Analysis the Prevalence Alzheimer's Disease in Patient Demographics

By Victoria Hernandez

Abstract

Currently Alzheimer's disease is ranked as the seventh leading cause of death in the United States and the purpose of the analysis is to find potential predictors can help assist with the early detection of the disease. (NIH National Institute on Aging, 2023) The dataset analyzed in this report dataset contains health records of various types of patients. The dataset includes patient attributes such as demographical details, lifestyle factors, and medical history.

Introduction

The dataset analyzed in this report contains patient data including an individual's age, gender, MMSE, education level, family medical history. Alzheimer's disease is a brain disorder that affects someone's memory and other cognitive abilities. (NIH National Institute on Aging, 2023) The disease affects more than 6 million Americans, mostly older that 65 years of age. (NIH National Institute on Aging, 2023) Understanding the possible contributing factors early are important for management and treatment of this disease. The analysis of this dataset can help reveal possible health correlations which can be used as predictors for the disease. During this study we utilize statistical analysis and various datamining techniques. The next sections of this report will describe the data collected, how the data was prepared, methods, results, and a conclusion on the implications of the results of the study.

Data and Data Preparation

The name of the dataset being analyzed is "Alzheimer's' Disease Dataset". (Kharoua, 2024) The dataset was sourced from Kaggle, and the data was collected by Rabie El Kharoua. For this study we will be analyzing 6 different patient attributes.

Table 1. Attribute Descriptions

Attribute	**Attribute Description**	**Data Type**	Range
Age	Age of patient	Numeric	60 to 90 years
Education Level	Patient education level.	Numeric	1 to 5 (1 = None, 5 = Higher)
FamilyHistoryAlzheimer	Patient's family history of Alzheimer's.	Categorical	0 = No 1 = Yes

MMSE	Mini-Mental State Examination score, measures cognitive function	Numeric	0 to 30
Gender	Patient's gender.	Categorical	0 = Male 1 = Female
Diagnosis	Alzheimer's Disease diagnosis.	Categorical	0 = No 1 = Yes

Methods

Exploratory data Analysis Techniques

The descriptive statistics used in this analysis were measures of central tendency, dispersion, and skewness (Zumel et al., 2020). Central tendency was measured by finding the mean, median and mode of the data (Zumel et al., 2020). The visualization methods used for numeric variables were histograms, box plots, scatter pots, and overlaid density plots (Larose & Larose, 2017). The visualization methods used for categorical variables were stacked bar plots, side-by-side bar plots, and bar plots (Larose & Larose, 2017).

Results

Individual attribute distribution analysis:

Age: Table 2 below shows that the average age for the patients was 74.91 years based on the mean of the data. Table 2 also shows that most of the patients are about 8.99 years within of the average age based on the standard deviation. From table 2 the skewness is -0.03 and the histogram in figure 2 below shows the data is slightly skewed to the left. This means that most patient studied are older.

Education level: Table 2 below show that the average education level of the patients is ~1 or high school. Table 2 below also shows that most patient are about 0.82 within the average education level based on the standard deviation. The skewness is 0.95 in table 2 below and the histogram in figure 2 below shows skewness to the right meaning most patients have a lower education level.

Gender: Table 3 and the bar chart in figure 3 below shows that the most common gender amongst the studied patients is female.

MMSE: Table 2 below shows that the average MMSE score is 14.76. Table 2 also shows that the most patient are about 8.61 within the average score based on the standard deviation.

From table 2 below the skewness is 0.11 and the box plot chart in figure 4 shows below shows a slight skew to the left meaning more patients have a higher score.

Diagnosis: Table 3 and the bar chart in figure 5 below shows that the most common diagnosis amongst the studied patient is for a patient not to be diagnosed with Alzheimer's.

Family Alzheimer's History: Table 3 and the bar chart in figure 6 below shows that the most common answer amongst patient about if any of the family members have been diagnosed with Alzheimer's is no.

Table 2. Numeric Attribute Descriptive statistics

Attribute	Mean	Median	Maximum	Standard Deviation	Variance	Skewness	Minimum
Age	**74.91**	**75**	**90**	**~8.99**	**~80.82**	**-0.03**	**60**
Education Level	**1.29**	**1**	**3**	**0.90**	**0.82**	**0.95**	**0**
MMSE	**14.76**	**14.44**	**29.99**	**8.61**	**74.19**	**0.11**	**0.01**

Table 3. Categorical Attribute Descriptive statistics

Attribute	Mode
Gender	**1(Female)**
Family History Alzheimer's	**0(No family history)**
Diagnosis	**0(Not diagnosed)**

Figure 1. Patient Age Distribution

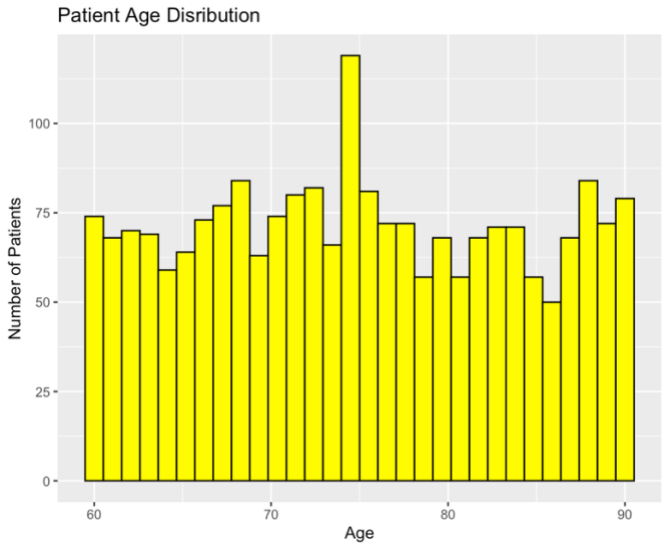


Figure 2. Patient Education Level

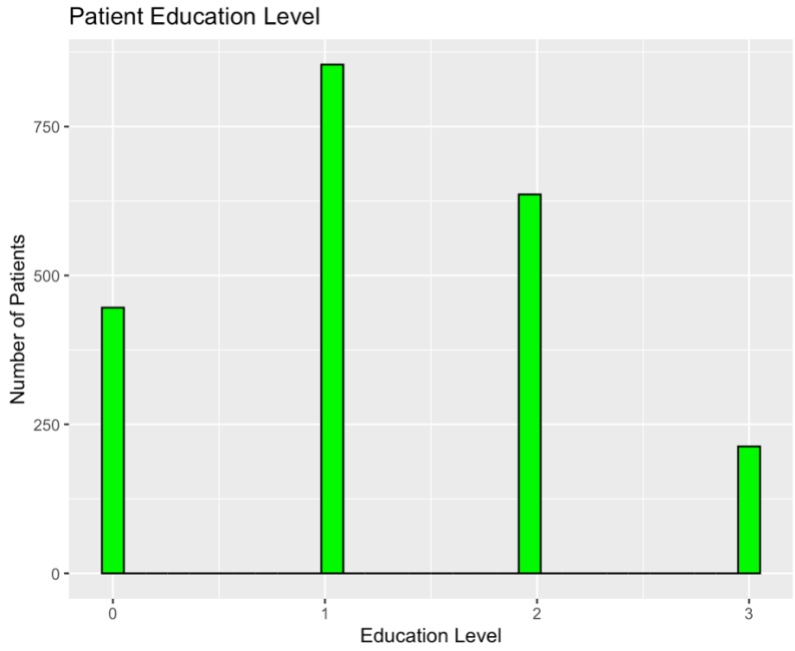


Figure 3. Patient gender distribution

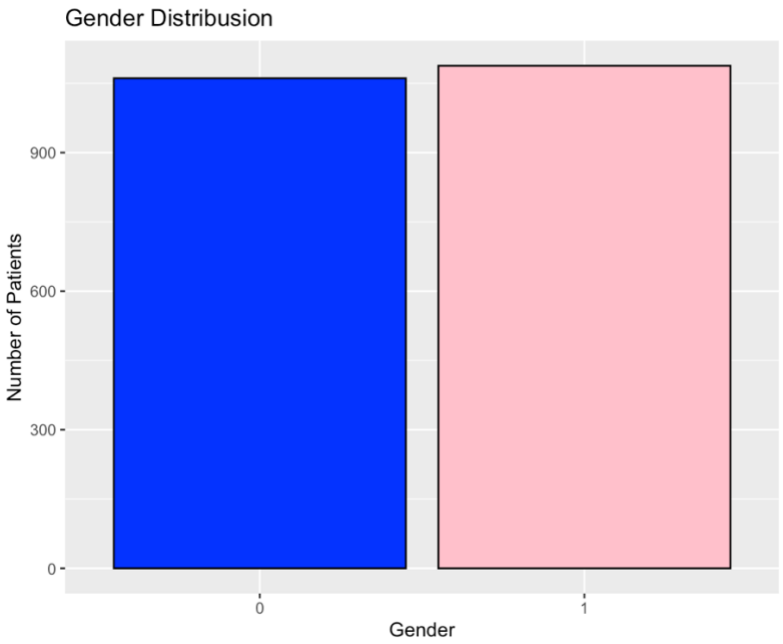


Figure 4. Patient MMSE Scores

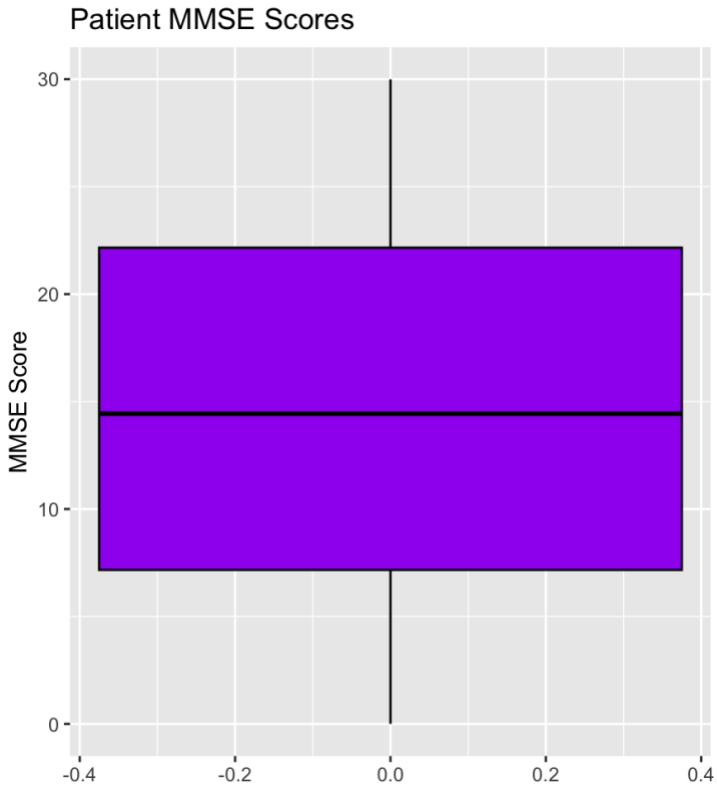


Figure 5. Patient Alzheimer's Diagnosis

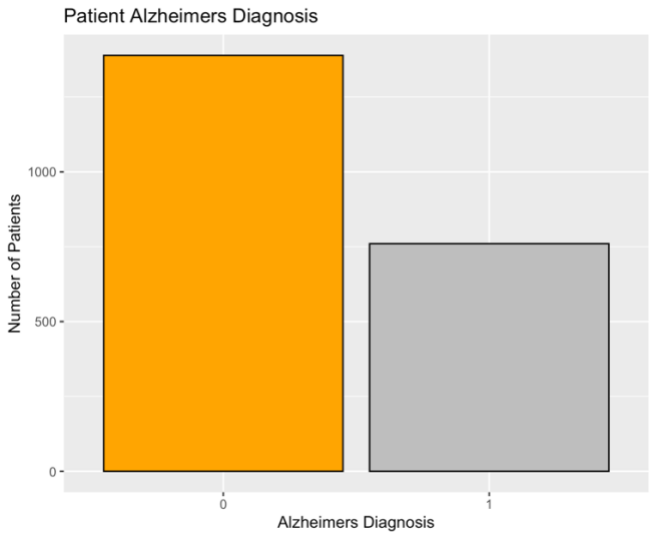
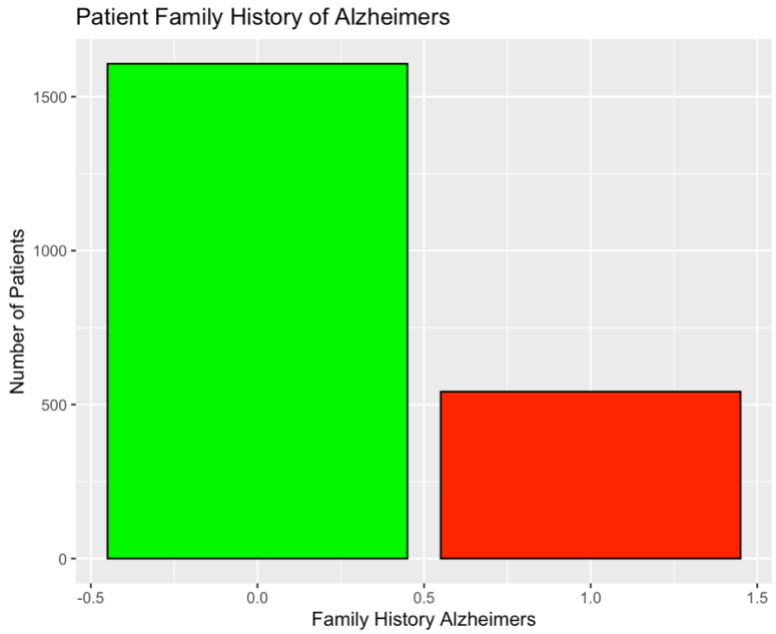


Figure 6. Patients Family history of Alzheimer's



Relationship between attributes analysis:

Family History and Diagnosis: The side-by-side bar chart in figure 7 below show the relationship between family history and diagnosis. The chart shows that patients without a

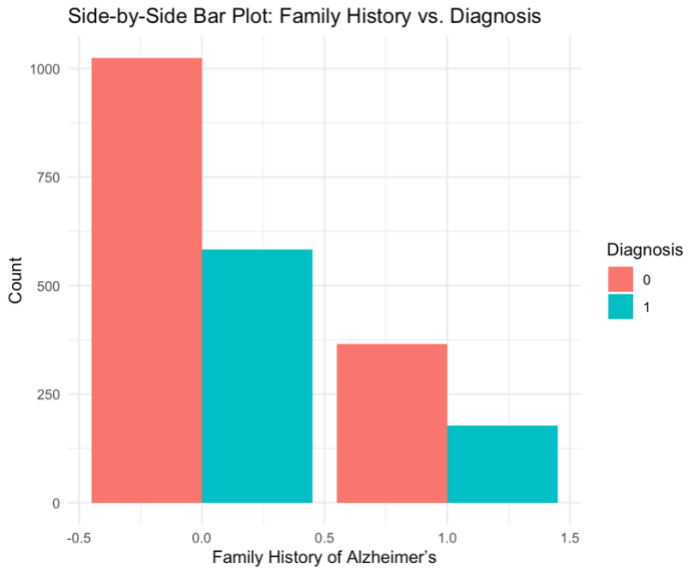
family history of Alzheimer's are more like to have develop Alzheimer's that patient that do have a family history of Alzheimer's. The chart also shows that patients without a family history of Alzheimer's are more likely to not develop Alzheimer's.

Gender and Diagnosis: The stacked bar chart in figure 8 below shows the relationship between gender and diagnosis. A higher number of females were diagnosed as not having Alzheimer's while males are more likely to be diagnosed with Alzheimer's.

MMSE and AGE: The overlaid density plot in figure 9 below shows the correlation between MMSE and AGE. For the 60- to 70-year-old age group the curve is peaks at about a MMSE of 28. For the 70-to-80-year-old age group the curve peaks at about an MMSE score of 25. For the 80 and older age group the curve peaks at about an MMSE score of 20. These results suggest that the as age increases patients cognitive function decreases.

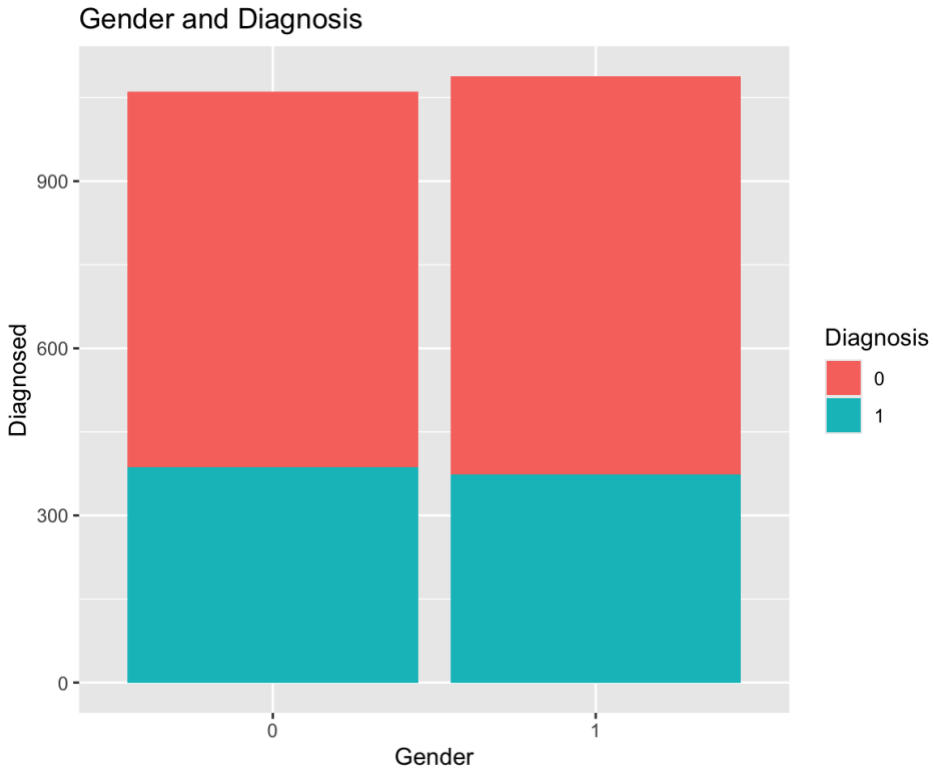
Education Level and Diagnosis: The stacked bar chart in figure 10 below shows the show the correlation between education level and diagnosis. The chart shows that patient with a lower education level of none or high school are more likely to have a Alzheimer's diagnosis than patient with a higher education level of bachelors or higher.

Figure 7. Side-by-side bar chart showing relationship between Family History and Diagnosis



Note: The pink bars are patients without diagnosis and the blue bars are patient with a diagnosis. The left side is without a family history of Alzheimer's and the right have a family history of Alzheimer's.

Figure 8. Stacked bar chart showing the relationship between Gender and Diagnosis



Note: The pink bars are patients without diagnosis and the blue bars are patient with a diagnosis. The left side are the males, and the right sides are the females.

Figure 9 Overlaid Density plot showing relationship between MMSE and Age

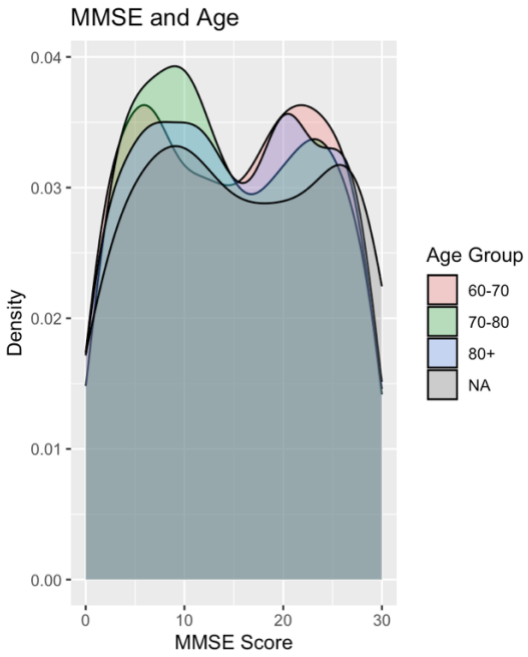
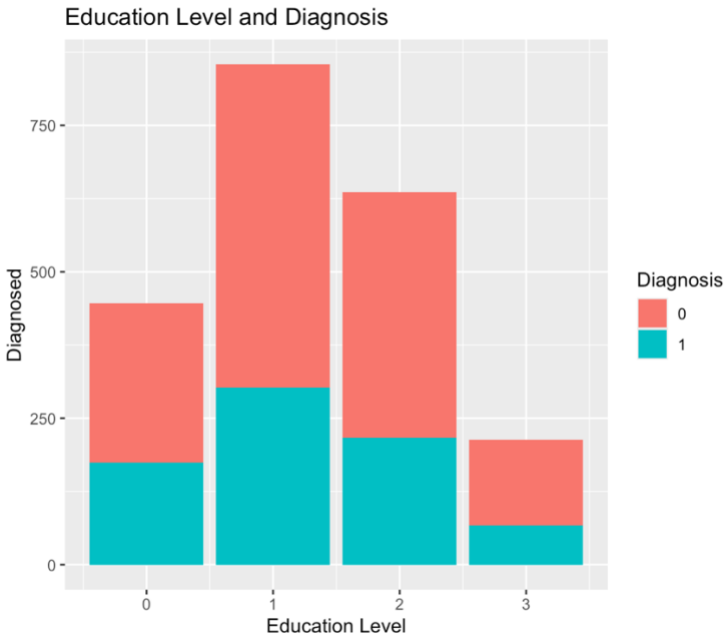


Figure 10 Stacked bar chart showing the relationship between Education Level and Diagnosis



Note:The pink bars are patients without diagnosis and the blue bars are patient with a diagnosis. The chart shows 0 = None, 1 = high school, 2 = bachelors, 3 = higher.

Conclusion and Discussion

This report analyzed the potential correlations between a patient's different health attributes and the likelihood of an Alzheimer's disease diagnosis. The analysis was done using various statistical analysis methods such as measuring central tendency, data dispersion, data skewedness and data visualization. This report specifically analyzed the possible correlation between a diagnosis and a patients age, gender, family Alzheimer's history, education level, and MMSE score.

Key Findings:

- When analyzing gender and diagnosis it found that female patients are more likely to develop Alzheimer's disease compared to male patients.
- When analyzing education and diagnosis it was found that patients with a lower education level of high school and below were more likely to develop Alzheimer's disease compared to patient with a higher education of bachelor and higher.
- When analyzing age and MMSE it was found that older patient had lower MMSE scores suggesting that as a patient's age increases their cognitive functions decrease.

Limitations of the analysis:

The main limitation of this analysis is that it can't be used as an accurate representation of Alzheimer's diagnosis in patients. The size of the sample used for this analysis isn't big enough to be considered an accurate representation.

Possible Future analysis:

For future analysis maybe a more in-depth look at some of the other possible health contributors to the development of Alzheimer's. This would include a more in-depth look at a patient medical history or environmental factor that could affect patients from different regions.

References

Kharoua, R. E. (2024, June 11). *alzheimer's disease dataset* . Kaggle. https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset/data

Larose, D. T., & Larose, C. D. (2017). *Discovering knowledge in data: An introduction to data mining*. Wiley.

NIH National Institute on Aging (NIA). (2023, April 5). *Alzheimer's disease fact sheet | National Institute on Aging*. Alzheimer's Disease Fact Sheet. https://www.nia.nih.gov/health/alzheimers-and-dementia/alzheimers-disease-fact-sheet

Zumel, N., Mount, J., Howard, J., & Thomas, R. L. (2020). *Practical data science with R*. Manning.