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
```

## Histogram of Scoring Result
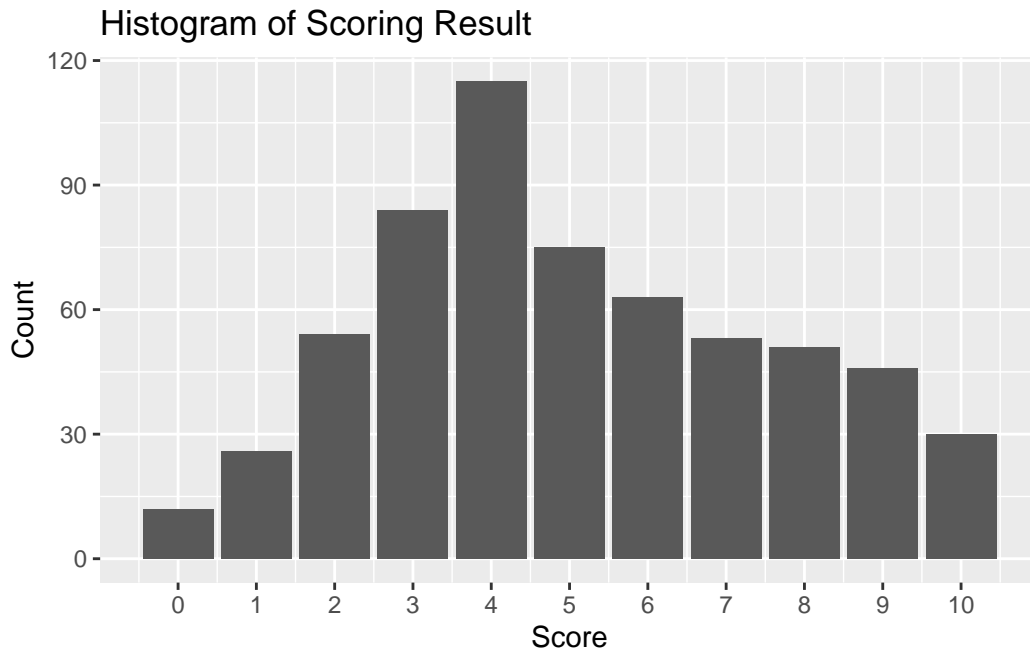

```
[1] 0.2955665
```

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.

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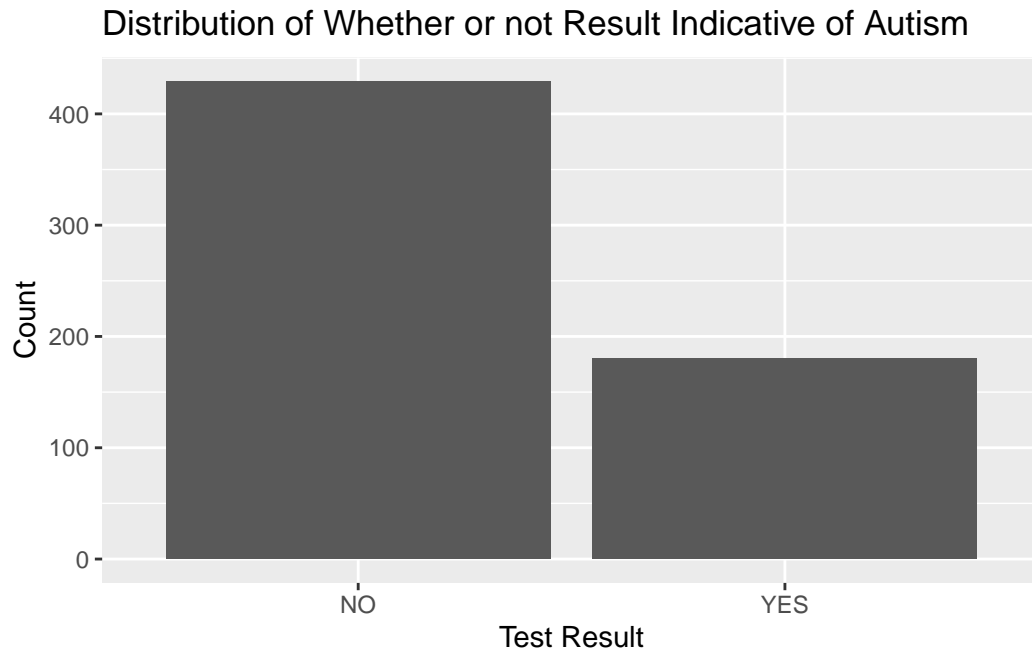
Additionally, here is a visual demonstration of the number of test takers receiving further encouragement to pursue a professional diagnosis of ASD. Though still significantly less than those not encouraged, a notable proportion of test takers were informed they had traits or behaviors indicative of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

## Distribution of Ethnicity in the Dataset
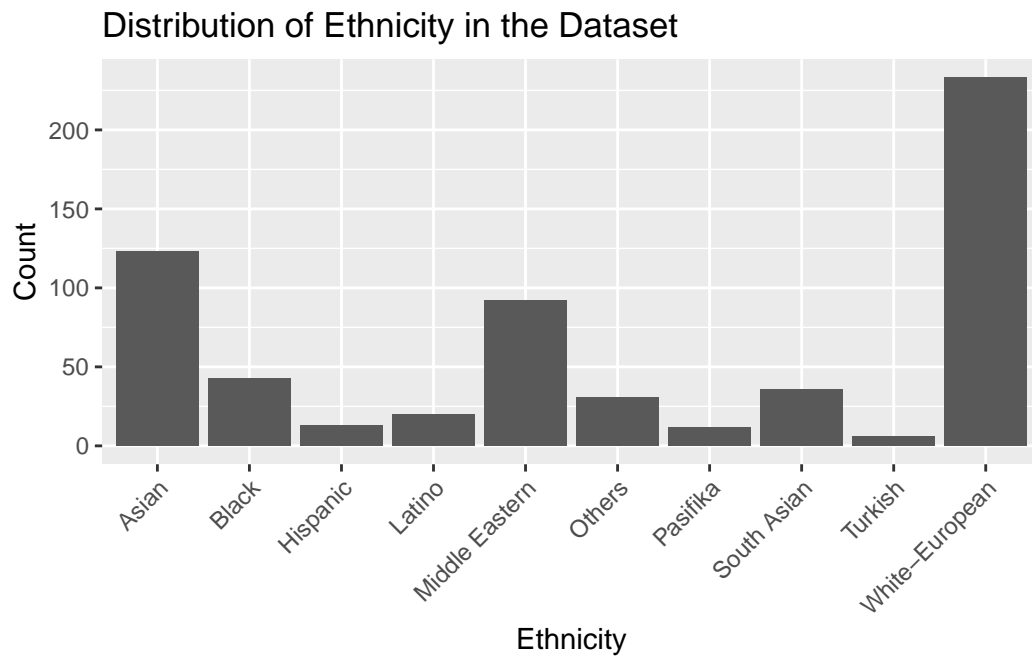
We observe here that the most common ethnicity was White-Europeans, with over 200 observations in our data set. Secondly and thirdly, we observe the Asian and Middle Eastern populations to be above 100 and slightly below 100 observations, respectively. All other ethnicities have less than 50 observations 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:


Proportion of Indicative Test Scores
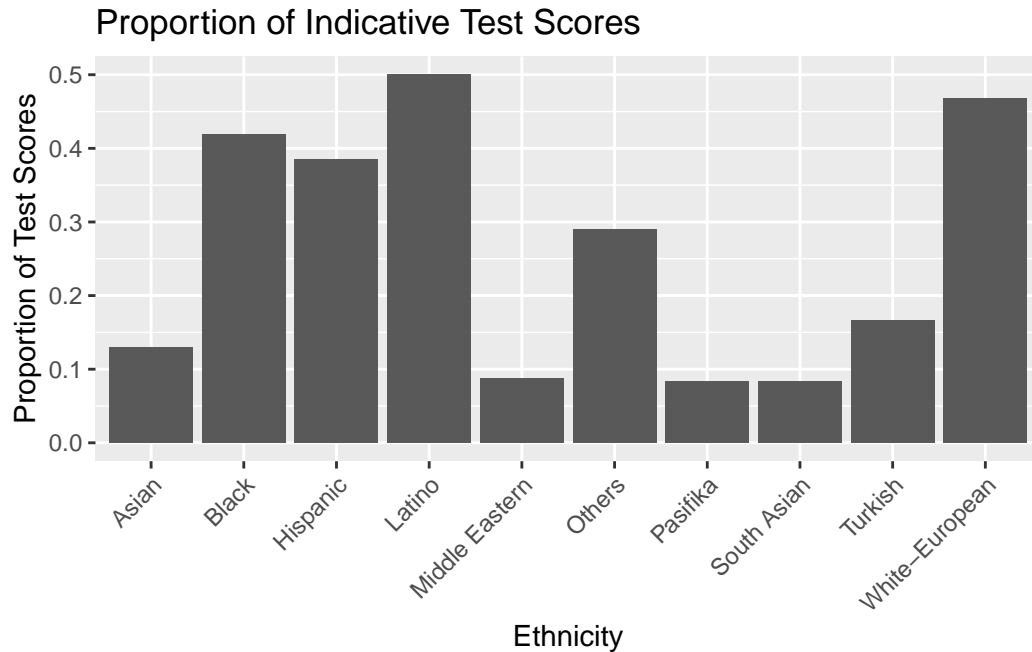
As we can see here, the population with the highest proportion of test takers receiving a high test score was the Latino population, with close to half achieving higher than a 6. Since this is a comparably uncommon population in our data set, it does raise questions of whether it is indicative of the 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.

Distribution of Age in the Dataset
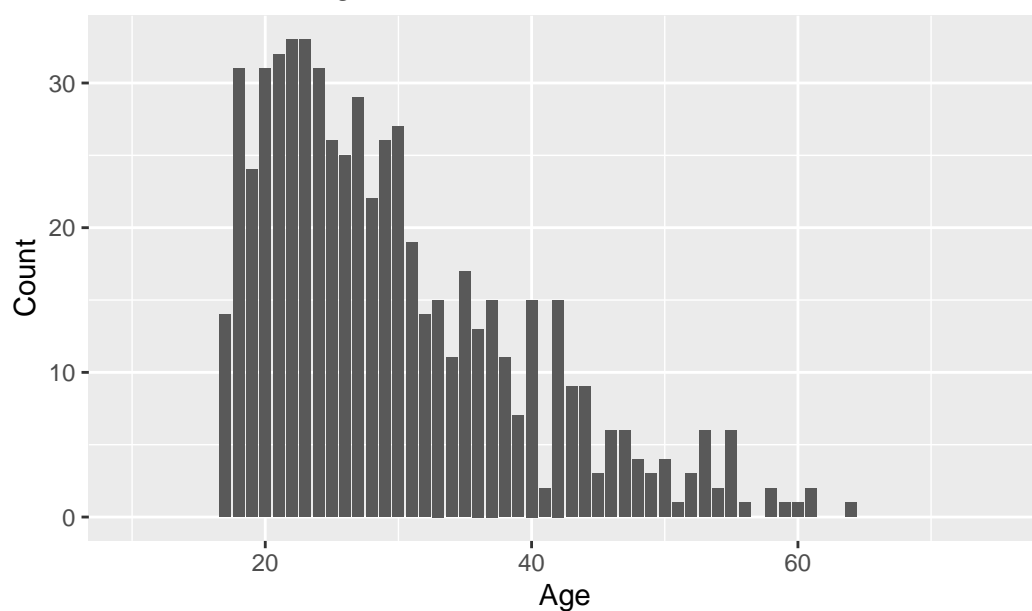

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
17.00   22.00   27.00   30.22   35.00  383.00
```

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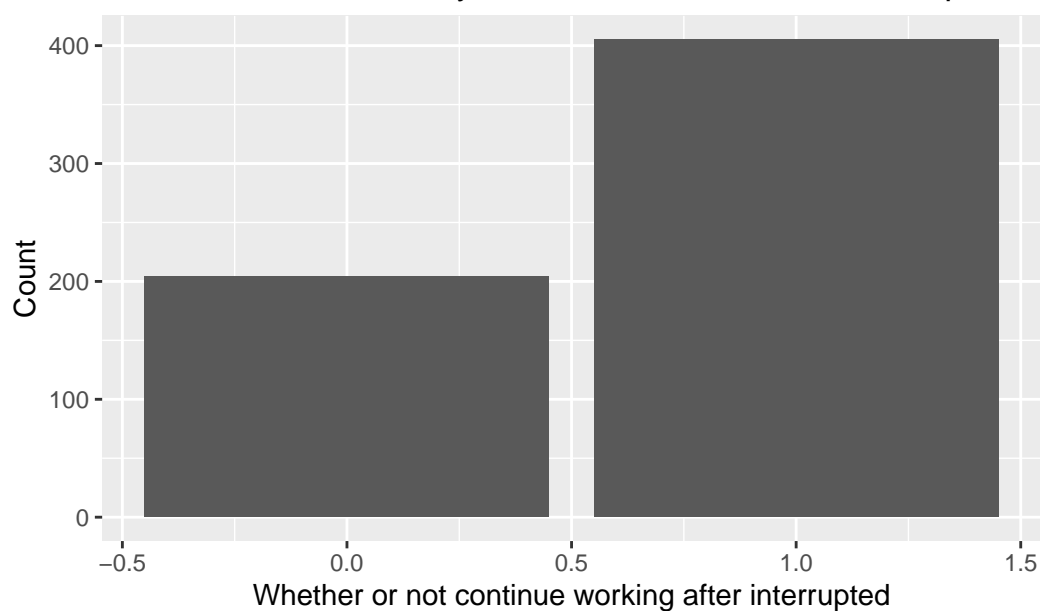
Generally, we observe the distribution of ages to possess a strong right-skewness, with a mean age of 30.22 and a median age of 27. There is a clear peak in the distribution roughly around early to mid 20's years of age. The ages possess a minimum of 17 up to a maximum age of 383, which is clearly a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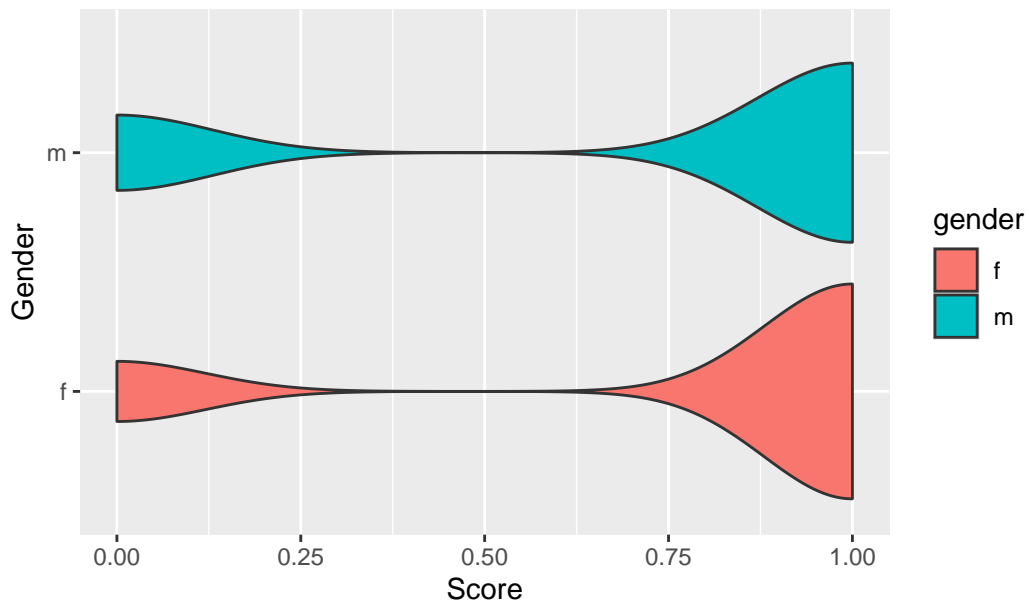
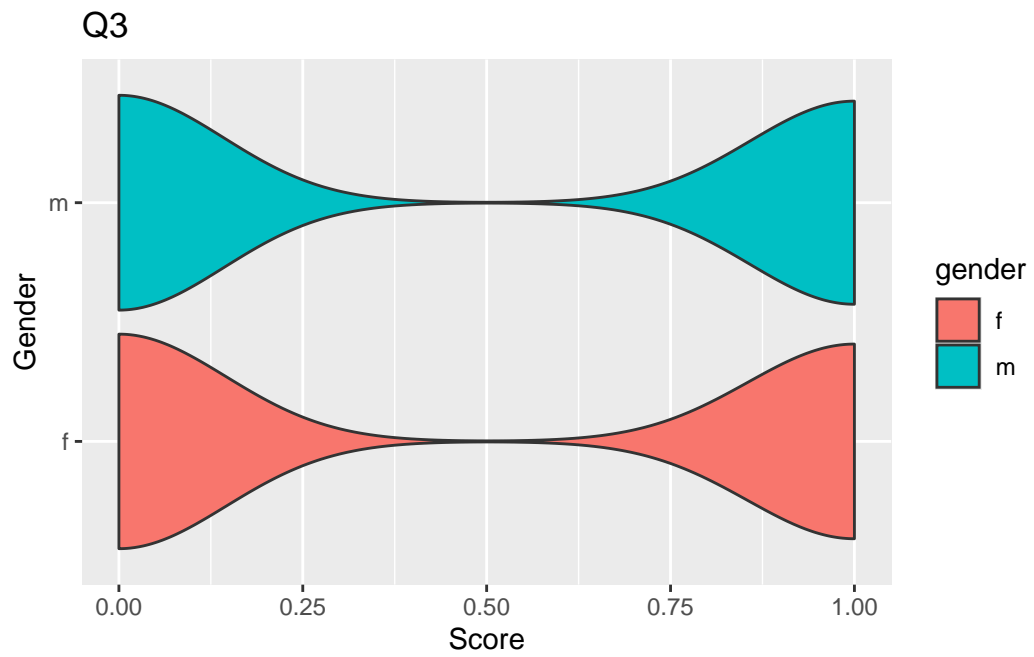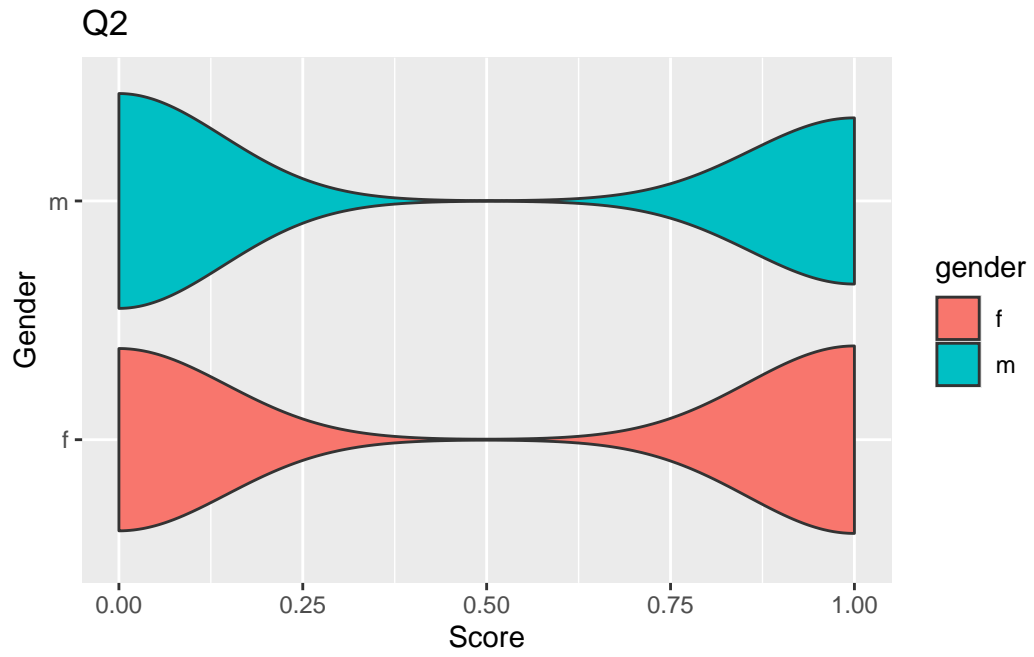**Bivariate EDA**

Distribution of Scores Across Ethnicity
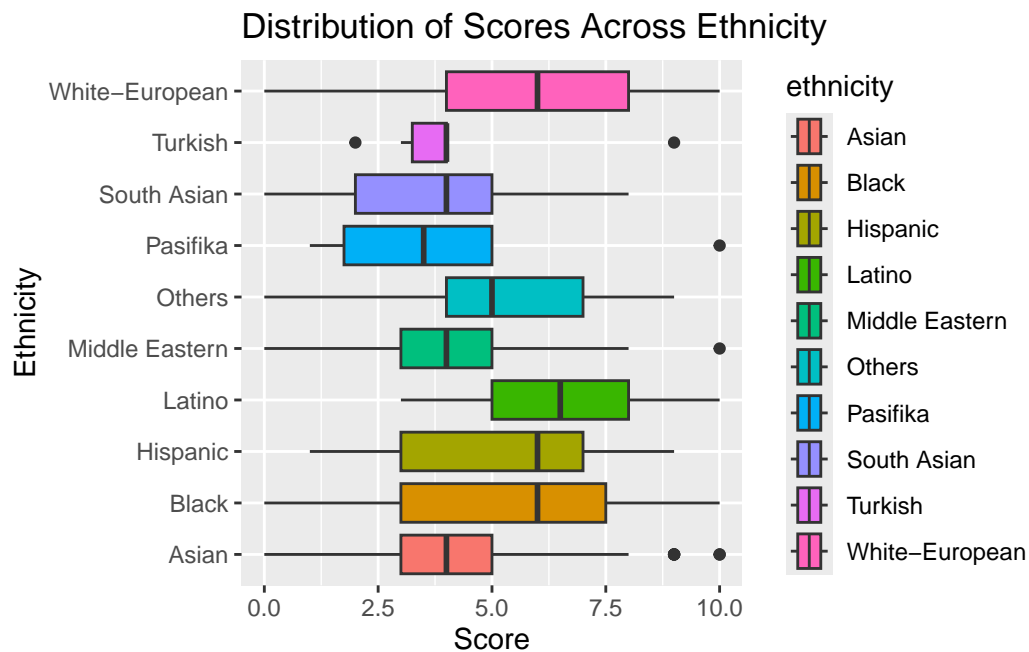
Distribution of Scores Across Person Filling Test

Q1

Q6

Q7

Q10

Points in Test vs. Age
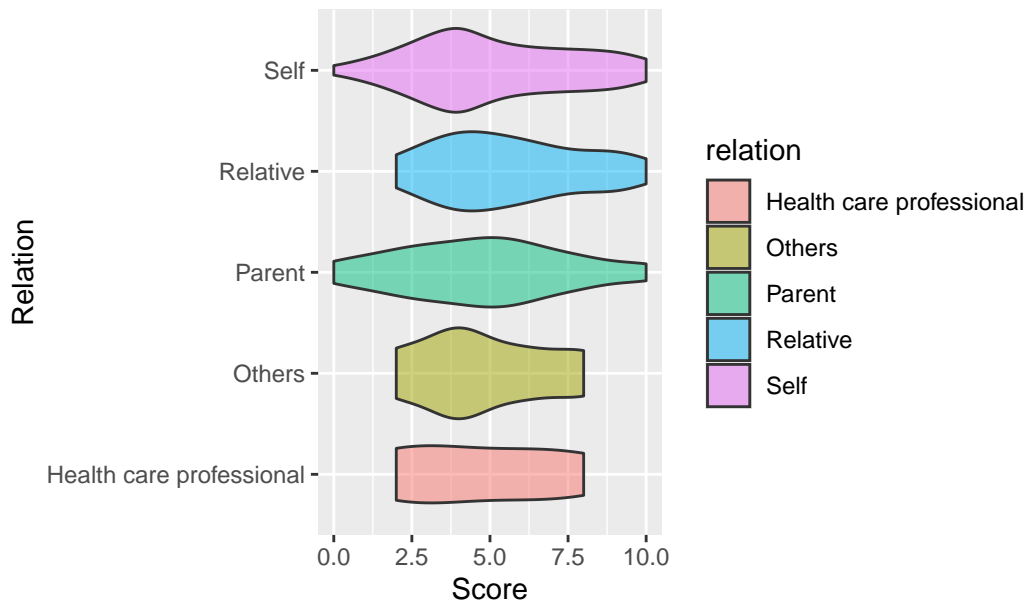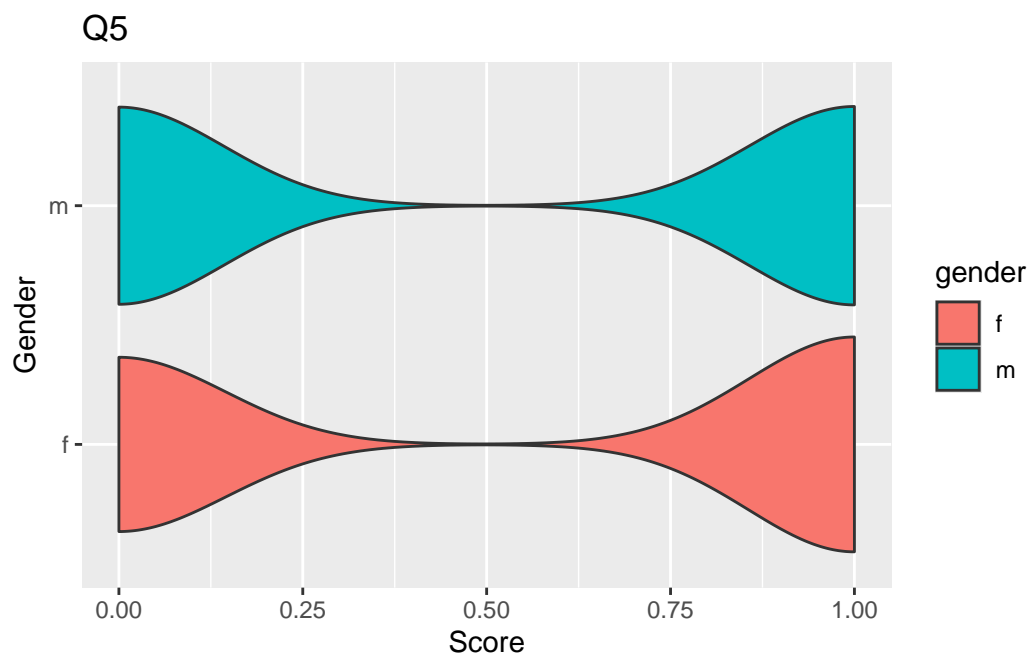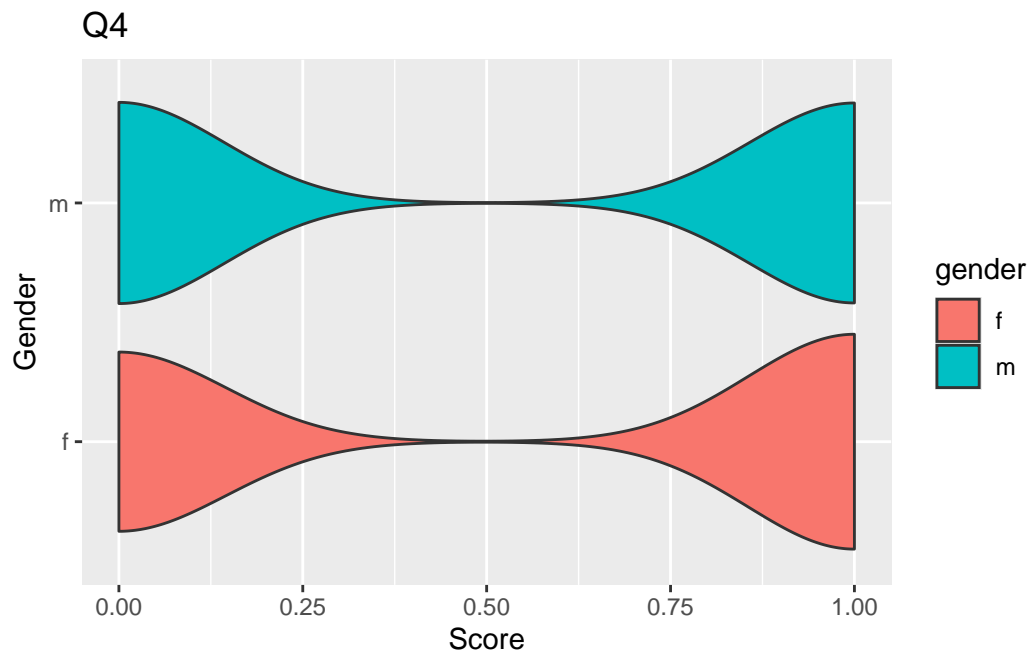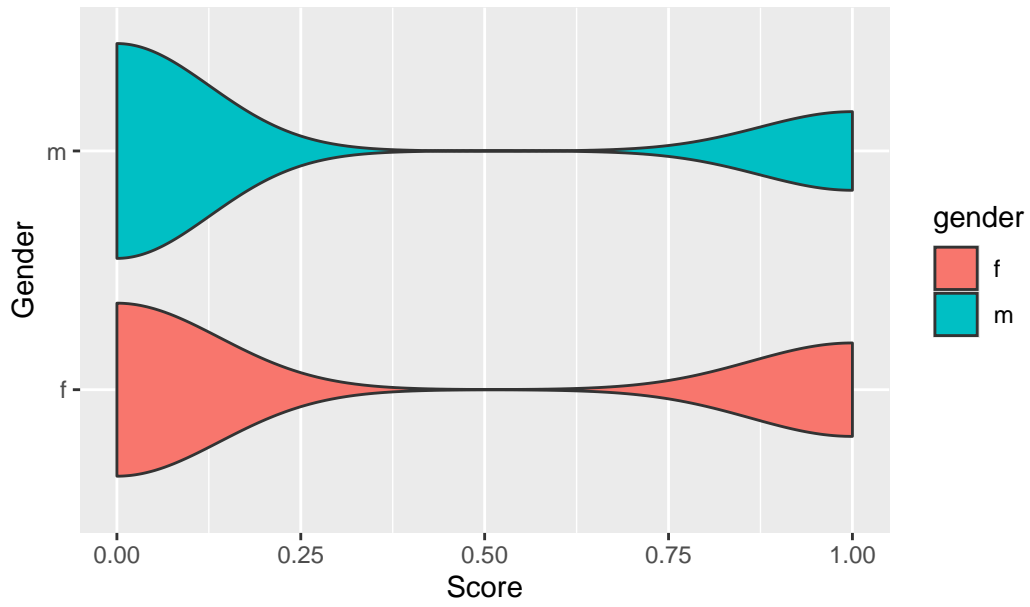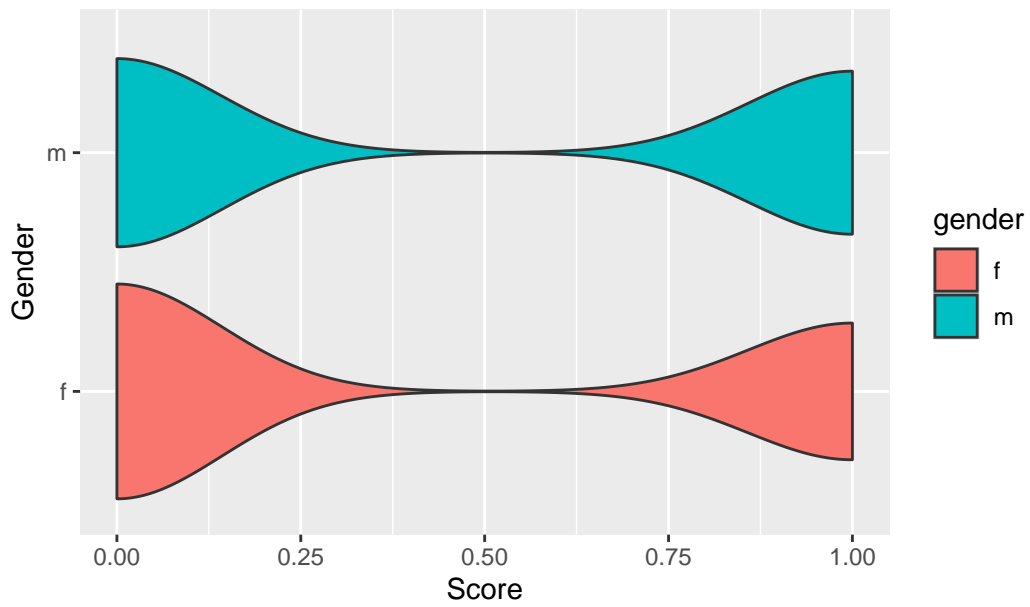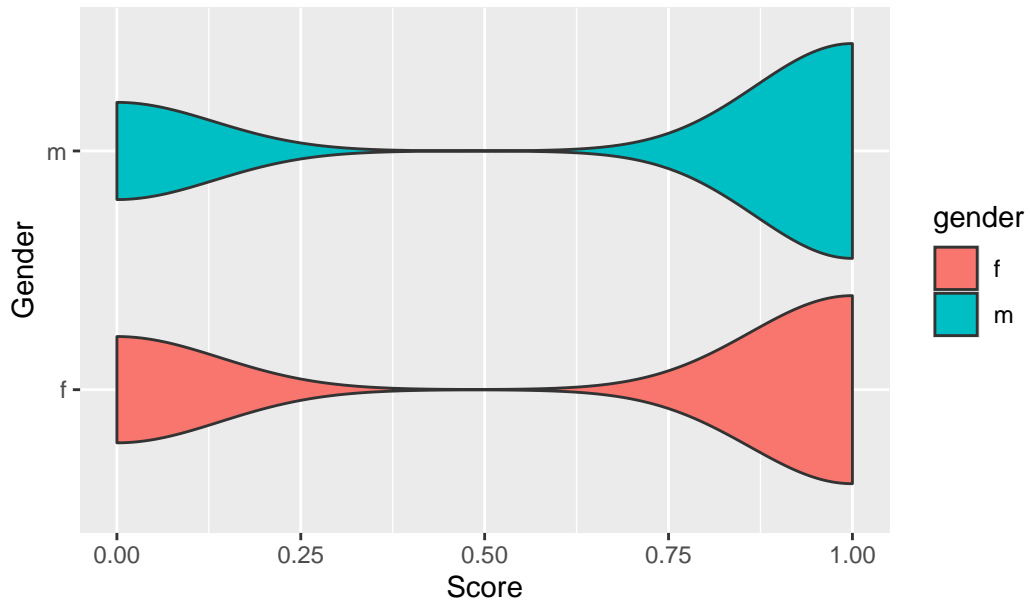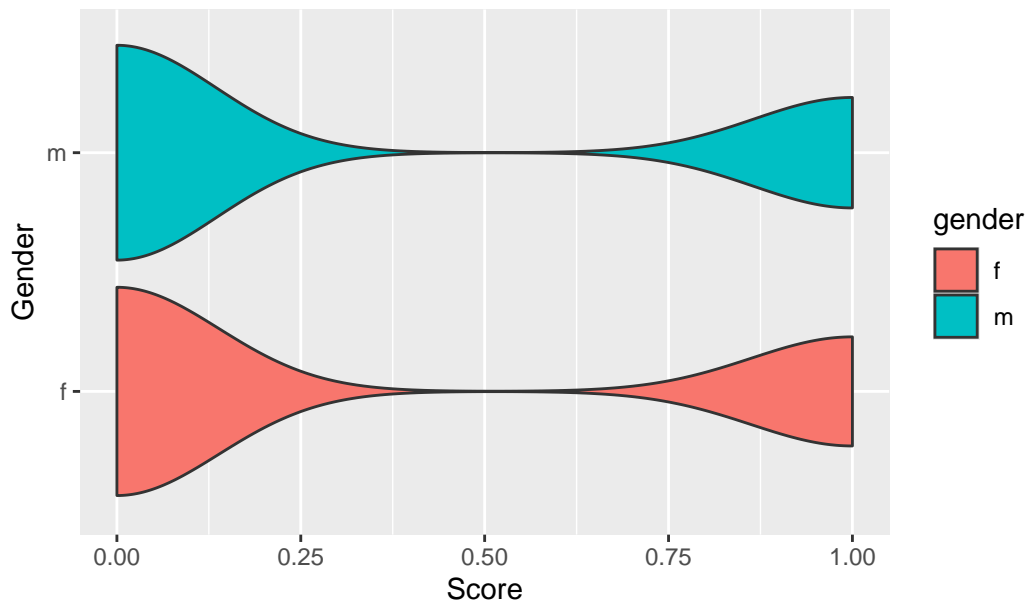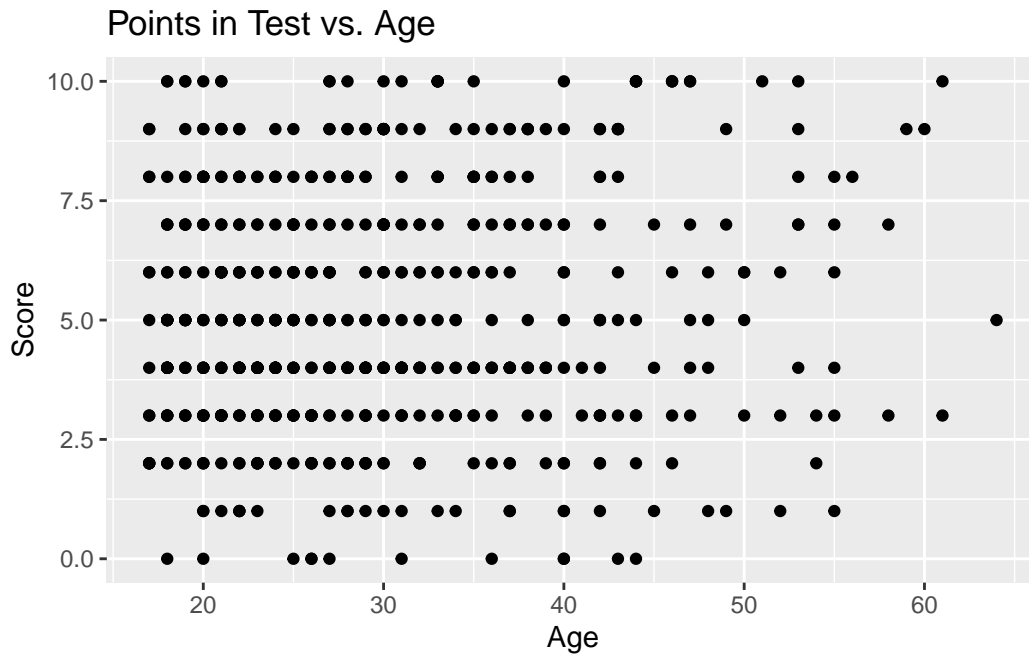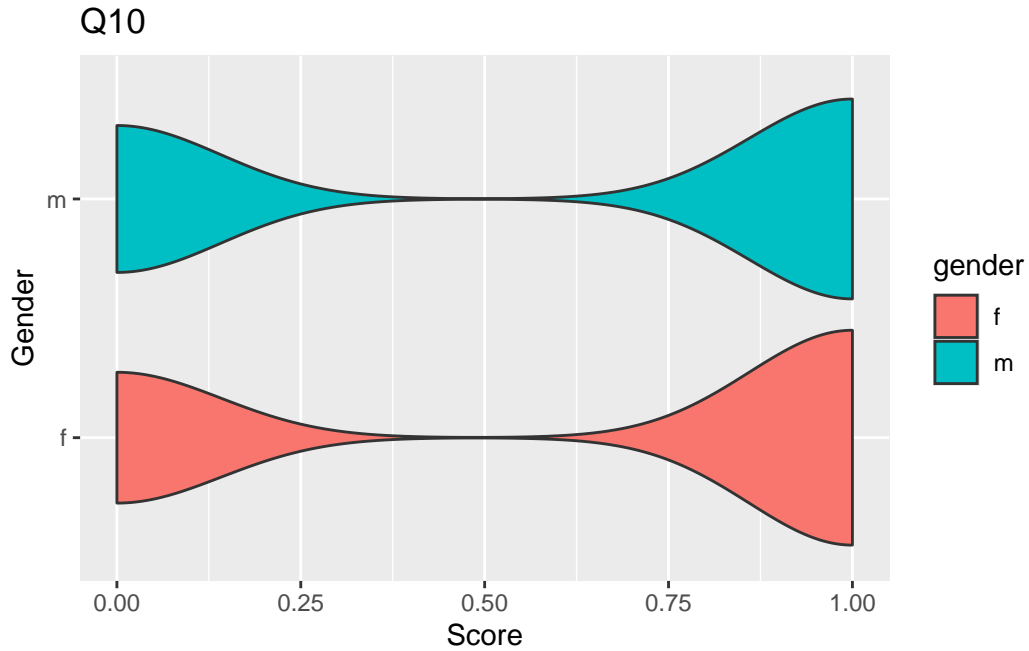
Here we present several graphs displaying bivariate relationships. With our first set of box plots, we find further evidence for our previous suspicion that test score distributions differ by ethnicity. Although most ethnicities have a median within the range of roughly 3 to 6, some ethnicities, like 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 all predictor variables as a means of assessing which predictors provide a statistically significant improvemnet in model fit against the null condition. The hypothesis for this test can be observed below:

$$H_0 : \beta_1 = 0$$

$$H_a : \beta_1 \neq 0$$

- 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
Q1	62.96	2.11x10^-15
Q2	62.96	2.11x10^-15
Q3	69.97	6.04x10^-17
Q4	121.68	2.70x10^-28
Q5	148.94	2.95x10^-34
Q6	217.14	3.81x10^-49
Q7	219.73	1.03x10^-49
Q8	77.51	1.32x10^-18
Q9	253.13	5.39x10^-57
Q10	97.29	5.99x10^-23
Age	1.54	0.22
Ethnicity	94.1	2.44x10^-16

Predictor	G-stat	p-value
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 Q1-10, ethnicity, gender, and, quite surprisingly, jaundice when compared to a null model as the p-values are either significantly less than 0.05 (Q1-10 + ethnicity) or slighly below it (gender and jaundice). 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.

Fitting the model As said before, an additive logistic regression model of the following form was fitted:

$$log(\frac{\hat{\pi}}{1-\hat{\pi}}) = \beta_0 + \beta_1 X_{Q1} + ... + \beta_{10} X_{Q10} + \beta_{ethnicity} X_{ethnicity} + \beta_{gender} X_{gender} + \beta_{jaundice} X_{jaundice}$$

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	step	estimate	lambda	dev.ratio
(Intercept)	1	-74.964	0	0.999
small_sounds	1	11.673	0	0.999
difficult_to_understand_char	1	11.259	0	0.999
ease_to_read_between_lines	1	11.252	0	0.999
focus_on_whole_picture	1	11.644	0	0.999
i_can_tell_if_someone_bored	1	11.947	0	0.999
i_can_multi_task	1	11.483	0	0.999
i_can_tell_feelings_from_faces	1	11.570	0	0.999
i_can_go_back_to_work_when_interrupted	1	11.509	0	0.999
i_like_to_collect_info_on_categories	1	11.733	0	0.999
i_find_it_hard_to_figure_out_others_intentions	1	11.054	0	0.999
ethnicityBlack	1	0.077	0	0.999
ethnicityHispanic	1	0.028	0	0.999
ethnicityMiddle Eastern	1	-0.196	0	0.999
ethnicityWhite-European	1	0.059	0	0.999
genderm	1	-0.037	0	0.999

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))
small_sounds	1.633894	1	1.278239
difficult_to_understand_char	1.980424	1	1.407275
ease_to_read_between_lines	1.860601	1	1.364039
focus_on_whole_picture	1.868743	1	1.367020
i_can_tell_if_someone_bored	1.521808	1	1.233616
i_can_multi_task	2.087475	1	1.444810
i_can_tell_feelings_from_faces	2.007757	1	1.416954
i_can_go_back_to_work_when_interrupted	1.790655	1	1.338153
i_like_to_collect_info_on_categories	1.989245	1	1.410406
i_find_it_hard_to_figure_out_others_intentions	1.820451	1	1.349241
ethnicity	2.634246	9	1.055285
gender	1.210095	1	1.100043
jundice	1.268408	1	1.126236

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.

ROC Curve
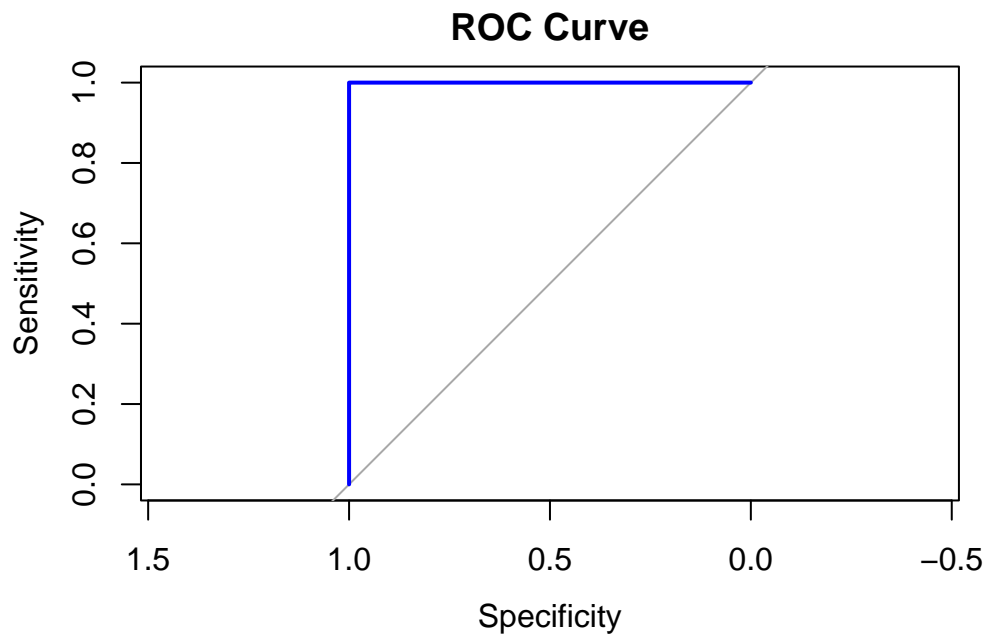

```
Area under the curve: 1
```

The area under the curve is 1, this tells us that the model is able to perfectly fit the data throughout all thresholds. This could very well be due to over fitting, however, we lack

training data as a means of assesing the model. Potential future direction would be to test this penalized model on another data set. ## Results

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. :::