Written Report on Factors that Help Diagnose Autism

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

2025-03-20

Introduction

Autism Spectrum Disorder (ASD) remains a highly prevalent condition despite modern strides made in medical technology. It is reported that nearly 2.2% of adults are affected by ASD, and growing awareness has led to an uptick in diagnoses, particularly in adults who went undiagnosed early in life (Hirota 2023). However, ASD screening tests for all age groups currently contain significant inaccuracies. For example, the most widely used toddler screening test, CHAT-R/FAs, was recently found to produce false negatives in 25% of cases. In contrast, the most commonly used adult autism screening test – the Autism-Spectrum Quotient (AQ) – was found to have limited predictive value in certain populations (Aishworiya 2023; Curnow 2023). Therefore, it has become critical to identify stronger predictors or explore underlying relationships to have more accurate tests and models to predict ASD in adults.

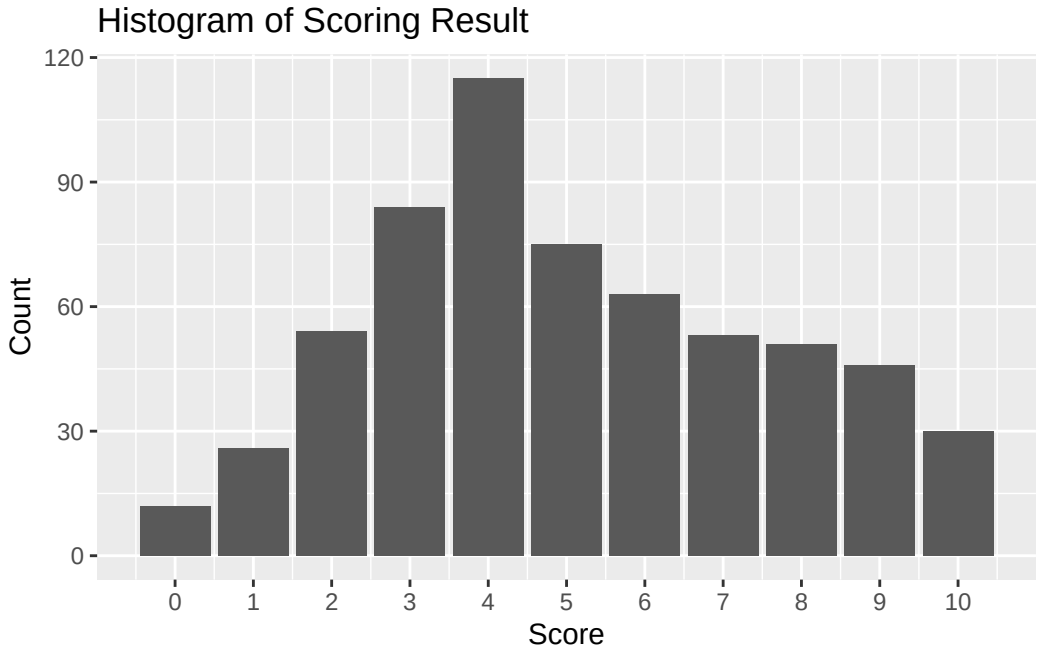
In this study, we will focus on identifying the features that most greatly affect the probability of being encouraged to pursue a diagnosis. The dataset being used will contain ten individual characteristics and ten behavioral questions, each of which has been associated with ASD diagnoses. Because ASD is difficult to identify and can significantly impair an individual's quality of life, understanding the relationship between demographics, certain behaviors, and their association with autism could encourage individuals to seek diagnosis and gain the self-understanding they need. These adults who receive a positive diagnosis can then access the necessary support and resources. We believe that individual characteristics such as gender and ethnicity, along with answers to questions assessing socialization abilities, are key predictors of autism.

The test scores were slightly right-skewed with no significant outliers, and despite the majority of test-takers not being encouraged to seek diagnosis, a significant proportion was. In the test, the potential ASD indicator of difficulty resuming work after a distraction was frequently noted. Within racial distributions, Latino participants had the highest proportion of high scores, White-Europeans were the most common, and certain groups had significantly higher IQRs, such as Black and Hispanic groups, compared to groups such as the Turkish and Middle

Eastern. However, some ethnicities had fewer than 50 observations, limiting generalizability. Between men and women, we also noted minimal differences.

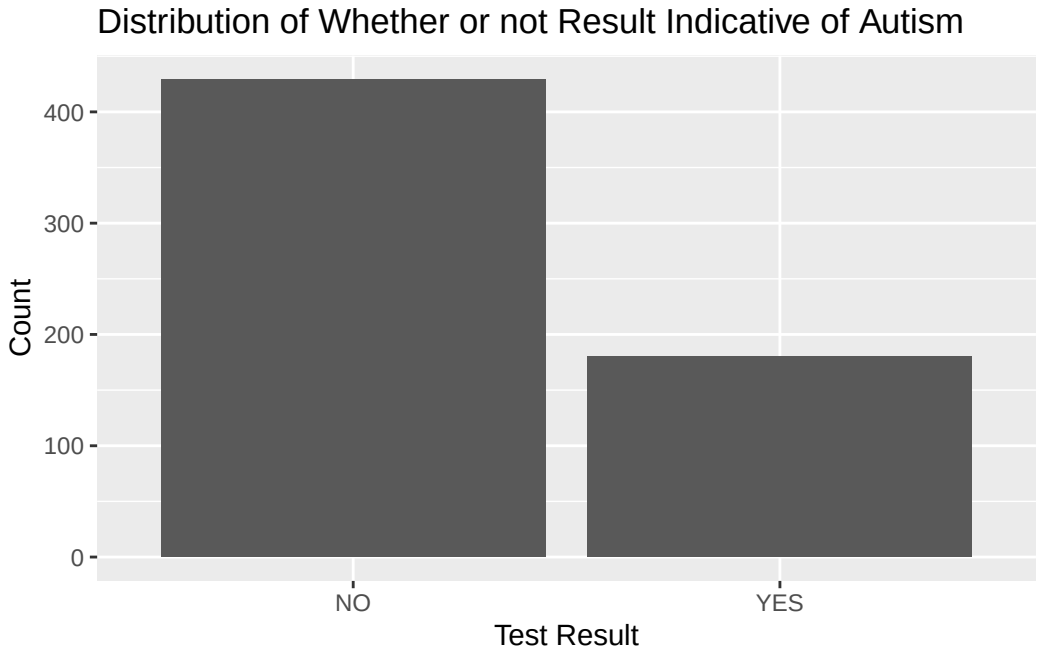
Univariate EDA

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

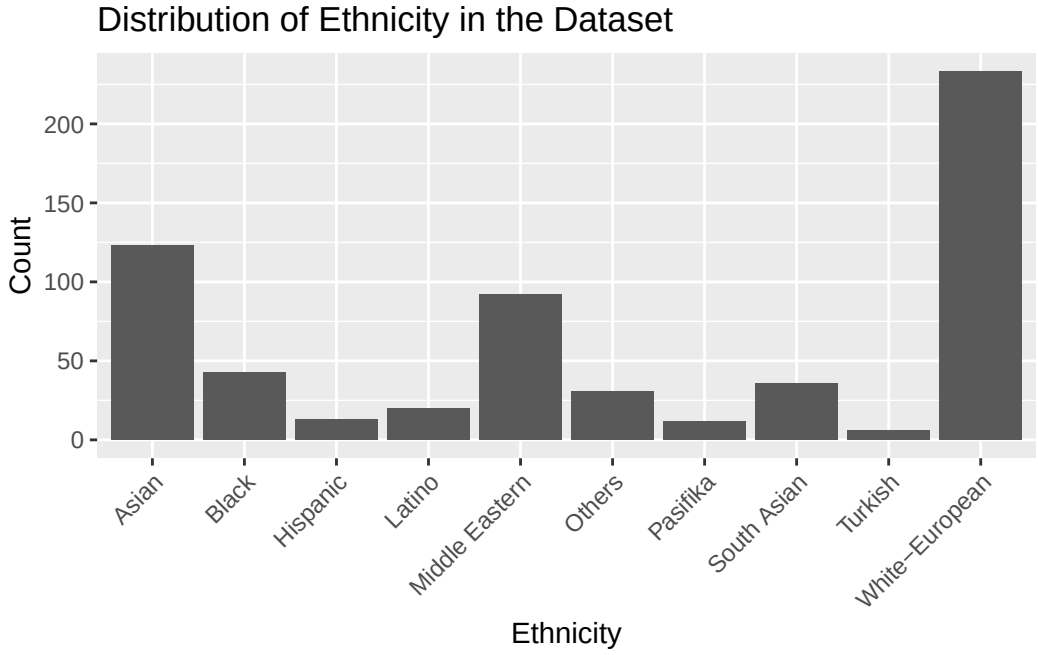



```
[1] 0.2955665
```

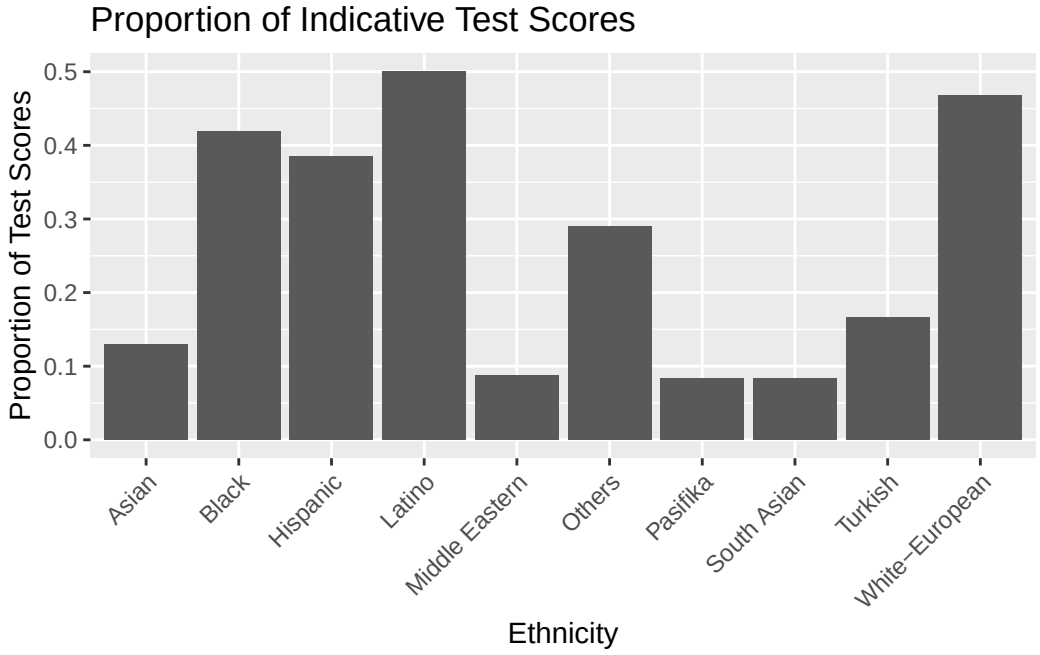
As we can see as a quick summary of our data, the variable we are most interested in – the final score of users – ranges from 0 to 10, which makes sense as there are 10 questions that can be answered either 'Yes' or 'No'. Generally, the scores are a bit right-skewed, with a single peak around 4 or 5 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.



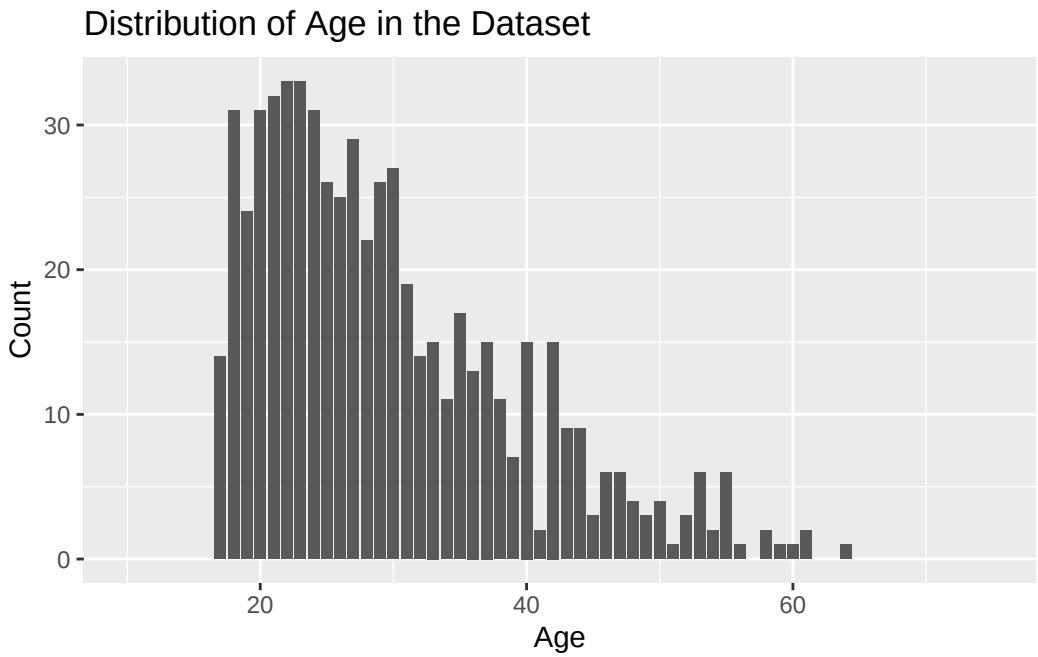
Additionally, here is a visual demonstration of the number of test takers receiving further encouragement to pursue a professional diagnosis of ASD. Though still significantly less than those not encouraged, a notable proportion of test takers were informed they had traits or behaviors indicative of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



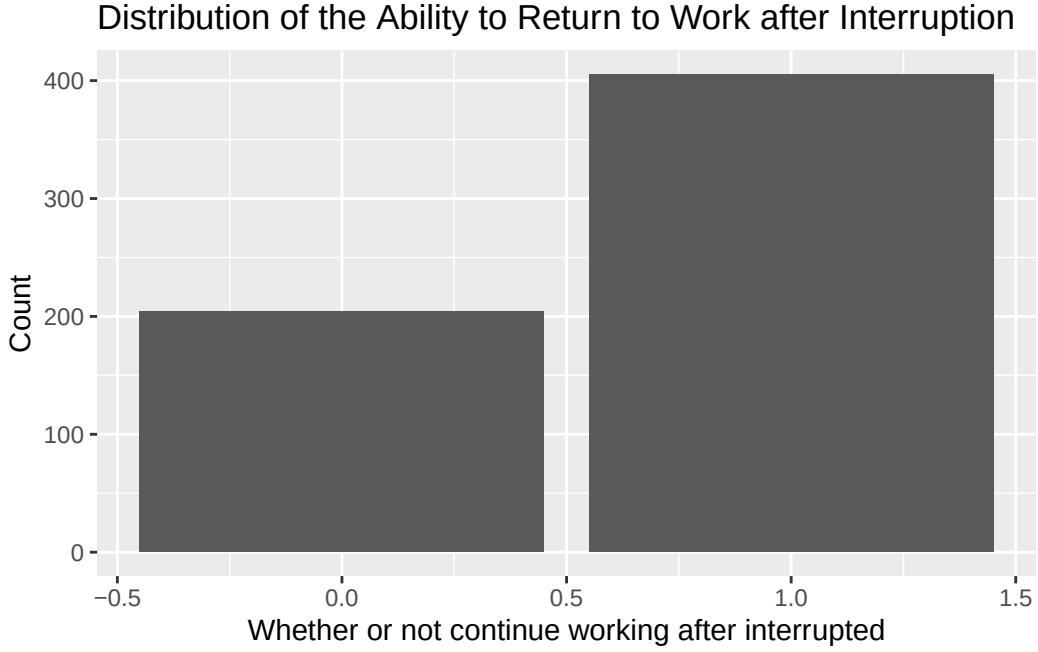
We observe here that the most common ethnicity was White-Europeans, with over 200 observations in our data set. Secondly and thirdly, we observe the Asian and Middle Eastern populations to be above 100 and slightly below 100 observations, respectively. All other ethnicities have less than 50 observations 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 warrants further exploration for how the distribution of test scores differs by ethnicity as well.



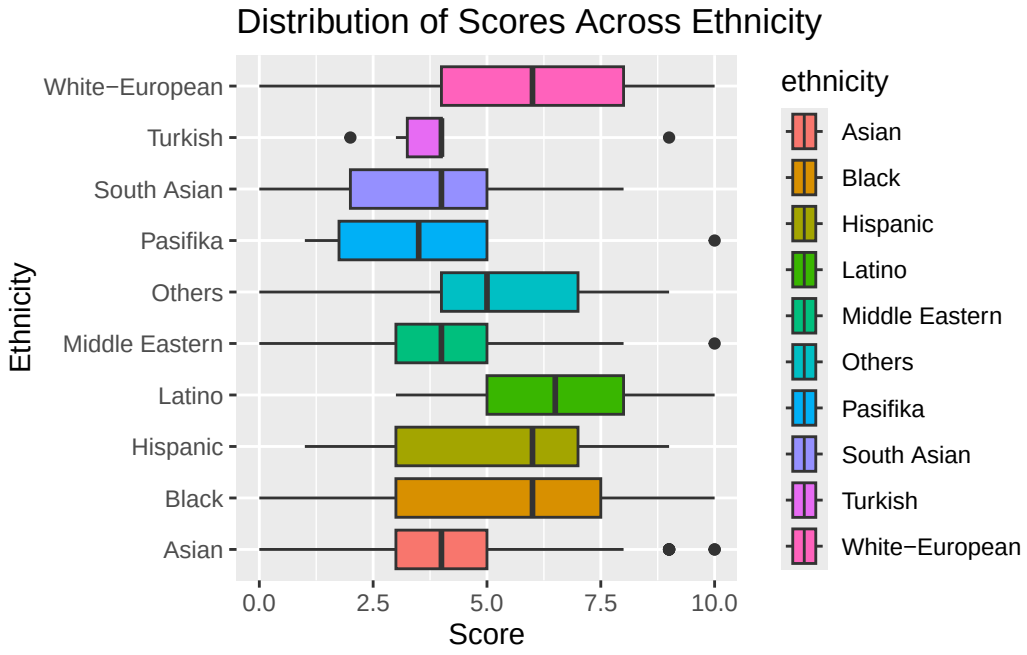

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 17.00   22.00   27.00   30.22   35.00  383.00
```

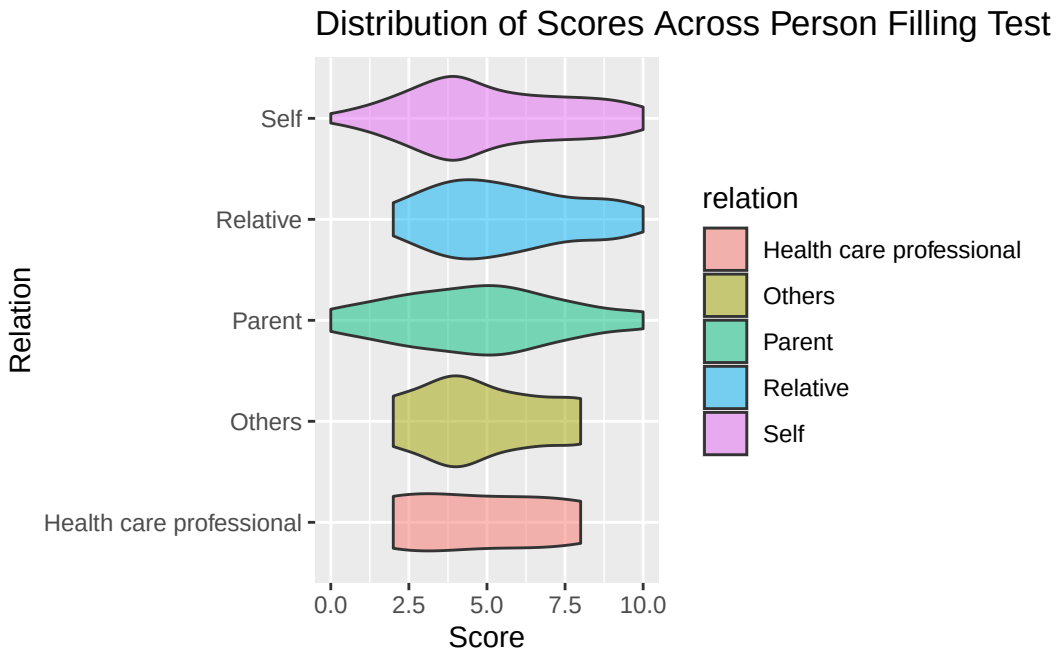


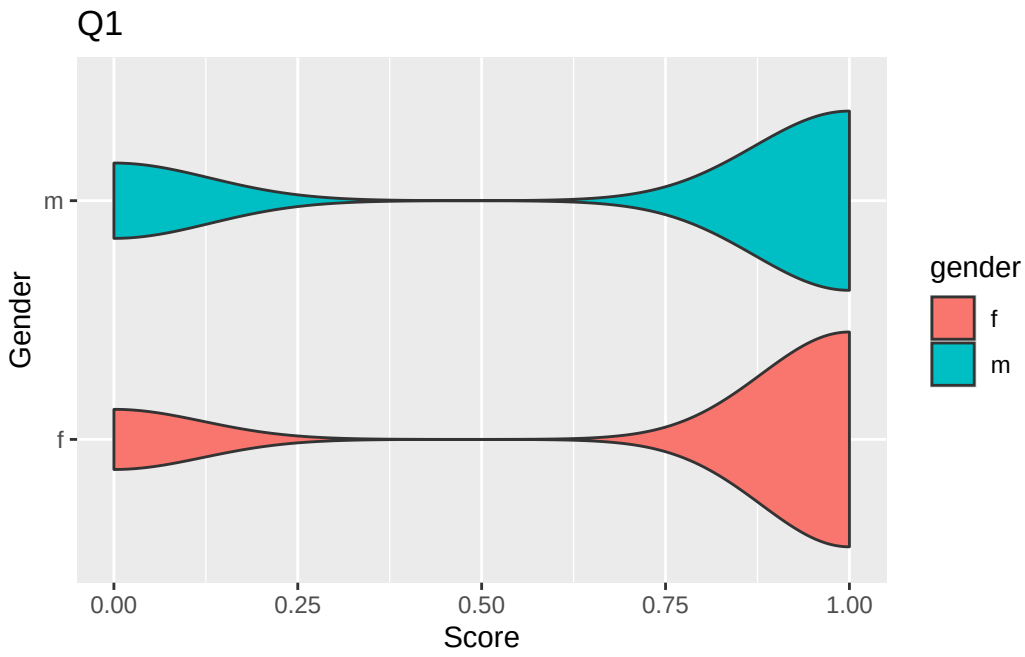
Generally, we observe the distribution of ages to possess a strong right-skewness, with a mean age of 30.22 and a median age of 27. There is a clear peak in the distribution roughly around early to mid 20's years of age. The ages possess a minimum of 17 up to a maximum age of 383, which is clearly a 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.

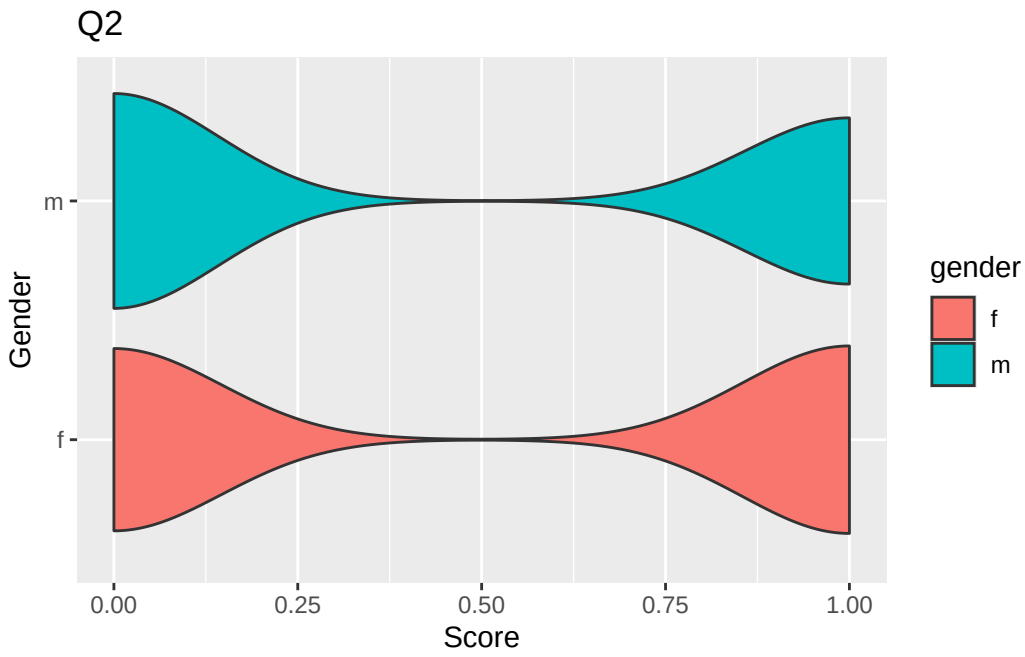
Additionally, as a brief glance at one of the ten questions on the test, we see that "the ability to continue working after an interruption" question had majority of test takers indicate that this was an area of difficulty, which acts as a potential indicator for ASD.

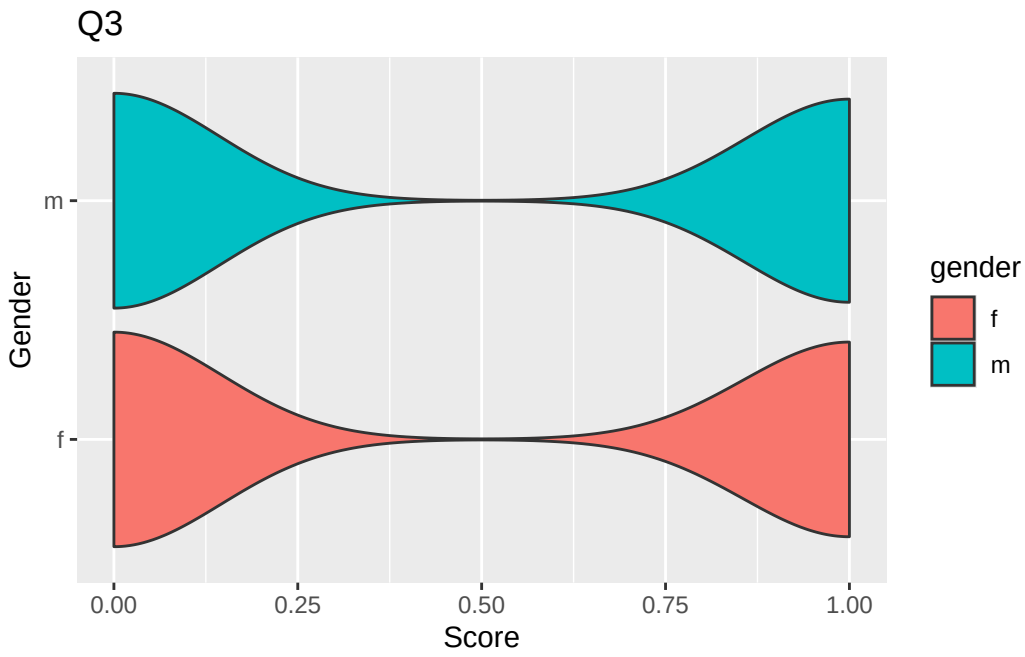
Bivariate EDA

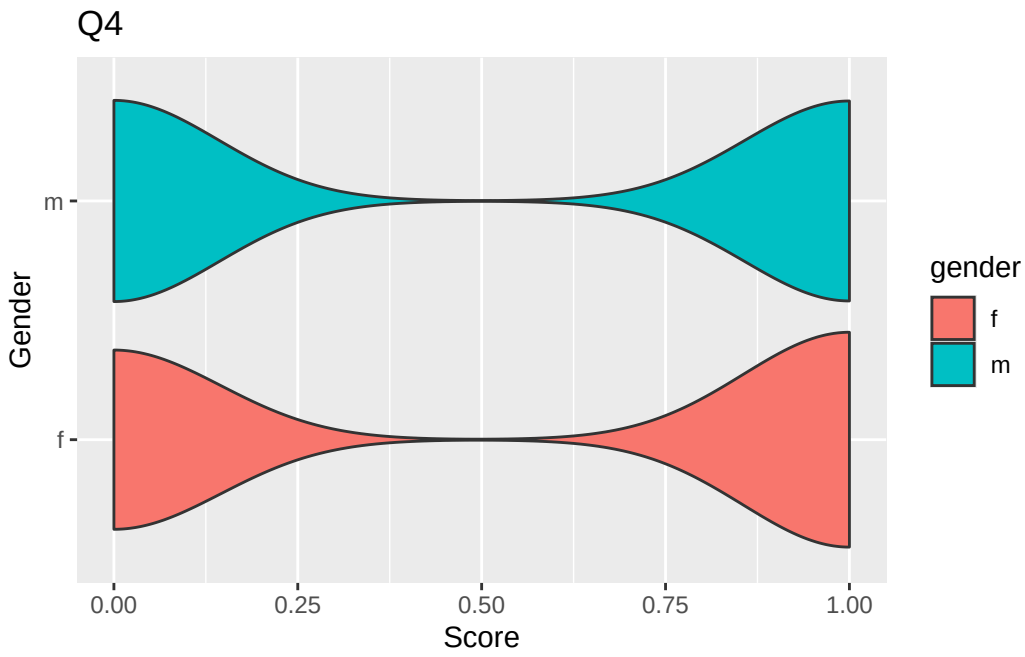


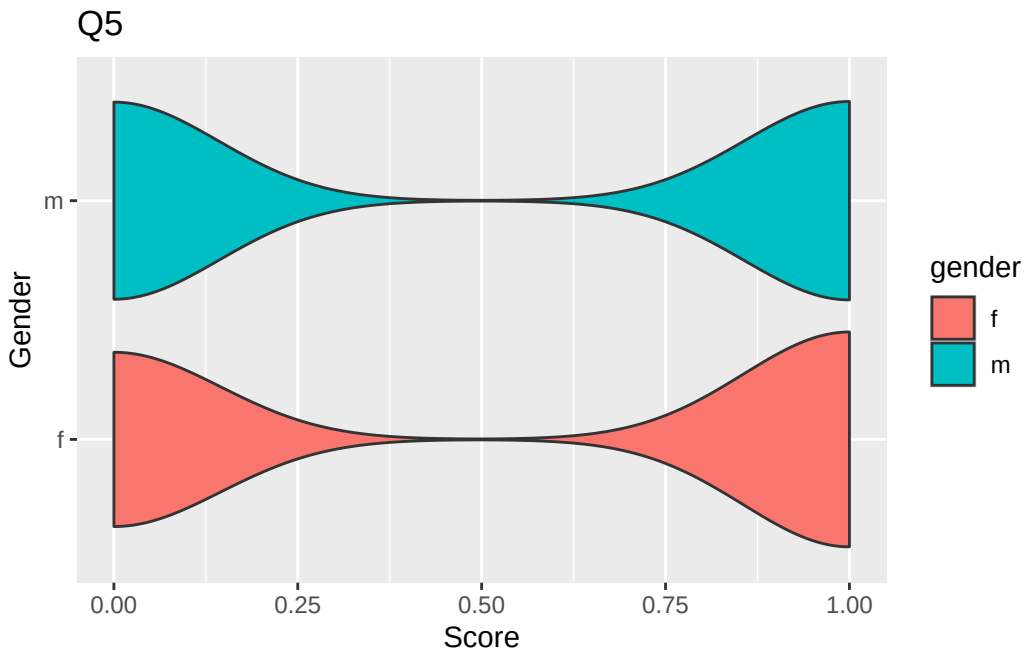

Distribution of Scores Across Person Filling Test
Relation
Self
Relative
Parent
Others
Health care professional
0.0
2.5
5.0
7.5
10.0
Score
relation
Health care professional
Others
Parent
Relative
Self

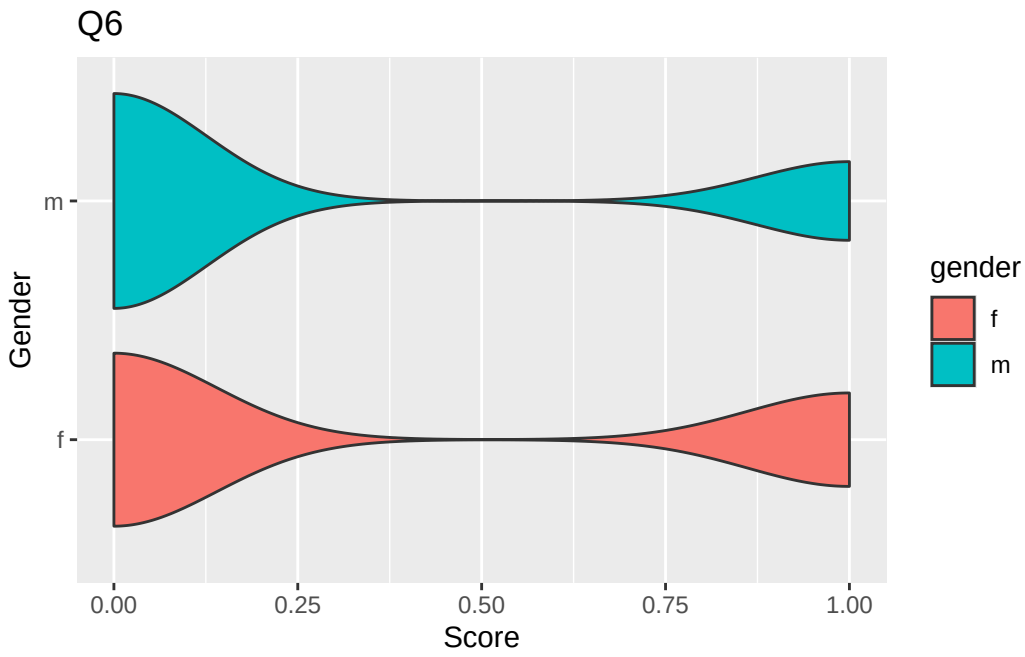

Q1
Gender
m
f
0.00
0.25
0.50
0.75
1.00
Score
gender
f
m

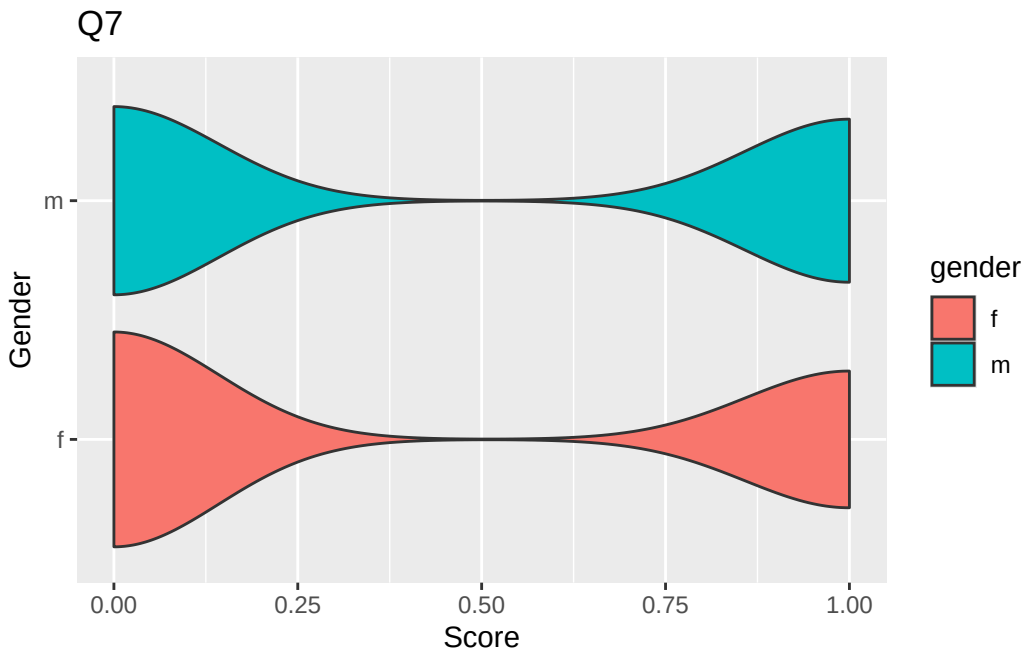

Q2
Gender
m
f
0.00
0.25
0.50
0.75
1.00
Score
gender
f
m

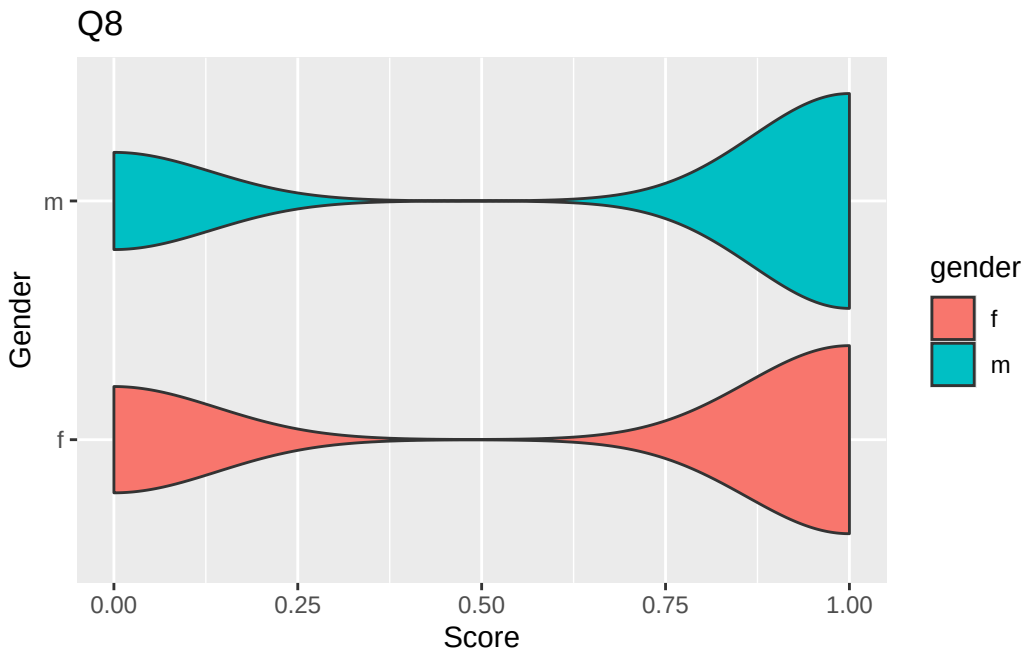

Q3
Gender
m
f
0.00
0.25
0.50
0.75
1.00
Score
gender
f
m

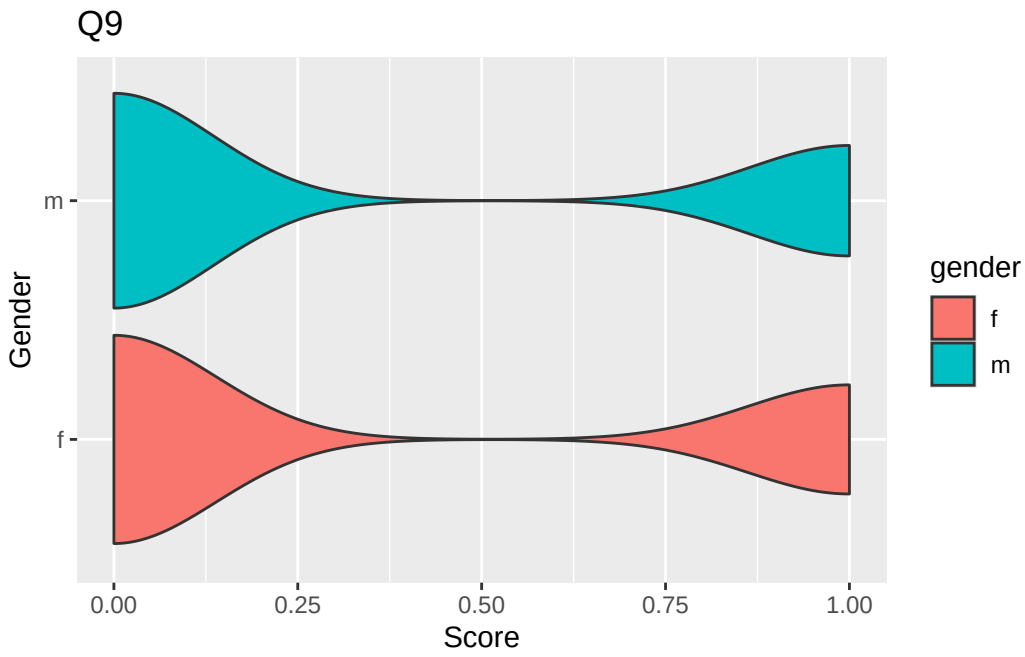

Q4
Gender
m
f
0.00
0.25
0.50
0.75
1.00
Score
gender
f
m

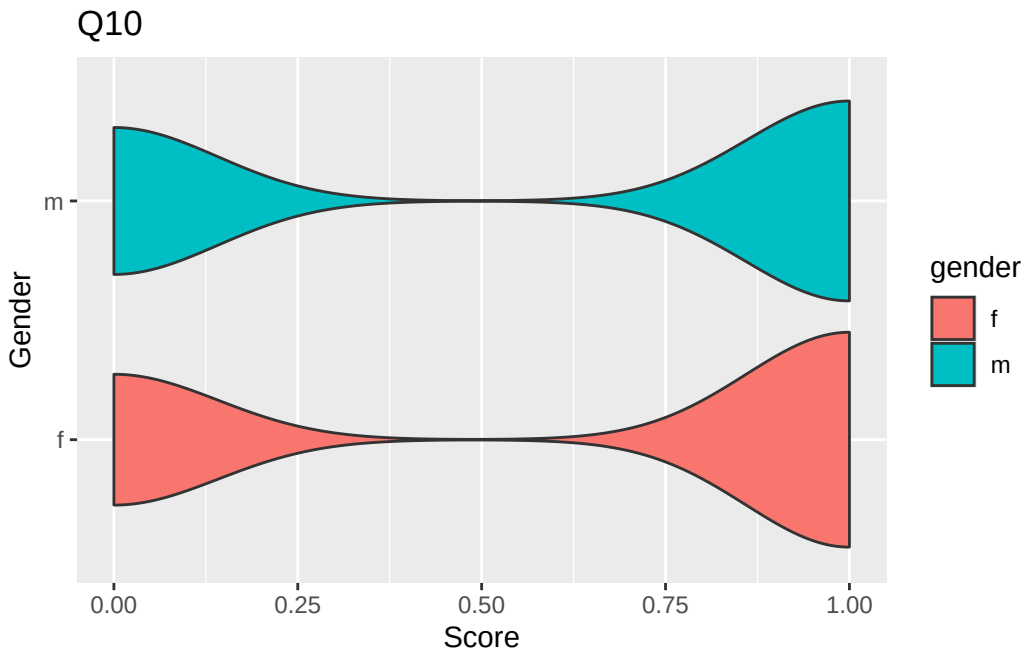

Q5
Gender
m
f
0.00
0.25
0.50
0.75
1.00
Score
gender
f
m

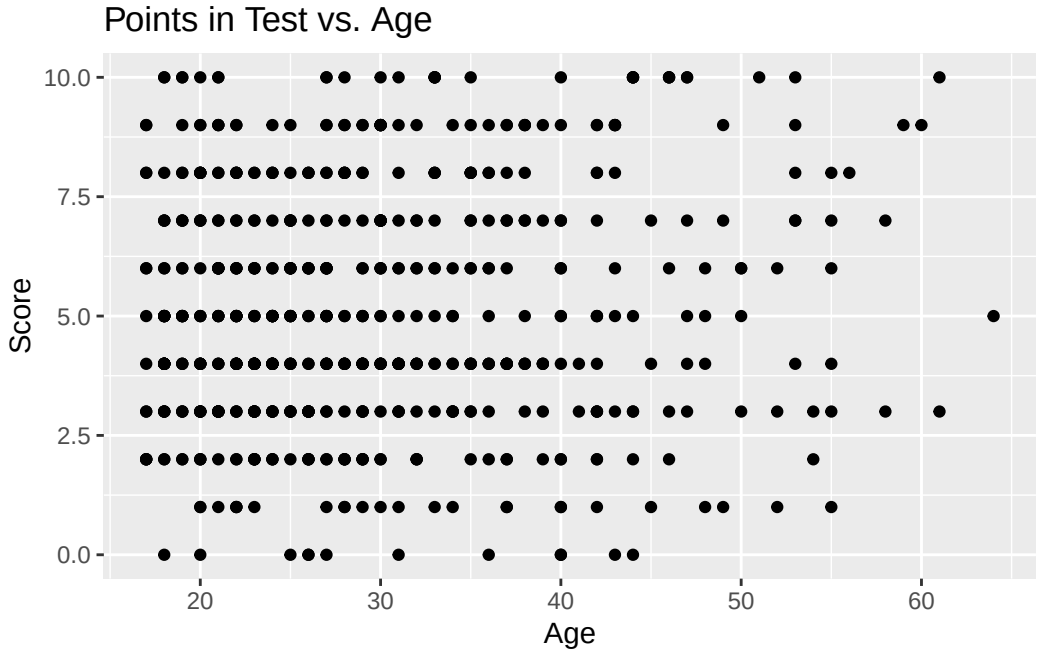

Q6
Gender
m
f
0.00
0.25
0.50
0.75
1.00
Score
gender
f
m


Q7
Gender
m
f
0.00
0.25
0.50
0.75
1.00
Score
gender
f
m


Q8
Gender
m
f
0.00
0.25
0.50
0.75
1.00
Score
gender
f
m


Q9
Gender
m
f
0.00
0.25
0.50
0.75
1.00
Score
gender
f
m





Here we present several graphs displaying bivariate relationships. With our first set of box plots, we find further evidence for our previous suspicion that test score distributions differ by ethnicity. Although most ethnicities have a median within the range of roughly 3 to 6, some ethnicities, like the White-European, Black, and Hispanic populations, demonstrate a

greater spread through their IQRs, while the Turkish, Middle Eastern, and Asian populations are much more concentrated around their medians. However, most ethnicities appear to have values almost entirely across the range of 0 to 10 in their test scores.

Interestingly, we can also observe how the relationship between the test taker and the subject of the questions may lead to different distributions of test scores. In the case when it is filled out by a relative or health care professional for example, the observed test score is 2 or greater, while in the case of the test taker being the subject or the parent of the subject, some still received a score of 0. This may reflect how personal biases or context affect truthfulness during the test.

Surprisingly, we can also observe minimal differences in how each gender answered the ten questions, with generally comparable distributions of 'Yes' and 'No'. This may be a point of further investigation as typically women tend to be underdiagnosed relative to men.

Methodology

Choosing Predictors

A drop-in-deviance test between a logistic null model without predictor variables and a logistic model with a single predictor was systematically conducted across ethnicity, gender, presence of neonatal jaundice, and relationship as a means of assessing which predictors provide a statistically significant improvement in model fit against the null condition. The hypothesis for this test can be observed below, where β_1 represents the coefficient for the associated predictor variable:

$$H_0 : \beta_1 = 0$$

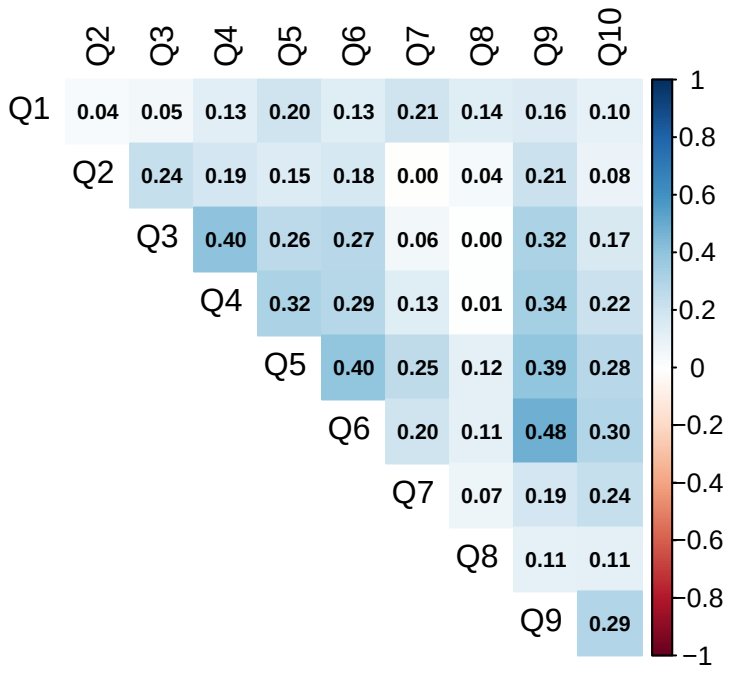
$$H_a : \beta_1 \neq 0$$

- Table 1: Drop-in-deviance test results comparing null model $log(\frac{\hat{\pi}}{1-\hat{\pi}}) = \beta_0$ to single-predictor model $log(\frac{\pi}{1-\pi}) = \beta_0 + \beta_1 X_1$

Predictor	G-stat	p-value
Age	1.54	0.22
Ethnicity	94.1	2.44x10^-16
Relation	1.21	0.87
Gender	4.46	0.035
Jaundice	9.33	0.0023

From the results of this test, it is evident that the parameters that are statistically significant for model fit are ethnicity, gender, and, quite surprisingly, neonatal jaundice when compared to a null model as the p-values are either significantly less than 0.05 (Q1-10 + ethnicity) or slightly below it (gender and jaundice). We also note that there appears to be strong correlation

between answers to questions 5, 7, and 10. Due to this consideration, the first model to be analyzed will be a simple additive model that takes into account all these statistically significant predictors.



By computing a correlation heatmap among the ten behavior questions to check multicollinearity, we note that Q8 is the least correlated to other questions, which is `i_can_go_back_to_work_when_interrupted`. Therefore, we consider adding this to the model.

Fitting the model

An additive logistic regression model of the following form will be fitted:

$$log(\frac{\hat{\pi}}{1-\hat{\pi}}) = \beta_0 + \beta_{ethnicity}X_{ethnicity} + \beta_{gender}X_{gender} + \beta_{jaundice}X_{jaundice}$$

Given that the condition we established for the "high probability" (i.e. total score greater than 6) allows the logistic model to perfectly separate the response variable in the data into "high" and "low" probabiities (thus making the respective MLE for β_j go to infinity), a penalized logistic regression model (LASSO) was employed (Marin, 2021).

term	estimate	std.error	statistic	p.value
(Intercept)	-1.783	0.290	-6.154	0.000
ethnicityBlack	1.542	0.411	3.748	0.000

term	estimate	std.error	statistic	p.value
ethnicityHispanic	1.507	0.635	2.371	0.018
ethnicityLatino	1.832	0.527	3.474	0.001
ethnicityMiddle Eastern	-0.482	0.459	-1.051	0.293
ethnicityOthers	0.998	0.480	2.079	0.038
ethnicityPasifika	-0.539	1.080	-0.499	0.618
ethnicitySouth Asian	-0.494	0.662	-0.747	0.455
ethnicityTurkish	0.361	1.130	0.320	0.749
ethnicityWhite-European	1.702	0.300	5.665	0.000
genderm	-0.291	0.195	-1.489	0.137
jundiceyes	0.640	0.305	2.099	0.036

Then to consider adding the behavior question Q8 or not, we first fit as following

$$log(\frac{\hat{\pi}}{1-\hat{\pi}}) = \beta_0 + \beta_{ethnicity}X_{ethnicity} + \beta_{gender}X_{gender} + \beta_{jaundice}X_{jaundice} + \beta_{Q8}X_{Q8}$$

term	estimate	std.error	statistic	p.value
(Intercept)	-2.989	0.366	-8.174	0.000
ethnicityBlack	1.808	0.435	4.153	0.000
ethnicityHispanic	1.430	0.644	2.221	0.026
ethnicityLatino	1.573	0.536	2.934	0.003
ethnicityMiddle Eastern	-0.381	0.466	-0.818	0.414
ethnicityOthers	0.959	0.495	1.936	0.053
ethnicityPasifika	-0.456	1.099	-0.415	0.678
ethnicitySouth Asian	-0.504	0.671	-0.751	0.453
ethnicityTurkish	0.197	1.145	0.172	0.864
ethnicityWhite-European	1.872	0.311	6.017	0.000
genderm	-0.382	0.206	-1.855	0.064
jundiceyes	0.735	0.321	2.288	0.022
i_can_go_back_to_work_when_interrupted	1.574	0.251	6.260	0.000

```
              df      AIC
model_one_fit 12 662.4442
model_two_fit 13 618.1985

              df      BIC
model_one_fit 12 715.3860
model_two_fit 13 675.5521
```

To compare the two models, where model one does not include the variable `i_can_go_back_to_work_when_inter` and model two does, both AIC and BIC of model two are smaller than the values of model one. Therefore, we choose model two.

The intercept means that for a test taker who is an Asian woman, was not born with jaundice, indicate on the questionnaire that they can go back to work after being interrupted, the odds of such a person to have a high probability of autism are expected to be 0.05 (calculated from (exp(-2.989))).

The coefficients of ethnicity are compared to the baseline group, which is Asian. For instance, for the coefficient of `ethnicityBlack`, it means that the odds of someone identified as Black to have a high probability of autism are expected to be 6.1 (calculated from exp(1.808)) times the odds of someone identified as Asian, holding all else constant.

The odds of a man to have a high probability of autism are expected to be 0.68 (calculated from exp(-0.382)) times the odds of a woman, holding all else constant.

The odds of someone who was born with jaundice to have a high probability of autism are expected to be 2.085(calculated from exp(0.735)) times the odds of someone who was not born with jaundice, holding all else constant.

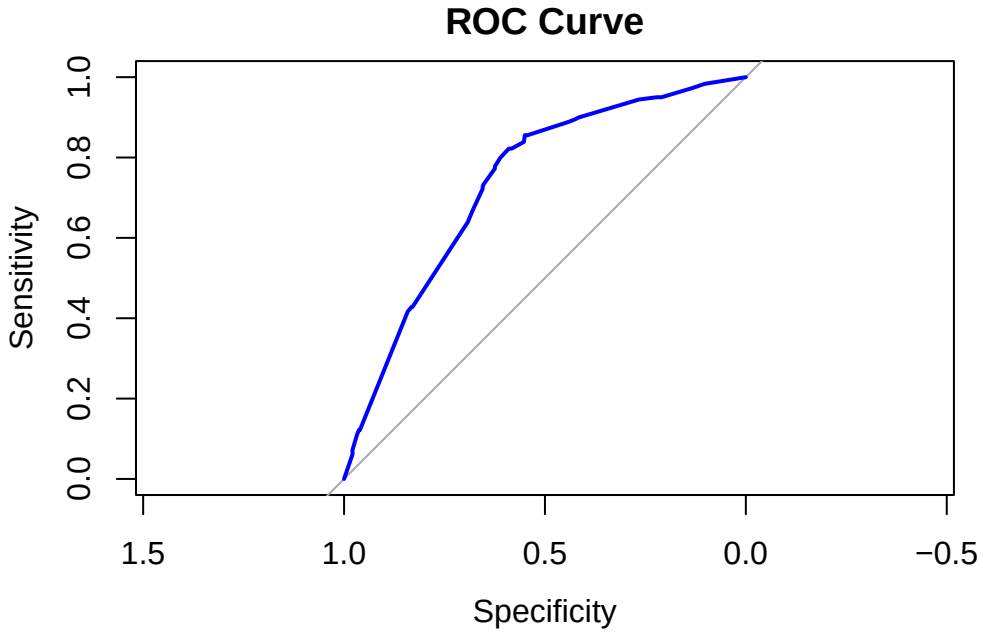
The odds of someone who indicates that they can go back to work after being interrupted to have a high probability of autism are expected to be 4.83 (calculated from exp(1.574)) times the odds of someone who indicates that they cannot go back to work after being interrupted, holding all else constant.

term	step	estimate	lambda	dev.ratio
(Intercept)	1	-1.642	0.008	0.133
ethnicityBlack	1	1.253	0.008	0.133
ethnicityHispanic	1	1.074	0.008	0.133
ethnicityLatino	1	1.490	0.008	0.133
ethnicityMiddle Eastern	1	-0.449	0.008	0.133
ethnicityOthers	1	0.653	0.008	0.133
ethnicityPasifika	1	-0.153	0.008	0.133
ethnicitySouth Asian	1	-0.327	0.008	0.133
ethnicityWhite-European	1	1.494	0.008	0.133
genderm	1	-0.201	0.008	0.133
jundiceyes	1	0.534	0.008	0.133

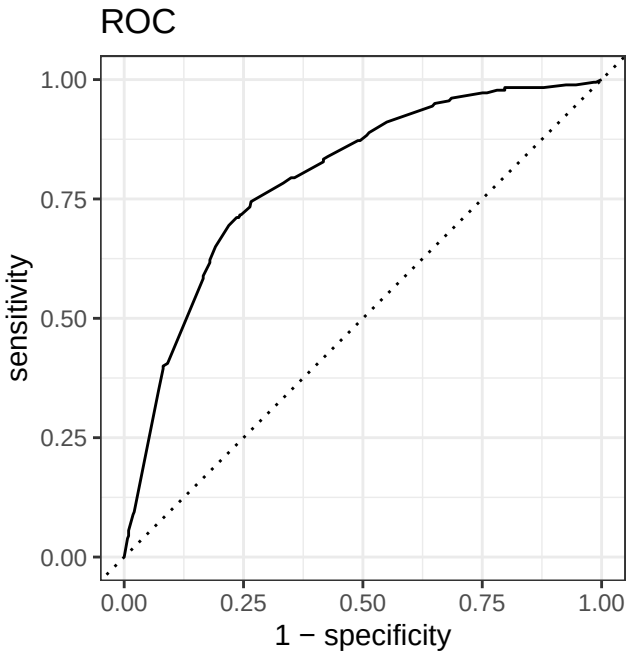
Considering the closely related questions regarding sensory and behavioral tendencies, multicollinearity is a major concern. Subsequently, a VIF test was conducted as a means of evaluating multicollinearity.

```
                                        GVIF Df GVIF^(1/(2*Df))
ethnicity                           1.092127  9        1.004908
gender                              1.035873  1        1.017779
jundice                             1.021815  1        1.010849
i_can_go_back_to_work_when_interrupted 1.067548  1        1.033222
```

This model was then employed to augment the original data frame and obtain the estimated probabilities for each observation, these would later be employed for an ROC curve to determine the optimal prediction threshold.




```
Area under the curve: 0.7363
```




```
# A tibble: 1 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 roc_auc binary         0.795
```

The area under the curve is 0.795, which for a model fitted to demographic data independent of questions on the quiz, indicates that further investigation into how these questions may over- or under-diagnose certain populations would be worth pursuing.

References

Aishworiya, R., Kim, V., MA, Stewart, S., Hagerman, R., & Feldman, H. M. (2023). Meta-analysis of the Modified Checklist for Autism in Toddlers, Revised/Follow-up for Screening. *PEDIATRICS*, *151*(6). https://doi.org/10.1542/peds.2022-059393

Curnow, E., Utley, I., Rutherford, M., Johnston, L., & Maciver, D. (2023). Diagnostic assessment of autism in adults – current considerations in neurodevelopmentally informed professional learning with reference to ADOS-2. *Frontiers in Psychiatry*, *14*. https://doi.org/10.3389/fpsyt.2023.1258204

Hirota, T., & King, B. H. (2023). Autism spectrum Disorder. *JAMA*, *329*(2), 157. https://doi.org/10.1001/jama.2022.23661

Marin, Z. (2021, April 26). GLM fit: Algorithm did not converge – How to fix it. Statology. https://www.statology.org/glm-fit-algorithm-did-not-converge/ ::: callout-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. :::