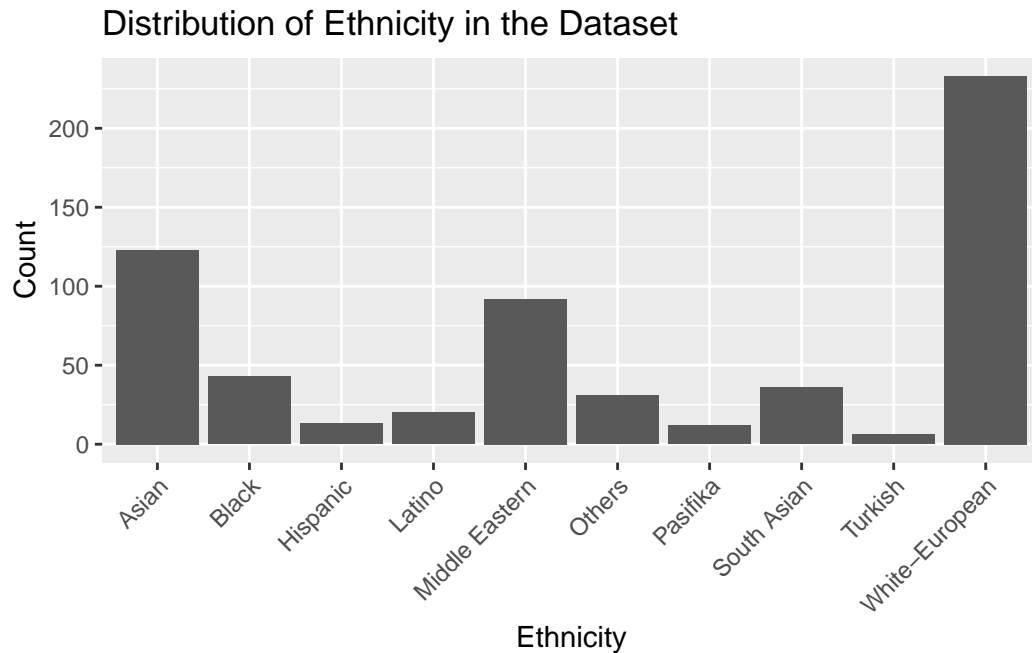


[1] 0.2955665

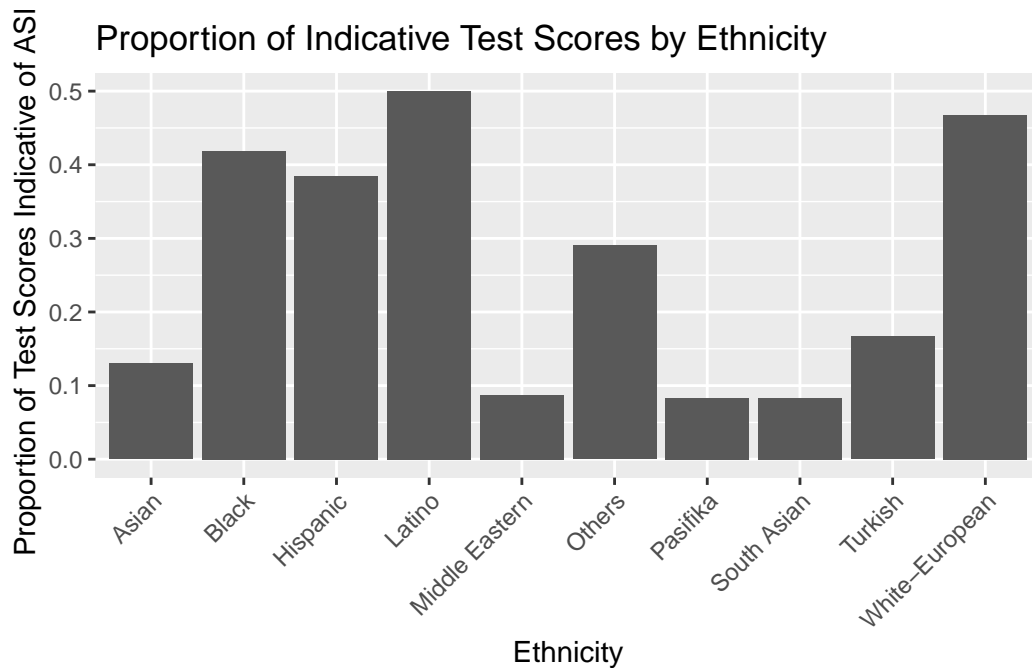
As we can see as a quick summary of our data, the variable we are most interested in – the final score of users – ranges from 0 to 10, which makes sense as there are 10 questions that can be answered either ‘Yes’ or ‘No’. Generally, the scores are a bit right-skewed, with a single peak around 4 or 5 points. The mean is found to be 5.077 points, and the median is 5 points, which is relatively high given that a score of higher than 6 warrants further investigation into a diagnosis. However, as suspecting a diagnosis is reason to take a test in the first place, that is a reasonable reflection of the population taking this test. We then observe that roughly 30% of test takers are encouraged to seek a diagnosis, due to a score higher than 6. We also observe the IQR to be 4 points, as most test takers have a score between 3 and 7 points inclusive. This is a reasonable spread in the context of the potential values, and in this context, no outliers exist.



Additionally, here is a visual demonstration of the number of test takers receiving further encouragement to pursue a professional diagnosis of ASD. Though still significantly less than those not encouraged, a notable proportion of test takers were informed they had traits or behaviors indicative of autism.

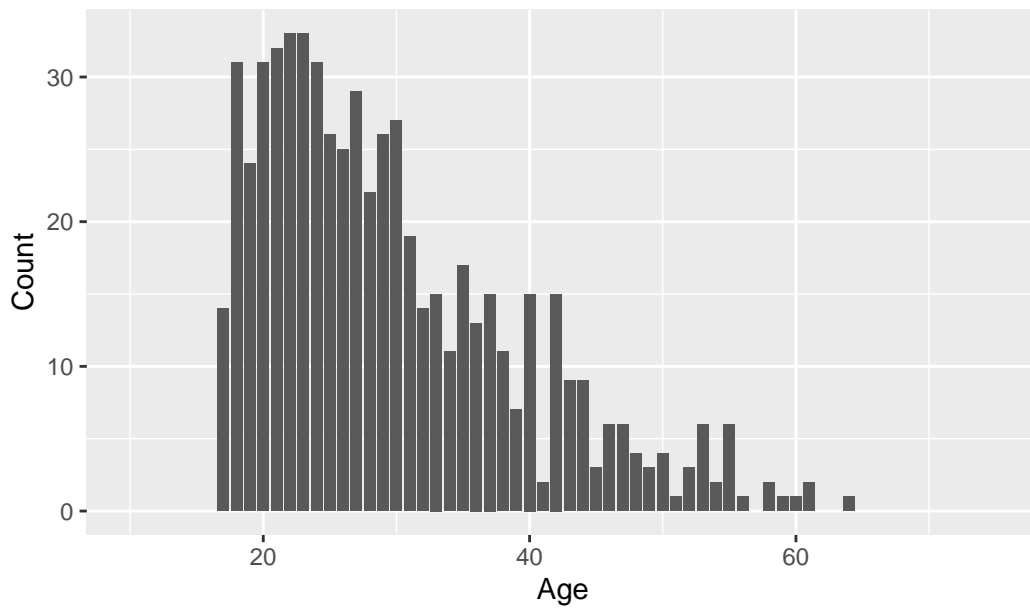


We observe here that the most common ethnicity was White-Europeans, with over 200 observations in our data set. Secondly and thirdly, we observe the Asian and Middle Eastern populations to be above 100 and slightly below 100 observations, respectively. All other ethnicities have less than 50 observations in our data set, meaning it may be more difficult to draw conclusions about these populations. We can now check and see how the percentages of being encouraged to seek a diagnosis differ by ethnicity:



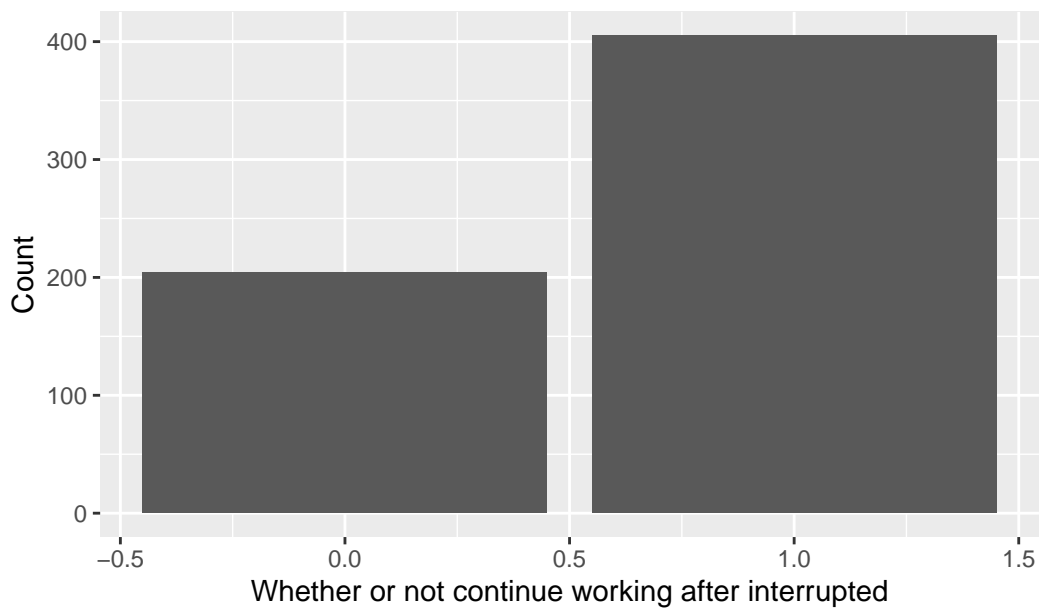
As we can see here, the population with the highest proportion of test takers receiving a high test score was the Latino population, with close to half achieving higher than a 6. Since this is a comparably uncommon population in our data set, it does raise questions of whether it is indicative of the greater Latino population or merely due to the smaller sample size in our data set. Additionally, the White-European, Black, and Hispanic populations also demonstrate a greater proportion of high test scores. The variation in these proportions according to ethnicity suggests that ethnicity may be a potential predictor for the results of the test, or even for test scores.

Distribution of Age in the Dataset



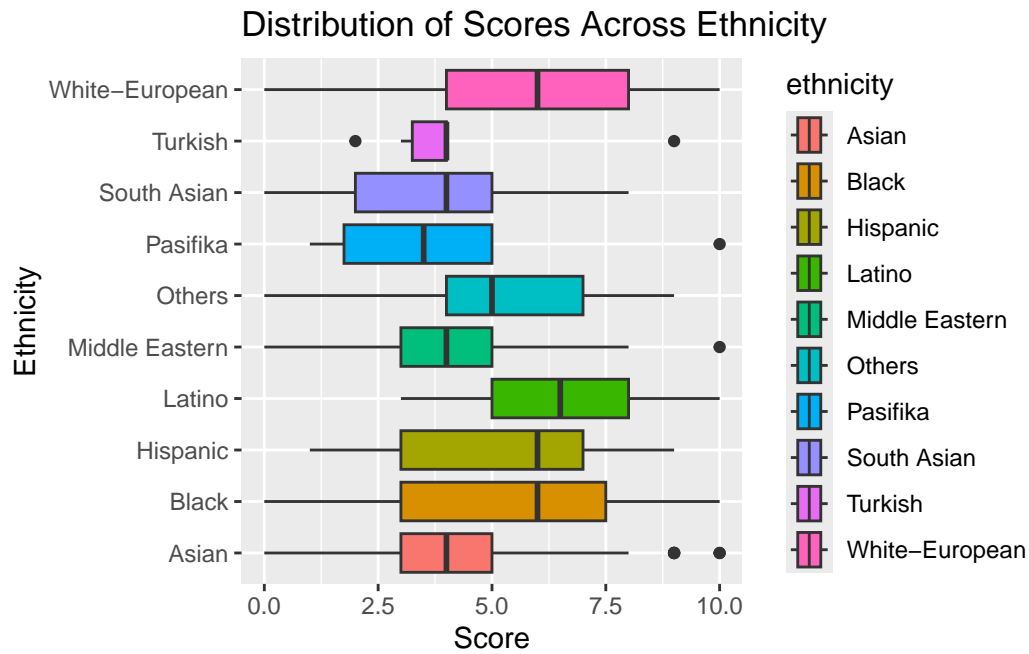
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
17.00	22.00	27.00	30.22	35.00	383.00

Distribution of the Ability to Return to Work after Interruption in

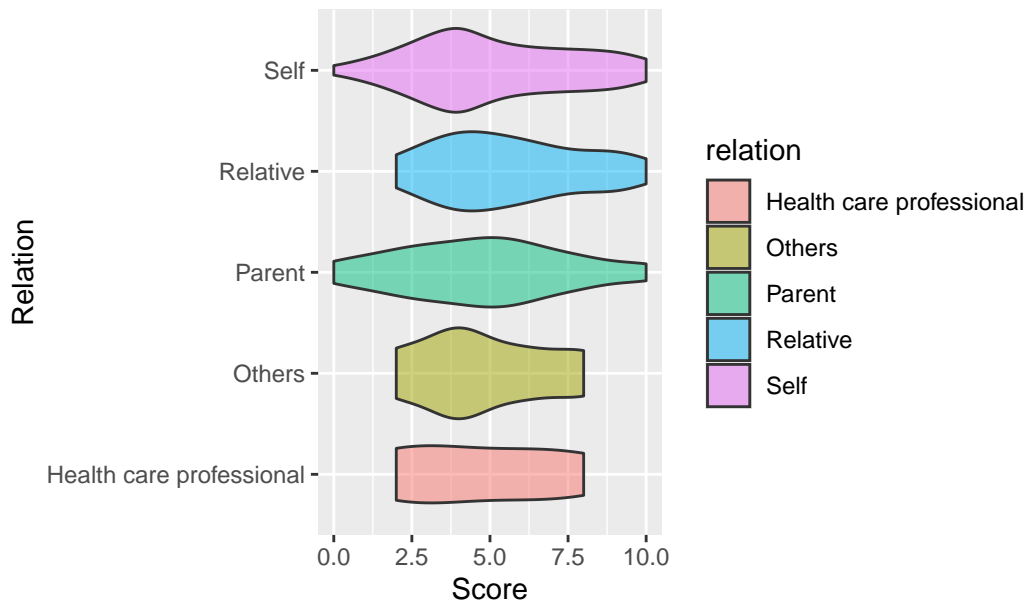


Generally, we observe the distribution of ages to possess a strong right-skewness, with a mean age of 30.22 and a median age of 27. There is a clear peak in the distribution roughly around early to mid 20's years of age. The ages possess a minimum of 17 up to a maximum age of 383, which is clearly a false observation that needs to be filtered from the data. The IQR is 13 years, which is a fairly small spread given the range of ages, as we observe that it seems the majority of test-takers are under 30 years old.

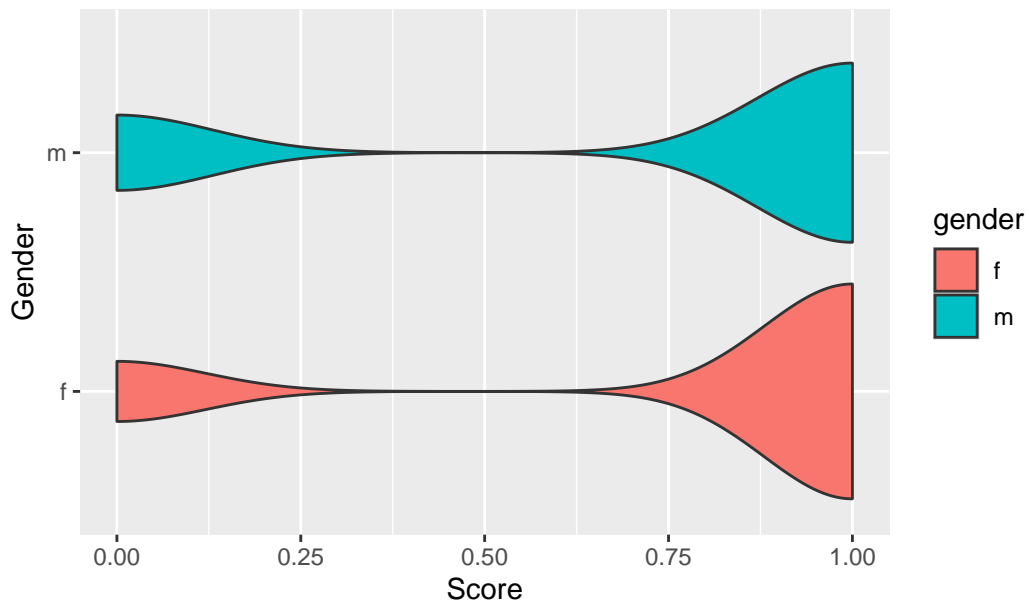
## Bivariate EDA

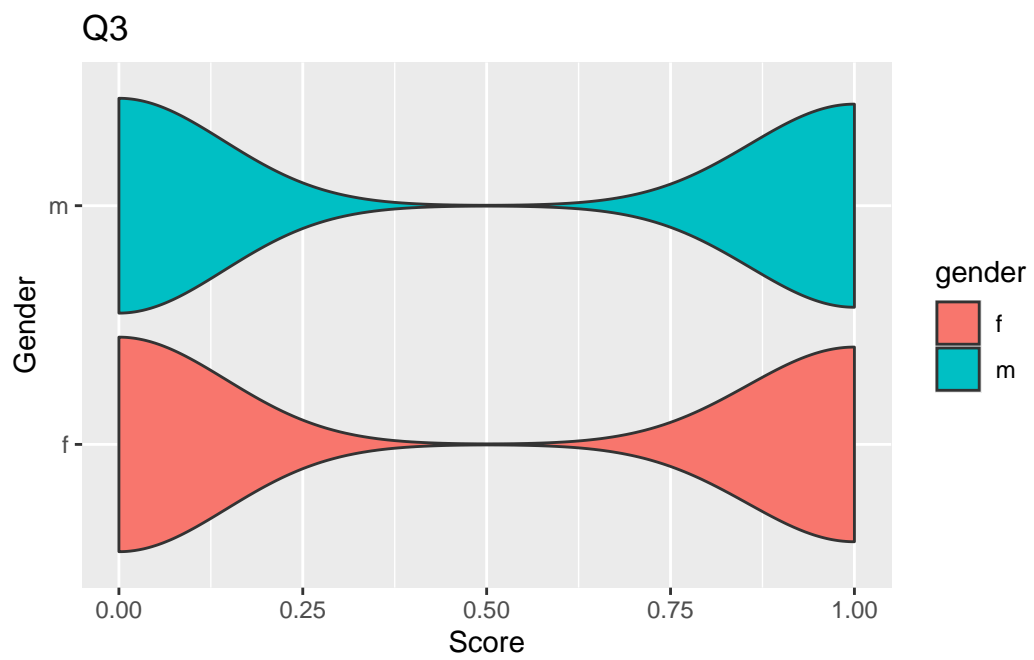
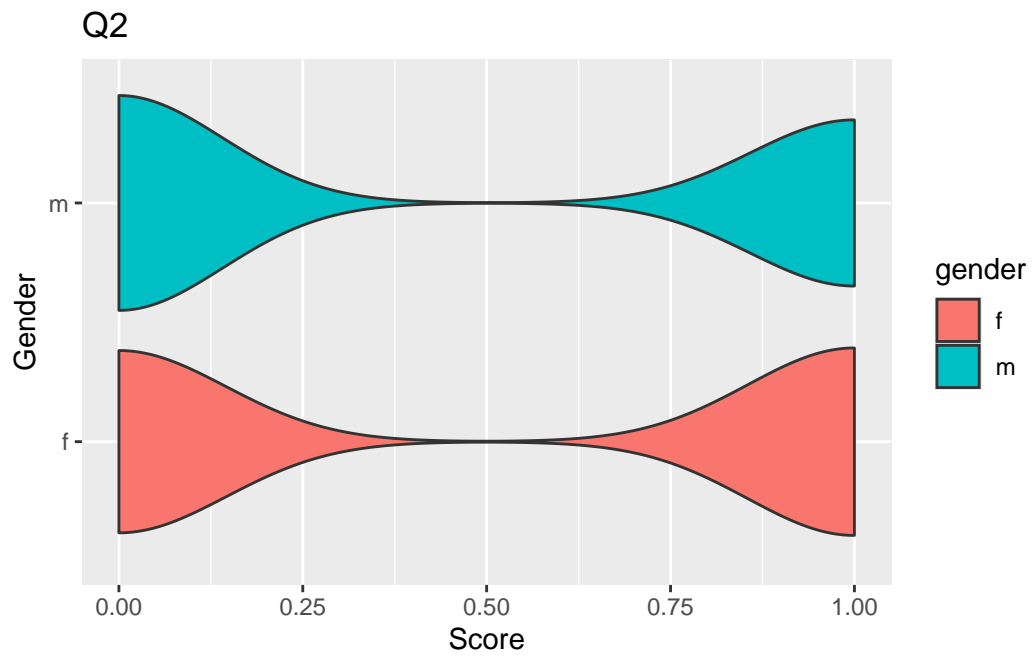


Distribution of Scores Across Person Filling Test



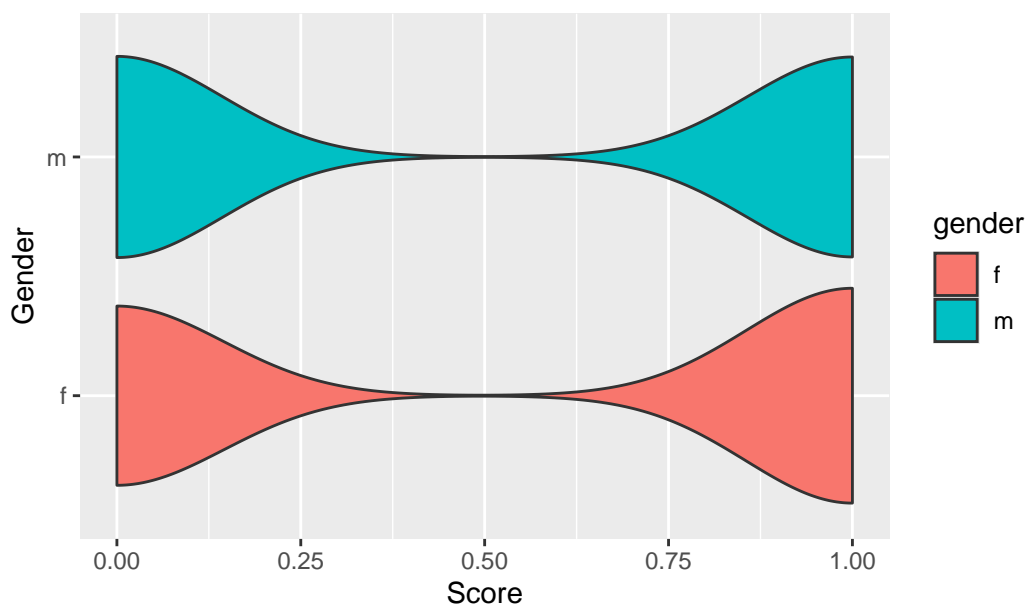
Q1



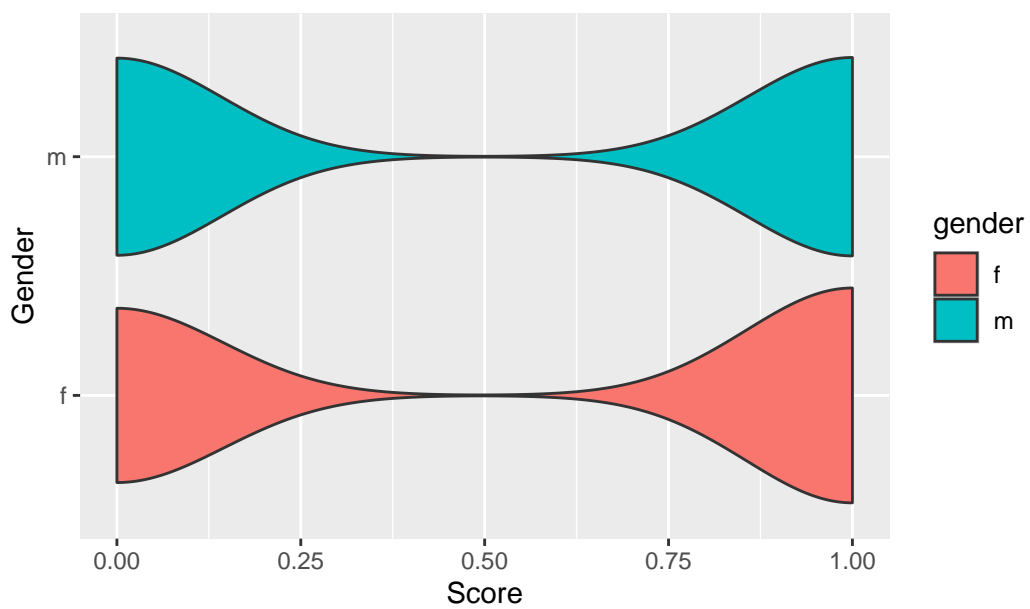




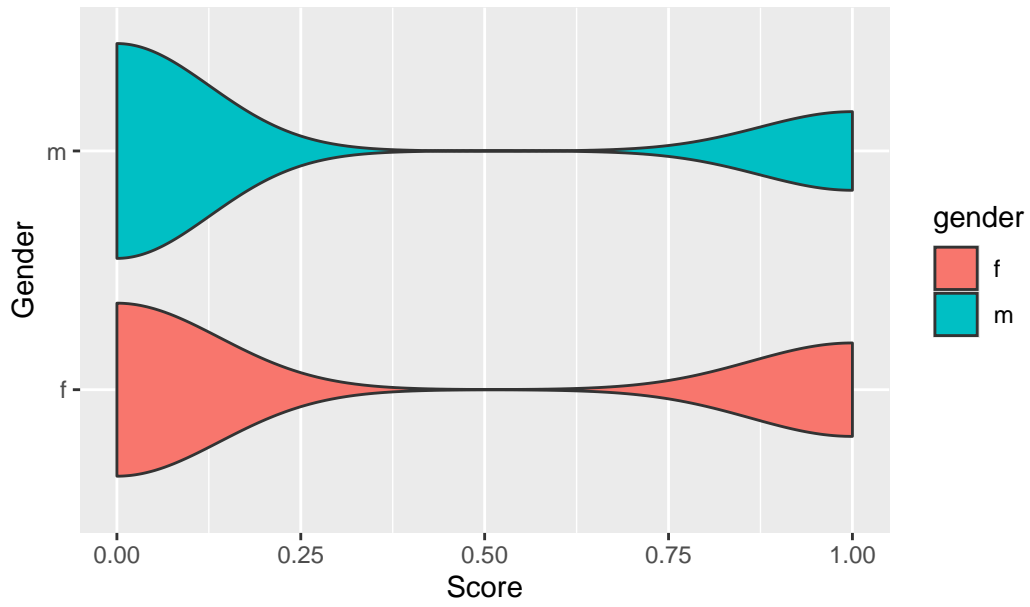
Q4



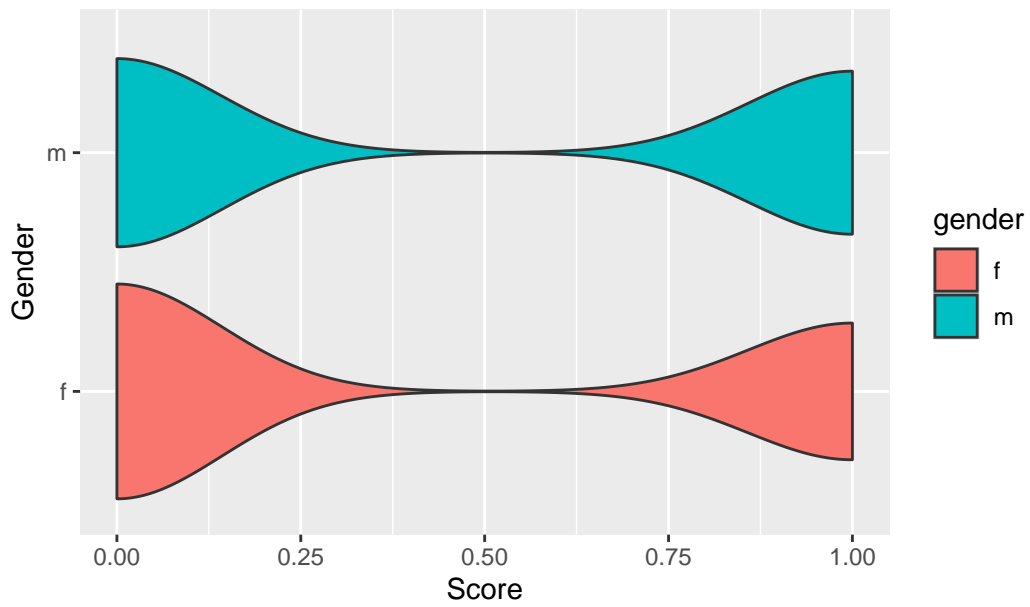
Q5



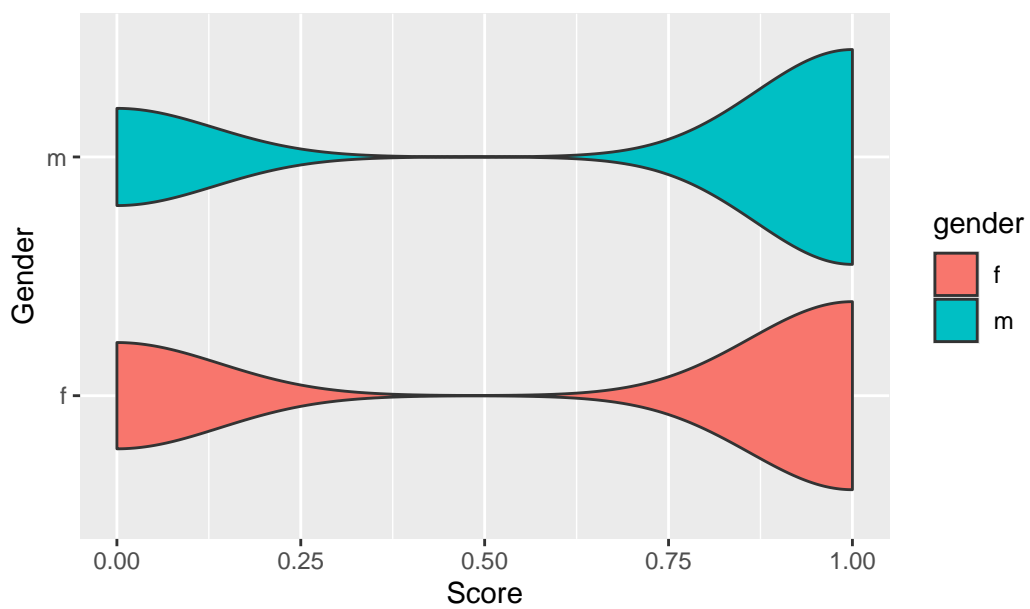
Q6



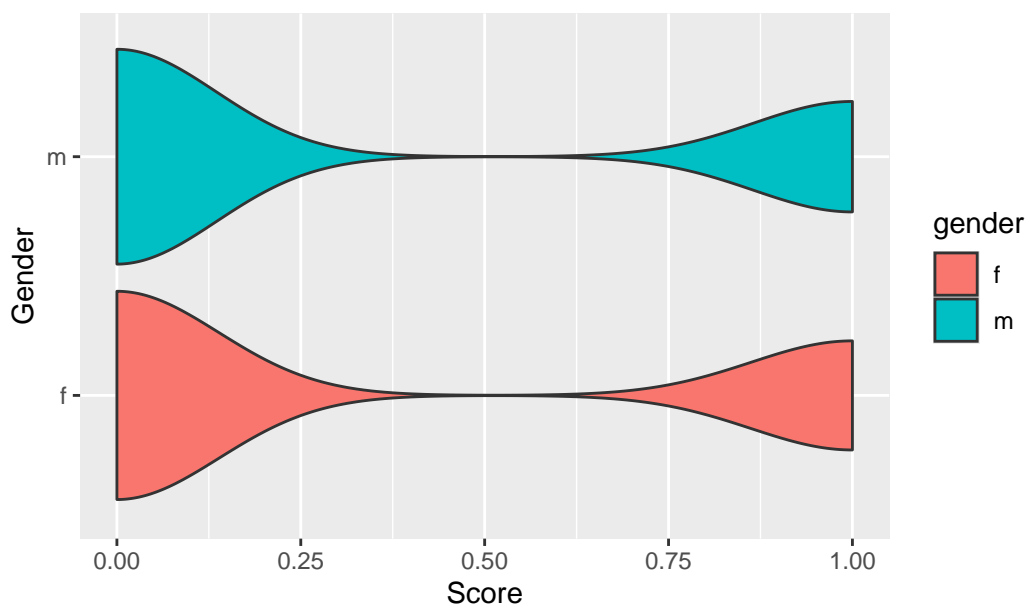
Q7

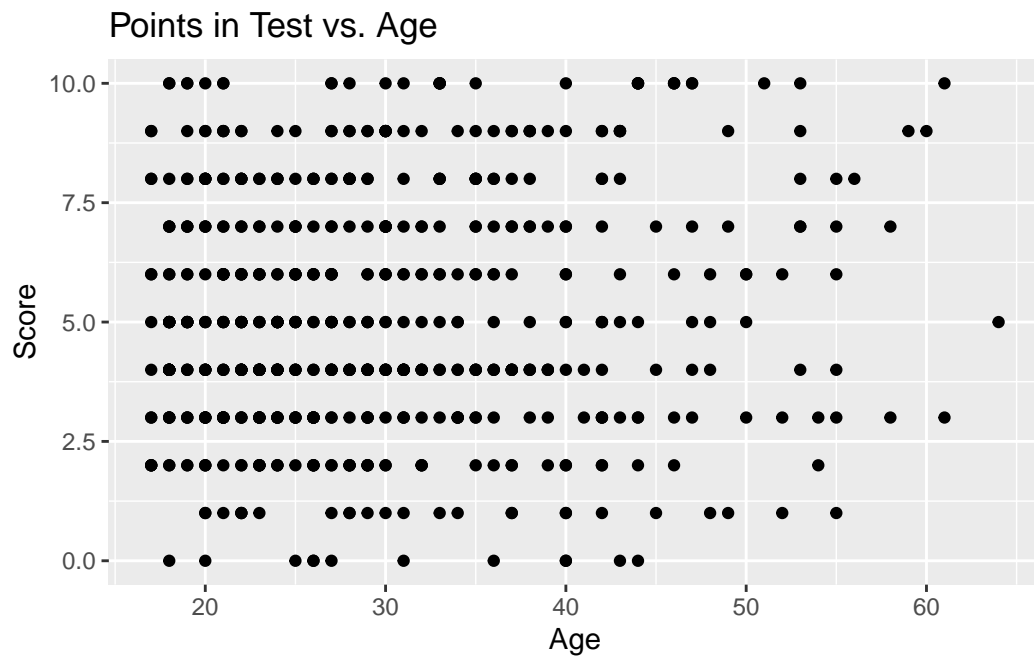
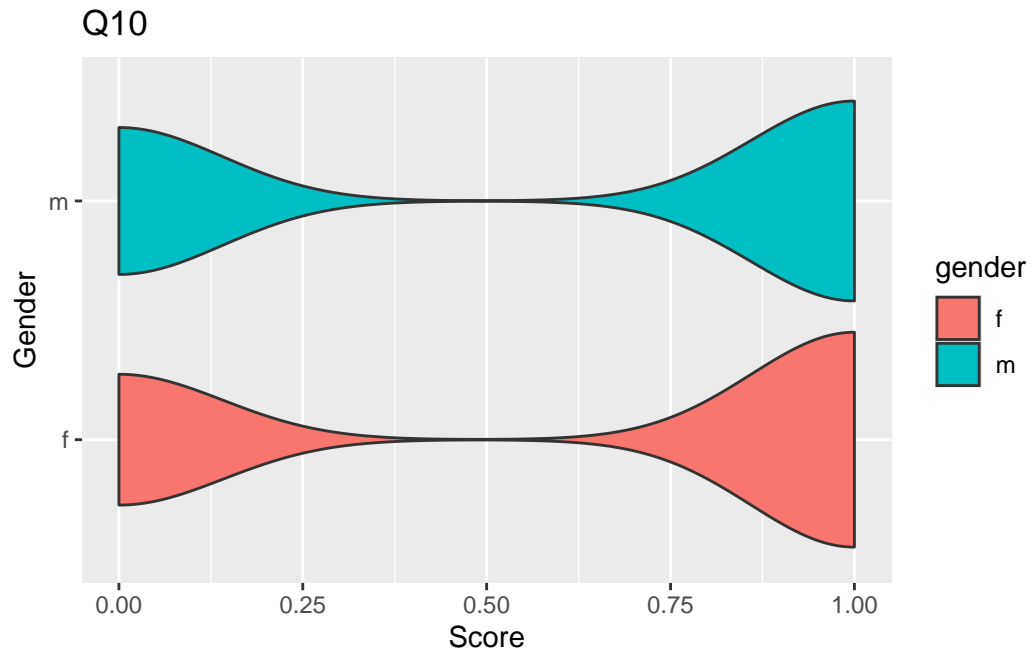


Q8



Q9





! Important

Before you submit, make sure your code chunks are turned off with `echo: false` and there are no warnings or messages with `warning: false` and `message: false` in the YAML.