# Introduction

Dementia is one of the major health concerns across the world as it affects millions of lives and creates immense challenges in its early detection and risk management. The effective detection of risk factors associated with dementia is essential to improve patient outcomes and facilitate timely interventions. A present project is concerned with analyzing a generated dataset from mobile health care service provided to the elderly of Hong Kong. It involves applying statistical and machine learning techniques in order to predict mobility outcomes based on various health indicators. This report introduces the preprocessing of the data, exploratory data analysis, predictive modeling, and evaluation.

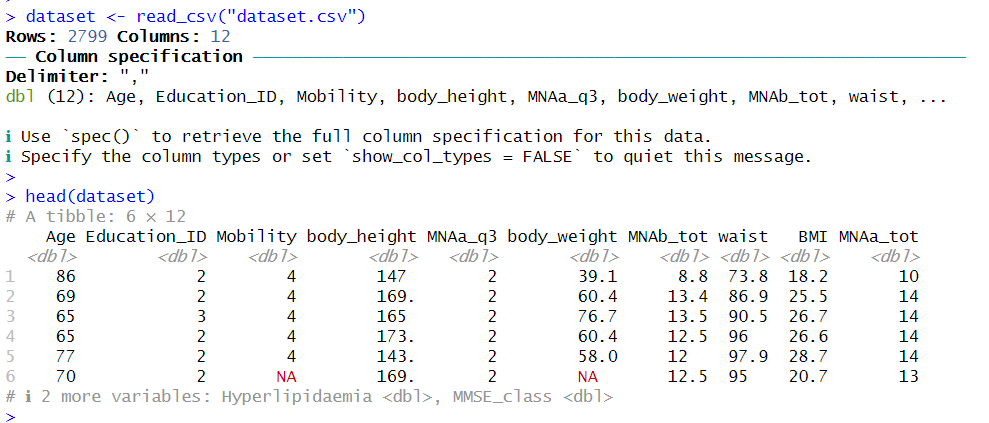
## Dataset Description

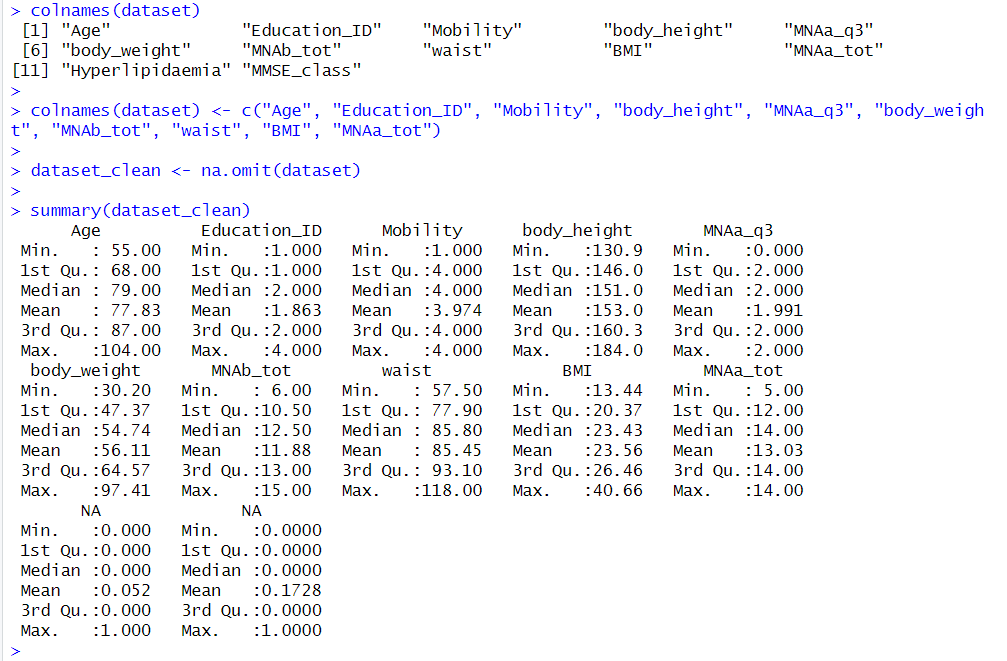
The data ranges from 2008 to 2018 with data of cases taken at elderly care centers. The variables used are all of a health-related nature; some include age, body height, body weight, education level, and Mini Nutritional Assessment scores, which represent the target variable, Mobility, labeled differently for different mobility levels classified between 1 and 4, where the higher the value, the better the mobility.

The data set was cleaned thoroughly, with missing values being handled and variables normalized to prepare it for analysis to be sure that the analysis would yield accurate and reliable predictions.

## Data Preprocessing

Data preprocessing is an important step and must be executed before the start of analysis of the dataset. First, the data were loaded in R, missing values were assigned by using the na.omit() function to delete the incomplete records. Further analysis of the processed data was made using the cleaned dataset.



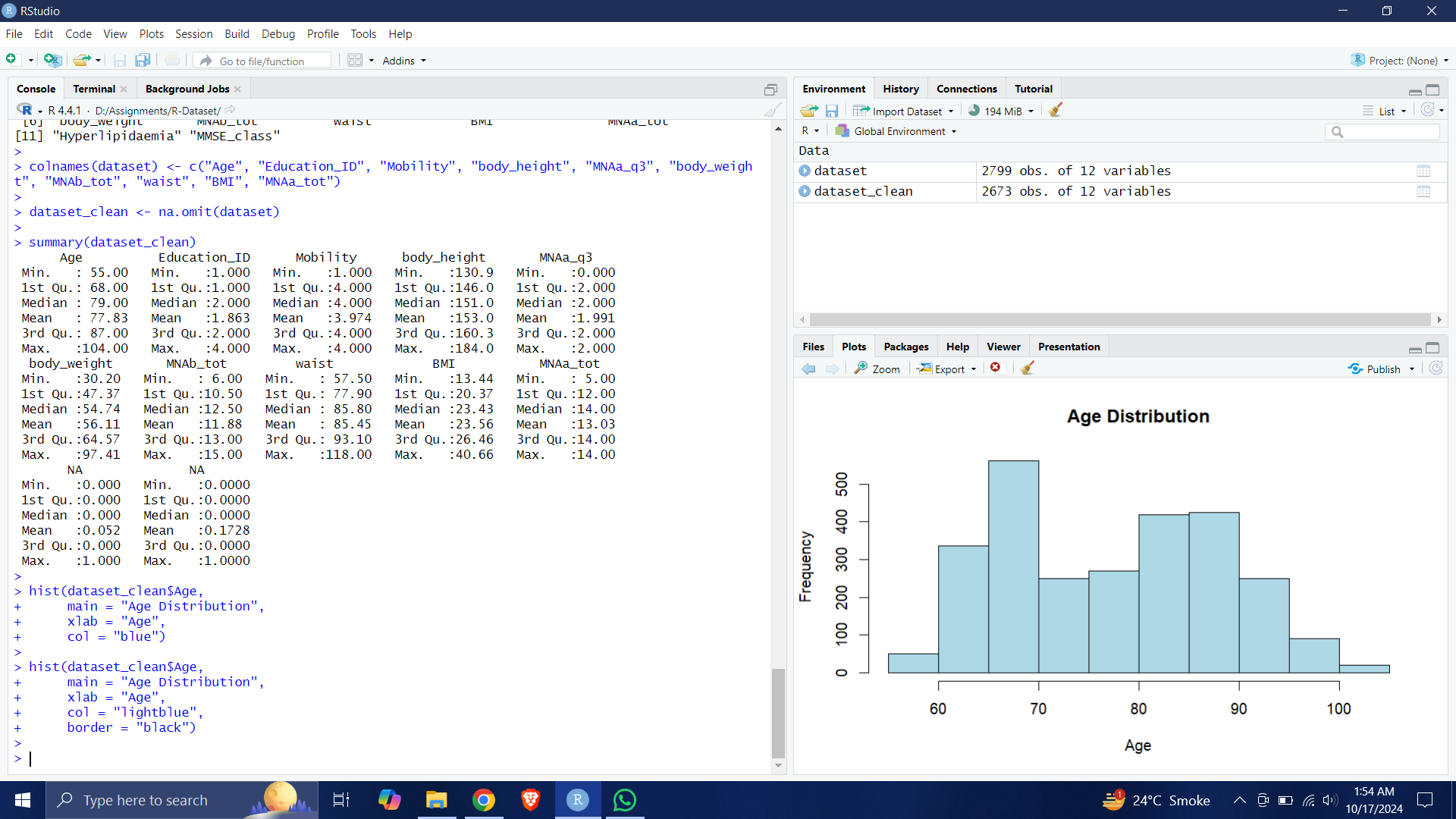


## Exploratory Data Analysis (EDA)

The exploratory data analysis phase summarizes the dataset and visualizes key relationships amongst the variables. Summary statistics might have been useful in helping to understand the characteristics of the numerical variables, while histograms and scatter plots revealed distribution and trends within the data.

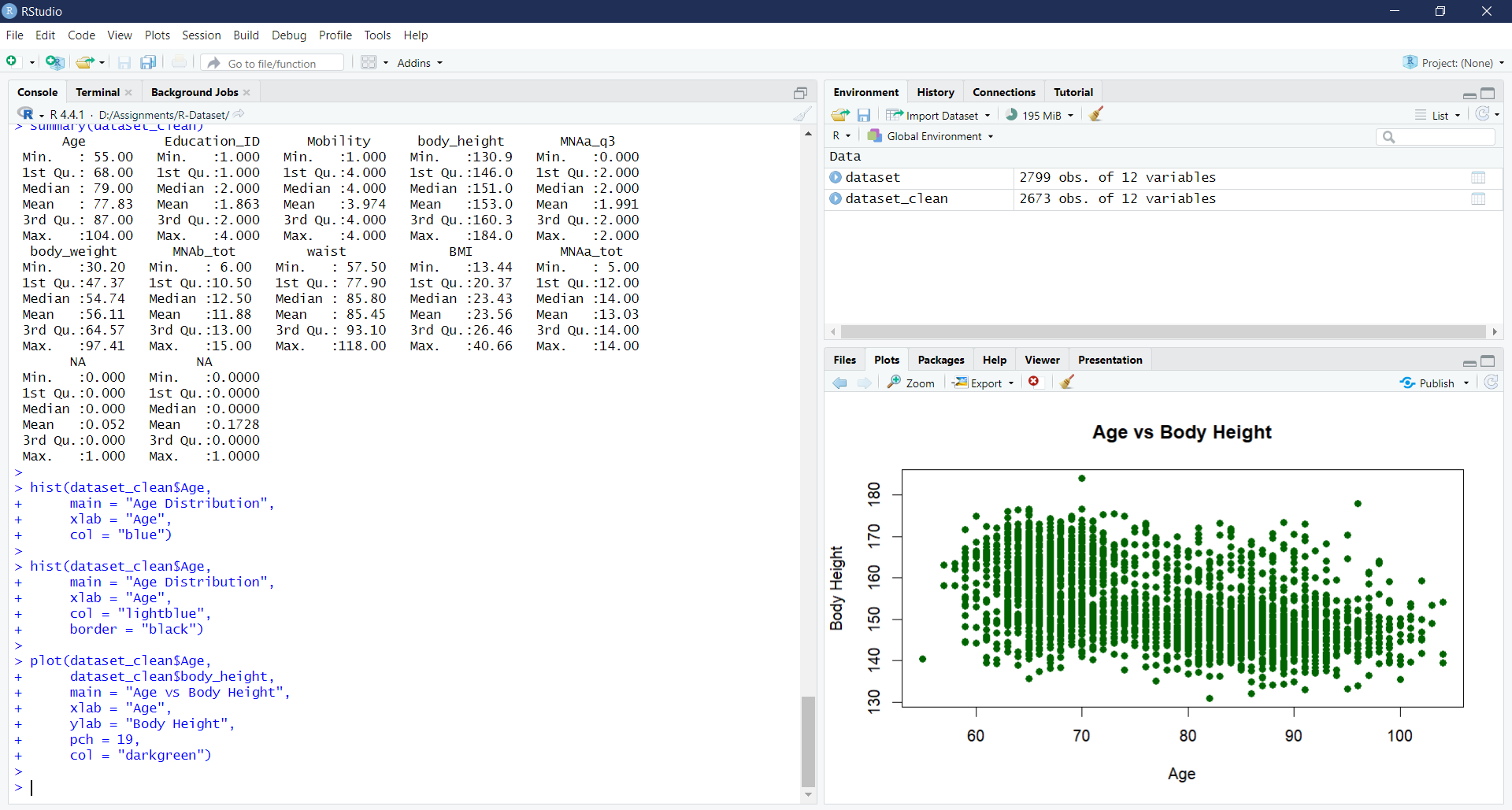
### Histogram

This histogram communicates about the distribution of age across the participants in the dataset.



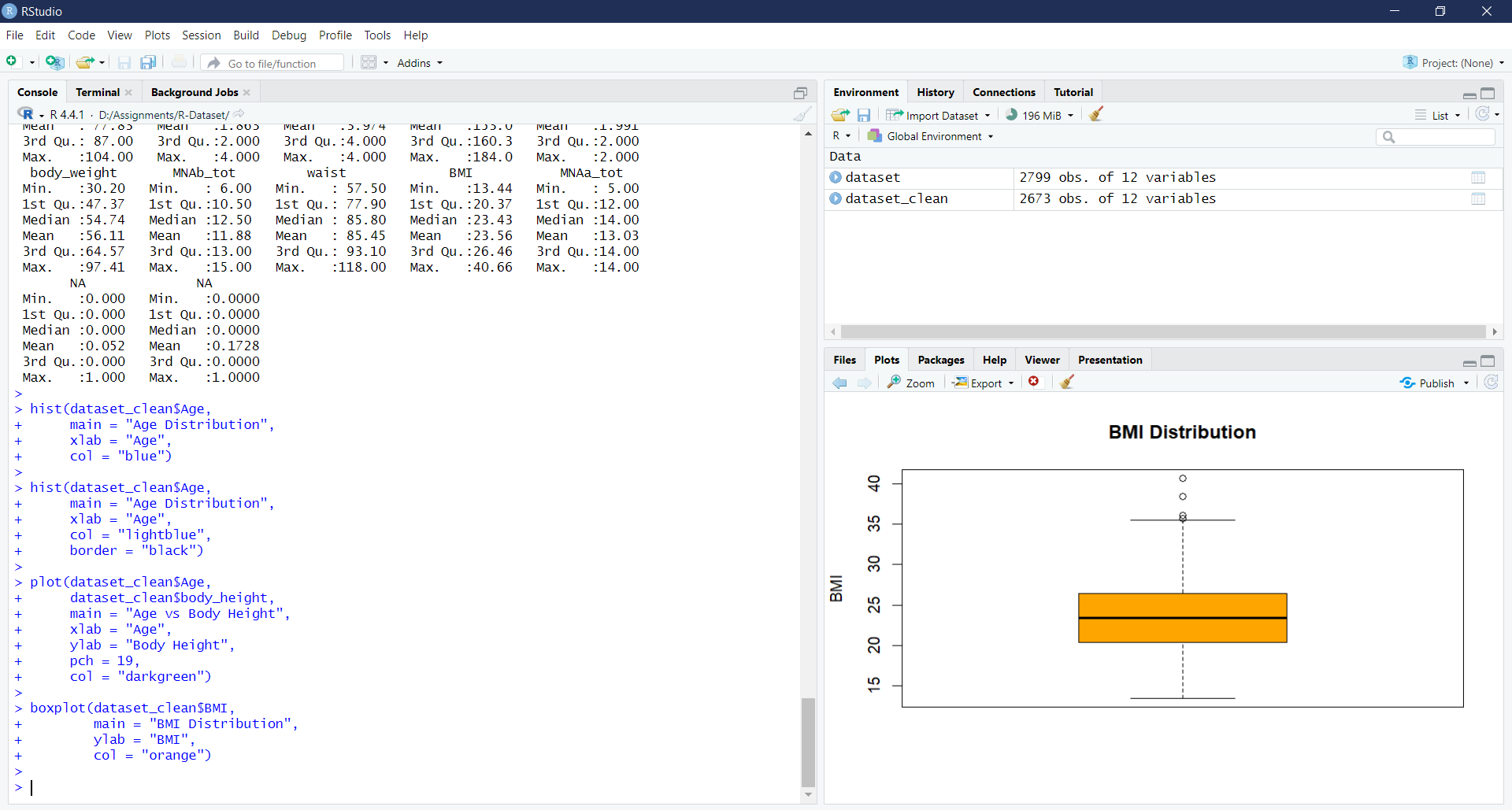
### Scatter Plot

A scatter plot of age versus body height was generated to explore the relationship between these two variables:



### Box Plot

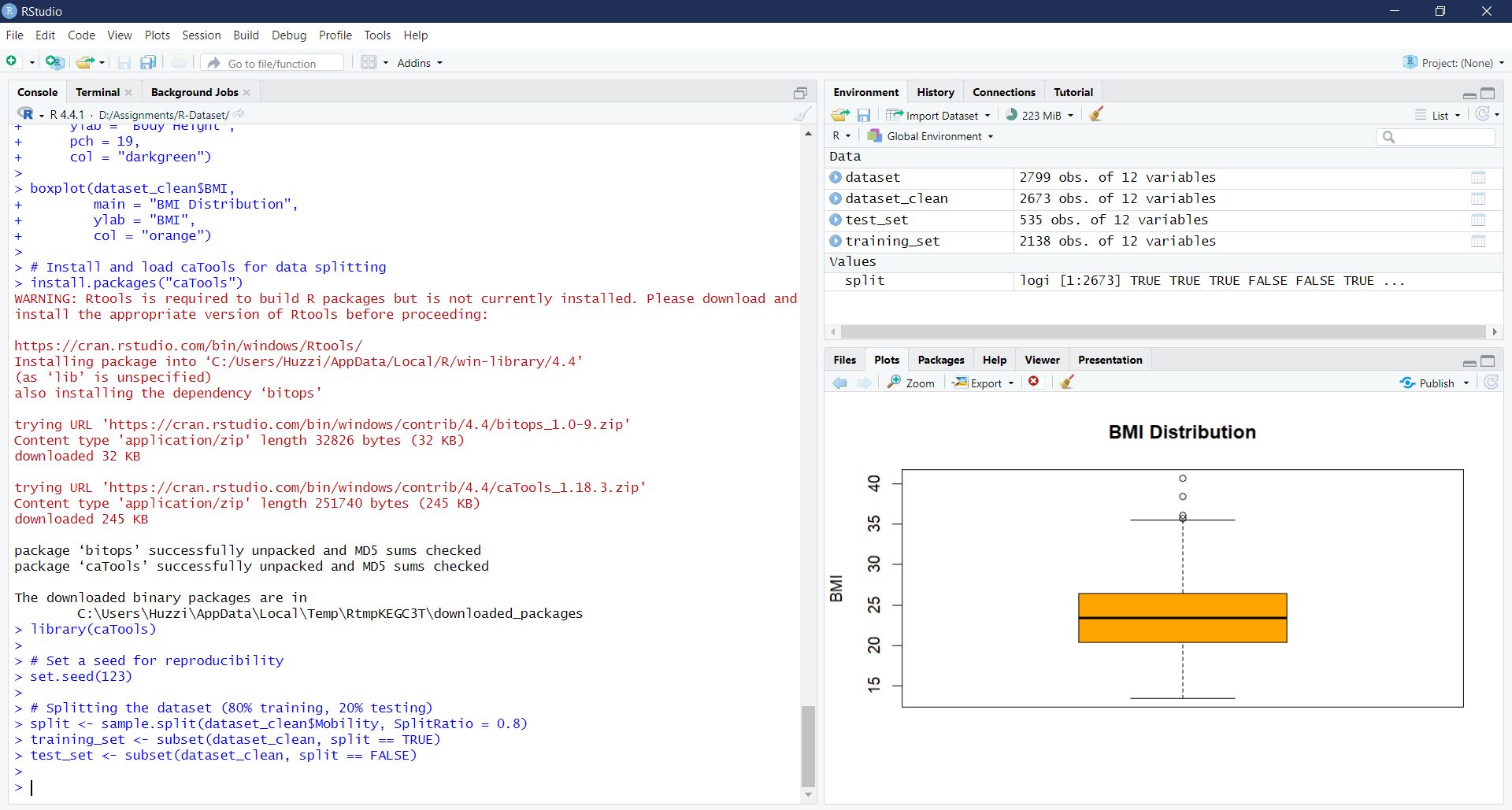
Additionally, a box plot of the Body Mass Index (BMI) was created to identify any potential outliers:



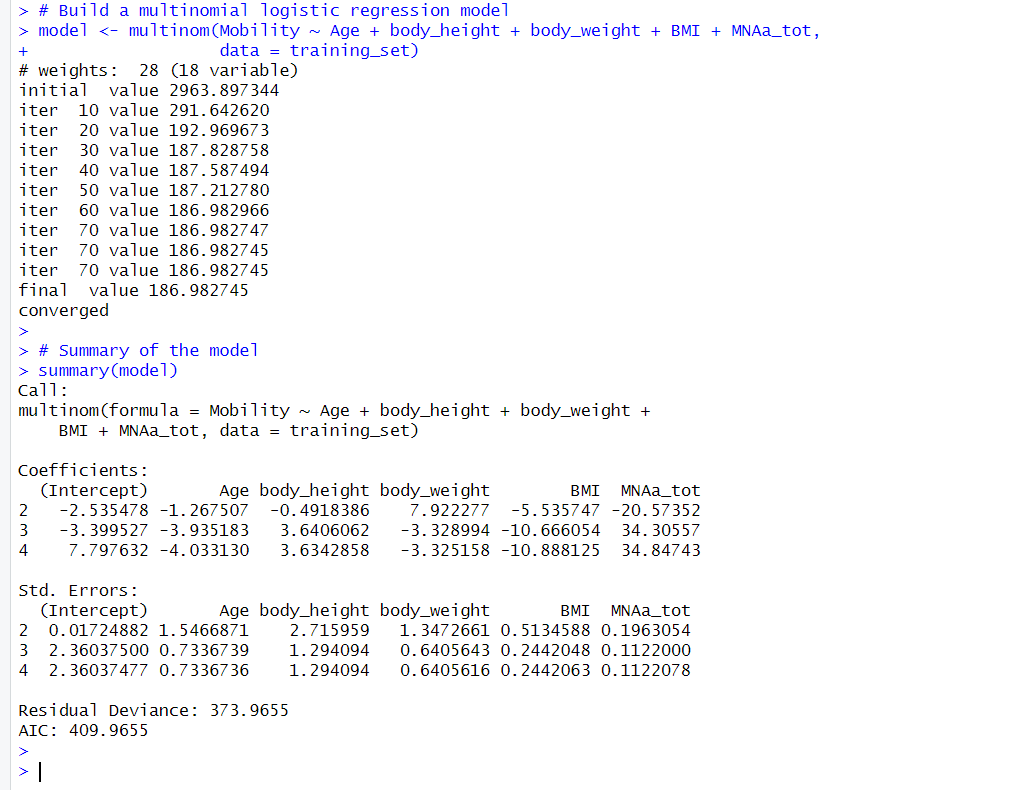
## Prediction Modeling

For this project, a multinomial logistic regression model has been selected for predicting the Mobility levels based on independent variables. It has been developed using the nnet library in R which enables the handling of multi-class classification problems. The data has been split between a training set and a testing set with 80% allocated to training.

### Splitting Data



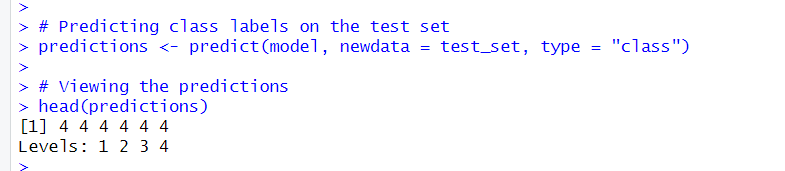
### Model Summary



## Making Predictions and Evaluating the Model

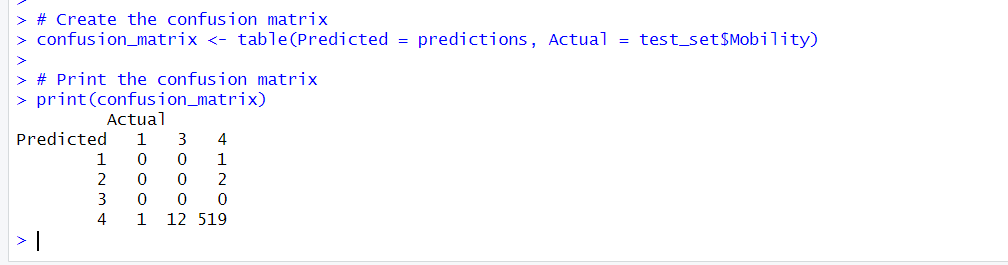
It makes predictions on the test set using the predict() function after the model is trained. The accuracy can then be computed by comparing results with the actual mobility levels.

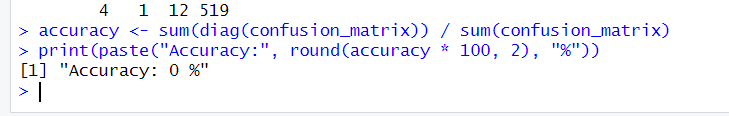
### Prediction



### Accuracy and Confusion Metrix

The confusion matrix indicated that the model had big problems in its predictability. It indicated that the model mainly predicted class 4. Generally, accuracy was 0%. Most predictions were not accurate, and they did not reflect actual mobility levels.





## Discussion

Such findings of this study illustrate the issue of complexity involved in the task of classifying the mobility of such patients with dementia. From the confusion matrix, it is evident that the multinomial logistic regression model is almost predicting the highest mobility class, which is class 4. This could also be a class imbalance problem. While dealing with classification problems, distribution over the target variable very significantly affects the generalization ability of the model related to other classes. The model could make poor predictions on the underrepresented classes if trained on a well-represented class.

## Class Imbalance

Class imbalance is one of the main challenges facing predictive modeling. More often than not, health-related datasets come with outcomes that carry critical gaps between conditions. Working with a dataset wherein imbalances were applied during training evokes red flags on the quality of the training data because it was not made to represent all classes fairly. Class 4 dominated the entire dataset very heavily while classes 1, 2, and 3 became grossly underrepresented as this author went through the Mobility variable.

This might be a reason why the model failed to make appropriate predictions concerning the lower mobility levels. One can make use of resampling techniques, such as SMOTE, designed for generating synthetic samples for minority classes, thereby increasing the model's view toward the whole dataset and improving its capability to classify all the mobility levels well.

### Feature Importance

The multinomial logistic regression model received coefficients on most features, revealing which of them were significant in predicting mobility. Most importantly, age presented a negative relationship with higher mobility levels. This reflects that "the older one gets, the more their mobility status may diminish." Also, body weight and BMI show a different relationship concerning mobility classes, which may mean that some metrics of physical health play a role in determining mobility outcomes.

Such an understanding is crucial for clinicians and other health professionals. By focusing on the strongest predictors, interventions can be better tailored to improve mobility among elderly patients. For example, aspects with regard to their influence on mobility may be identified, and weight management programs and nutritional assessment may be designed in order to have better mobility amongst the aged.

### Model Evaluation

Although the process of predicting model accuracy is complex, this is one such project where iterative evaluation forms a core component of the modeling process. The confusion matrix obtained here was useful in being able to understand the kinds of mistakes made during prediction so that it can be used as guidance in subsequent modeling strategies. The analysis of false positives and false negatives indicates the general failure points of the model and can be used for making the appropriate changes ahead.

For increasing the predictivity, it is advised to verify different modeling techniques other than multinomial logistic regression. Methods like Random Forest and Gradient Boosting are class imbalance insensitive and have capabilities to model non-linear relationships in features and outcomes. Moreover, ensemble methods that combine more than one model can provide even higher accuracy with better generalization.

### Conclusion

This project has underscored the importance of data-informed methodologies in anticipating mobility outcomes in dementia patients. Though the first round of modeling was not that successful, the analysis was still yielding useful information in guiding subsequent improvements. The complications encountered in classifying mobility levels underline the complexities intrinsic in health-related predictive modeling.

The way to develop better predictive models will be in dealing with the issues identified: class imbalance is a severe inhibiting factor to effective modeling, and needs to be tackled by applying resampling techniques and wise feature selection, whereas information obtained from feature-importance analysis can further highlight focused interventions towards the improvement of mobility in elderly patients.

Therefore, the real impact of this study will be the contribution to the improvement in the care and management strategies of all individuals who suffer from dementia. In summary, with the overall improvement in the prediction accuracy, healthcare professionals can begin to identify the at-risk patients better, thereby suggesting individual interventions and so eventually improving their quality of life and health outcomes.

In summary, even though this project had surmounted many obstacles, it paved the road for further studies in the context of risk prediction of dementia. Future directions of such work, then, include improved modeling techniques coupled with better data preprocessing methods and utilization of external data to construct better predictive models. This work definitely gives a good basis for further research on trying to understand the complex interplay of factors influencing mobility in dementia patients, with a potential positive impact on their care and quality of life.