The Repos: STA221 Project

Jeffrey Bohrer, Alexandra Green,
Anna Zhang, Kevin Lee

Index

Topic, Motivation & Research Question

- Autism Spectrum Disorder (ASD) is a highly prevalent condition–2.2% of adults are affected by it and there is growing awareness (Hirota, 2023).
- Most screening tests are inaccurate–lots of false negatives and lack of predictive value (Aishworiya, 2023; Curnow, 2023).

Can we generate a better diagnostic model?

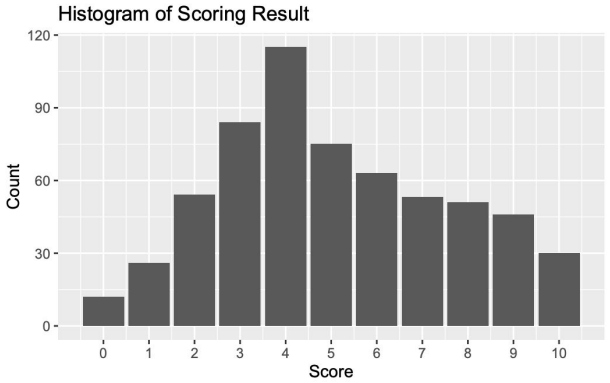
Dataset

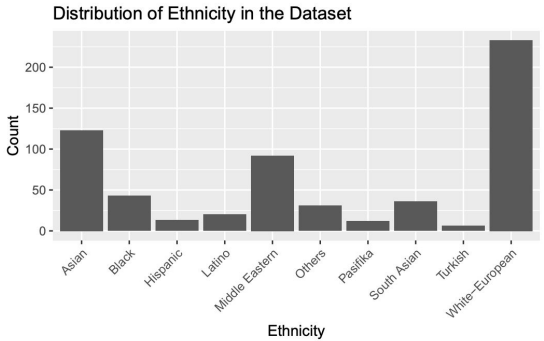
- UCI Machine Learning Repository (2017) Professor Fadi Fayez Thabtah from the Manukau Institute of Technology in New Zealand
    - Sourced from his app **ASDTests** – screens for Autism using 10 questions

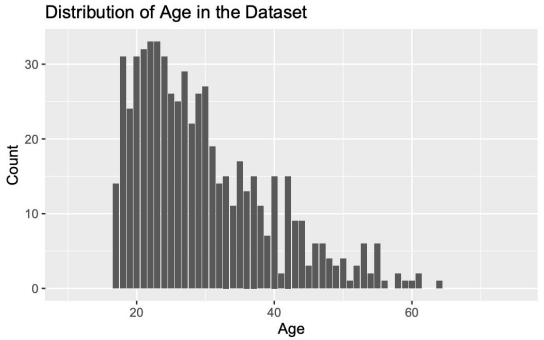
---

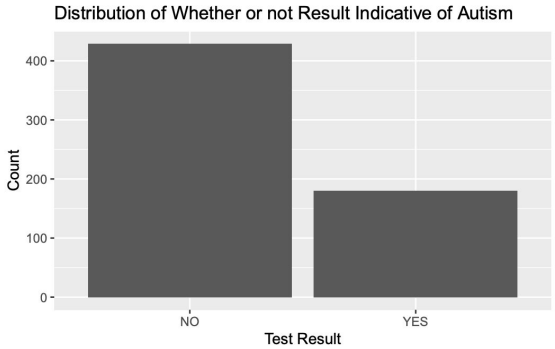
- 704 observations and 20 features
    - 10 features are questions
    - 10 are individual characteristics
- Presence of autism is generally categorized at a six or higher

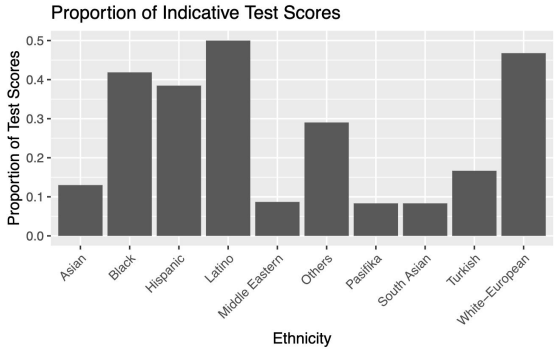
Univariate EDA

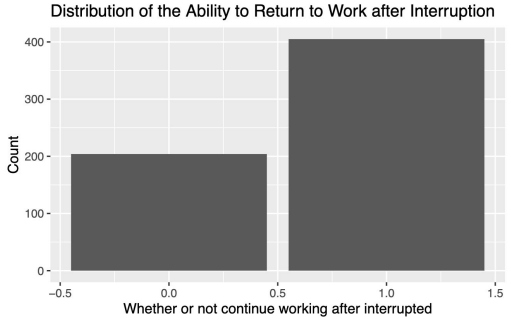




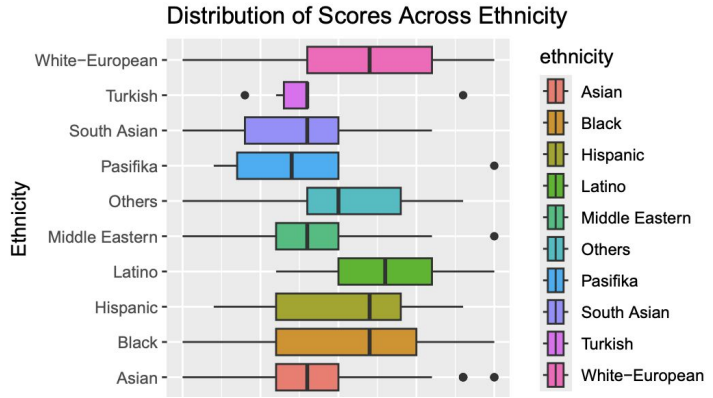


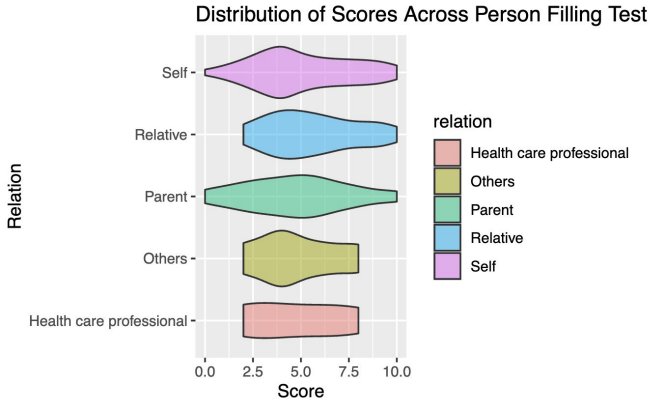


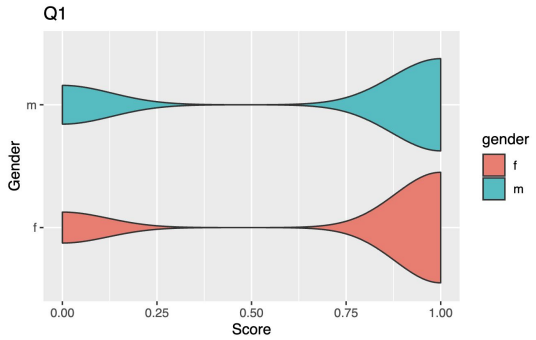


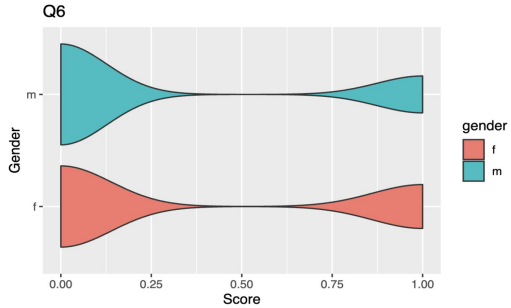


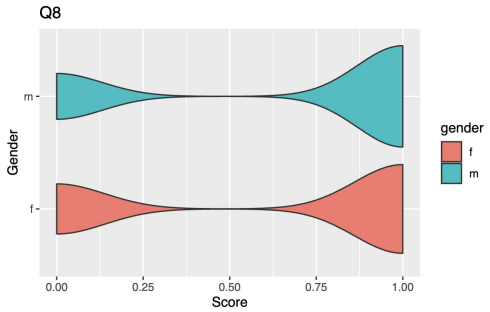
Bivariate EDA











Next Steps:

- Can we detect any differences in the answer distributions for different age groups?

- Further investigation of relationship of the test taker to the subject of the quiz and the subject's score – how does this impact answers to specific questions, and also how does the ethnicity or gender of the subject affect these distributions?

- Distributions of answers across all questions for subjects of different ethnicities – are there specific questions on the test that leads to these disparities between score distributions?