Application of Machine Learning Models in Diagnosis

John Awebwa\*, Hong Ding\*\*, Dianne Williams\*\*\*, Kunal Singh\*\*\*\*

\*[john.awebwa@marquette.edu](mailto:john.awebwa@marquette.edu), \*\*[hong.ding@marquette.edu](mailto:hong.ding@marquette.edu), \*\*\*[dianne.williams@marquette.edu](mailto:dianne.williams@marquette.edu), \*\*\*\*[kunal.singh@marquette.edu](mailto:kunal.singh@marquette.edu)

**Abstract-**  Medical conditions that are difficult to diagnose consume significant time and expense for both patients and medical professionals. More than ever, diagnostic models generated with machine learning algorithms play an important role in medical decision-making, helping physicians to provide a fast and accurate diagnosis, and helping patients to receive critical treatment sooner. Benign Paroxysmal Positional Vertigo -BPPV-is a widespread condition marked by vertigo affecting approximately 107 people per 1,000 per year.[1] BPPV has a high misdiagnosis rate. [1] Previous research in this area have effected electronic pre-visit electronic surveys and various models- machine learning and others- to guide diagnosis accurately to “BPPV” or “NO BPPV”. The intention of this study is to improve BPPV diagnosis by implementing machine learning methods to develop an electronic survey based clinical support system consisting of a minimum set of questions that achieve a 90% accuracy rate for BPPV.

**1 INTRODUCTION**

Complex medical diagnosis is frequently a lengthy and complicated process, occupying prolonged amounts of time for clients, physicians and healthcare staff. Subsequently, medical resources are often inefficiently allocated, and client costs may be excessive. Increasingly, healthcare organizations are finding solutions to lengthy referral and diagnosis issues with machine learning and electronic surveys that serve as a medical triage to classify symptoms and refer clients to appropriate diagnosis recourse. Faster accurate diagnosis results in improved client outcomes, as well as saving time and money for all involved, and allows for more clients to be seen and helped by healthcare professionals.

BPPV (Benign Paroxysmal Positional Vertigo) is a common dizziness disease with very high misdiagnosis rate. Consequently, clients and healthcare professionals alike stand to benefit from methods for faster more efficient diagnosis of BPPV. This study aims to improve BPPV diagnosis by implementing machine learning methods to procure an electronic survey based clinical support system consisting of a minimum set of questions that can achieve an accuracy of approximately 90% for BPPV diagnosis.

Related works have implemented a linear pre-visit questionnaire to expedite BPPV diagnosis (Friedland et. al)[5], and more recently a study by Richburg, Povinelli and Friedland developed an electronic survey questionnaire as a clinical support system for BPPV diagnosis with high accuracy and specificity using machine learning. [2] In addition, medical implementation of machine learning models to help diagnose illness is seen in studies implementing a self- referral decision support framework for low back pain as in “Evaluation of three machine learning models for self-referral decision support on low back pain in primary care”, and work by Ayeldeen et. al., “Prediction of Liver Fibrosis stages by Machine Learning model: A Decision Tree Approach”. [3][4]

This project work is built on the framework of early two-phase survey conducted to diagnose BBPV using decision tree [2]. Two iterations of survey and data analysis have been conducted and our project is moving to the phase three and four of our research. Our study will employ decision tree and ANN models to generate and classify diagnostic features of BPPV and predict outcomes using Weka based on survey results. Machine learning models such as Naive Bayes, and K-Nearest Neighbors, will compare decision trees and ANN models. Current accuracy rate using this model in diagnosis is around 70%. This study will implement the new models and examine accuracy for improvement, with considerations for the potential for certain noise in the data set. This project will continue to explore a novel ensemble approach to improve the accuracy of diagnosis for BPPV. Expectations are to develop a model with high accuracy, sensitivity and specificity in our prediction with the help of machine learning algorithms. Additional literary analysis will examine similar research to further understanding of potential methods to incorporate and deepen project understanding.

**2 RELATED WORK**

Relevant studies include work by Friedland et. al., which implemented a linear pre-visit questionnaire to expedite BPPV diagnosis. The questions were administered prior to the office visit and concerned patient history in an effort to garner enough information to allow for a narrowed differential diagnosis before the first clinic visit. Because the questionnaire was overly lengthy-10 pages- it was prone to questionnaire discrepancies i.e., skipped questions, dishonest reporting, lack of question meaning, etc. [5]

More recently a study by Richburg, Povinelli and Friedland, developed an electronic survey questionnaire as a clinical support system for BPPV diagnosis with high accuracy and specificity using machine learning. [2] The electronic survey allowed for keeping relevant questions, and the ability to pass over questions that were not relevant. Study objective was to employ machine learning models to classify features as BPPV or not and thereby providing clinical support and accelerating diagnosis and treatment.[2]

In addition, medical implementation of machine learning models to help diagnose illness is seen in studies implementing a self- referral decision support framework for low back pain as in “Evaluation of three machine learning models for self-referral decision support on low back pain in primary care”, where, in an attempt to accurately prescribe the right intervention at the appropriate stage to circumvent condition attaining chronic status, specific machine learning models consisting of decision tree, random forest, and boosted tree techniques to classify low back pain cases.[3] Here, the boosted tree model performed best on the classification of low back pain cases, however, the evaluation measures confirm that all models provided referral advice better than just a random guess, meaning that all models learned some implicit knowledge of the provided referral advices in the training dataset. Finally, research by Ayeldeen et. al., “Prediction of Liver Fibrosis stages by Machine Learning model: A Decision Tree Approach”, also used machine learning techniques to predict an individuals’ degree of liver fibrosis. Here, using decision tree classifier techniques, researchers were able to achieve accuracy rates of 93.7%, higher than studies with similar conditions were able to attain. [3][4]

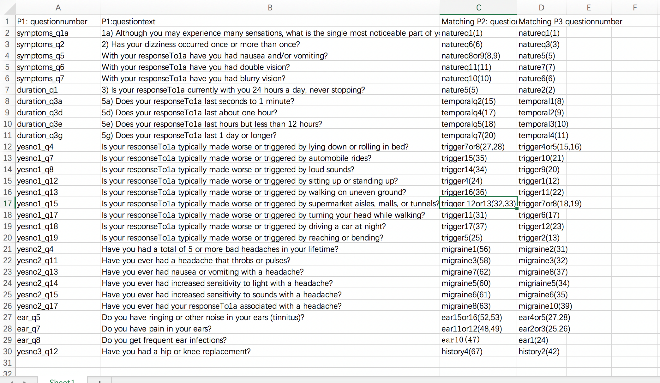
**3. DATASETS AND FEATURES**

1. **Description of Dataset**

The patient data is this study was collected from a survey questionnaire in Medical College of Wisconsin. The study coordinator in Medical College of Wisconsin would help the patients set up the tablets and assist them in filling out the survey in Android Tablets. The answers to the survey were stored in SQLite database when the survey was completed. The doctors in the hospital later would diagnose the patients if he or she has BPPV. Until the end of March in 2019, our research team has collected 74 patient surveys in phase one, 91 patient surveys in phase two, and 100 patient surveys in phase three. In total, there are 266 patient survey collected from three phases. The first phase of the survey contains 84 questions and began in May 2017. The second phase of the survey contains 68 questions and began in November 2017. The third phase of the survey contains 42 questions and begin in August 2019 and is still going on. The survey contains six sections of questions with focus on symptoms of dizziness, timing of attacks, feeling of ear pain and headaches as well as triggers for dizziness and medical history. The phase two survey contains supplemental section. The supplemental section contains six questions and is used to test how well two decisions trees which are generated based on phase one survey would perform.

1. **Combination of Three Survey Questions**

The project has implemented three versions of survey in three phases. Each version of survey has same overlapped questions with the other two. There are 84, 74, and 42 survey questions respectively in phase one, two, and three surveys. In order to efficiently use all of the patient data, we match the same questions from all three phases and get a combined data set.



**4 METHODS**

The following machine learning models were implemented in this study: Decision trees, ANN models, Naïve Bayes, and KNN. Models for this project were selected by the following criteria:

**Decision tree and ANN:** Recent work of a similar nature by Friedland et al., specifically choose decision tree for its ability to display the subject components and structure of the model. The tree root contains the base question and subsequent child nodes hold various questions ultimately leading to a BPPV diagnosis, or “No BPPV”. ANN results were delivered using WEKA for this project, similarly Friedland et al., utilized WEKA for similar analysis.

**Naïve Bayes:** Naïve Bayes reflects our decision to implement a model learned in Data Mining class, as a reliable indicator of model accuracy, sensitivity, and specificity. We felt it added valuable comparative insight.

**KNN:** Finally, KNN was selected as we strove to determine relationship insights in an attempt to narrow the dataset. Faithfully, it did deliver an accurate predictable feature of BPPV-age.

Basic Jitter plot analysis will be used to visualize feature selection.

**A. Decision trees**

The implementation of decision trees model consists of two steps. The first step is to obtain a subset of the most relevant questions using the correlation attribute evaluator in the Weka. The threshold of correlation attitude evaluator is 0.25. Then a J-48 decision tree algorithm in Weka will be used to classify each record as "Has BBPV" or " No BPPV".

1. **ANN model**

ANN model is another machine learning framework that is inspired by biological neural network. The model takes features as input and results as output. There would be hidden layers called perceptron to process the information. We plan to use Weka to adjust the number of layers, the number of perceptron and learning rate to achieve the best prediction we could have.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Number of records | Phase I (74) | Phase 2 (92) | Phase 3 (100) | Combined 266  (74+92+100) |
| Number of Questions | 84 | 74 | 42 | 29 |
| Accuracy | 0.86 | 0.67 | 0.56 | 0.73 |
| Sensitivity (TP) | 0.85 | 0.79 | 0.64 | 0.69 |
| Specificity (TN) | 0.88 | 0.49 | 0.45 | 0.75 |

**C. KNN**

KNN was a method of choice as the model structure is decided from the data- an outcome this study is interested in. This model makes predictions by calculating an input similarity to a training instance. [6] Not wanting to hold authentic medical data to assumptions, a goal of this research is to determine which survey questions provide the highest accuracy for BPPV diagnosis. KNN is often a go to classification method when prior knowledge about data distribution is scant. [6]

**D. Naïve Bayes**

Naive Bayes classification is based on Bayes theorem which provides us a way to calculate the probability of our classification based on existing data.

[7]Bayes’ Theorem is stated as:

P(h|d) = (P(d|h) \* P(h)) / P(d)

Where

**P(h|d)** is the probability of hypothesis h given the data d. This is called the posterior probability.

**P(d|h)** is the probability of data d given that the hypothesis h was true.

**P(h)** is the probability of hypothesis h being true (regardless of the data). This is called the prior probability of h.

**P(d)** is the probability of the data (regardless of the hypothesis).

In this study we are calculating the conditional probability of P(h|d) from the prior probability p(h) with P(D) and P(d|h). Once we calculate the probability of a different number of hypothesis then we select the hypothesis with the highest probability.

Naive Bayes theorem is called Idiot Bayes because of the probability calculation of each hypothesis is made simpler. It is assumed in this theorem that features are independent of each other and do not have relation to each other what so ever.

This is a very strong assumption and might not be true with most of the real-time data. However, this theorem seems to work well with most of the cases. [7]

To arrive at a reliable prediction several steps were completed.

Our dataset had most of the features as categorical and we converted into dummy variable for modelling. Next, jitter plot analysis was conducted to visualize feature selection. In this phase we identify the features that best distinguish the values of outcome variable. The jitter plot explains the spread of different symptoms across the dataset. We noticed that most people who had vertigo were diagnosed with BPPV.

**V EXPERIMENT RESULTS**

**A. Decision Tree**

In the decision tree model, 10 cross validation is used to measure accuracy. And table shows the results generated. For example, in phase one, there are 84 questions and 74 records and the achieved accuracy in phase 1 is 0.86. The table indicates that when the number of questions decreases from phase 1 to phase 2 to phase 3, the accuracy drops from 0.86 to 0.67 to 0.56 as well. Yet, when in the combined dataset which includes 29 questions and 266 records, the accuracy rises again to 0.73. It suggests that the increase of number of records can compensate the loss of accuracy due to the decrease of the number of questions.

**ANN Results**



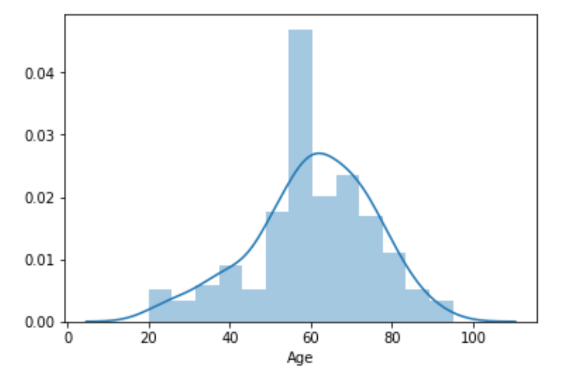
**B. ANN model**

In the ANN model, same 266 records are employed in the analysis. The multilayer perceptron function in Weka in used to generate the confusion matrix. With 10 cross validation, different combinations of layer numbers, node number and learning rate are tested. By adjusting the number of layers, the number of nodes and learning rate, we tested five ANN models. The result shows that if the layer number is 1 or 2, with 17, or 17,9 nodes, it achieved the highest accuracy of 0.71.

1. **KNN Model**

K-NN (k-nearest neighbors’ algorithm) in our project was used for classification as well. In k-NN classification, the output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive [integer](https://en.wikipedia.org/wiki/Integer), typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor. [8] K-NN does not directly work with categorical data and yet most of the questions answered by the patients were a binary category of ‘YES’ or ‘NO’ as seen with the exception of age. Being questionnaire response, the only numeric attribution that was available we to convert some of the responses into dummy variables and apply integers for ‘YES’ and ‘NO’ questions.

As part of preprocessing, features have been transformed as follows 1) Que stion1 q1a responses have been transformed into dummy binary for each answer 2) Ear related answers (0=no ear response, 1=one ear response 2=Two ear response 3) Gender has been mapped as 1=Male and 0=Female 4) The rest of the binary responses Yes=1 and No=0.[9]

One of the features that directly comply with K-NN that did not require transformation is Age.

Age is a very important factor while analyzing medical research. Exploring this data indicates that most of the participants were approximately between the age of 50 and 70. It would of great interest therefore to know how best this age range is a better or worse classifier. If this age range turns out to be a bad classifier, we might as well conclude that age does not matter, or we might need to examine age against some other features.

**KNN Results**

|  |  |
| --- | --- |
| Results with age included | **precision recall f1-score support**  **No 0.50 0.58 0.54 24**  **Yes 0.62 0.53 0.57 30** |
| Results without age | **No 0.92 0.73 0.81 33**  **Yes 0.68 0.90 0.78 21** |

Note that we chose to use **minkowski** distance which is a generalization or combination of both the Euclidian distances.

**KNeighborsClassifier(algorithm='auto', leaf\_size=30, metric='minkowski',**

**metric\_params=None, n\_jobs=None, n\_neighbors=5, p=2,**

**weights='uniform')[8]**

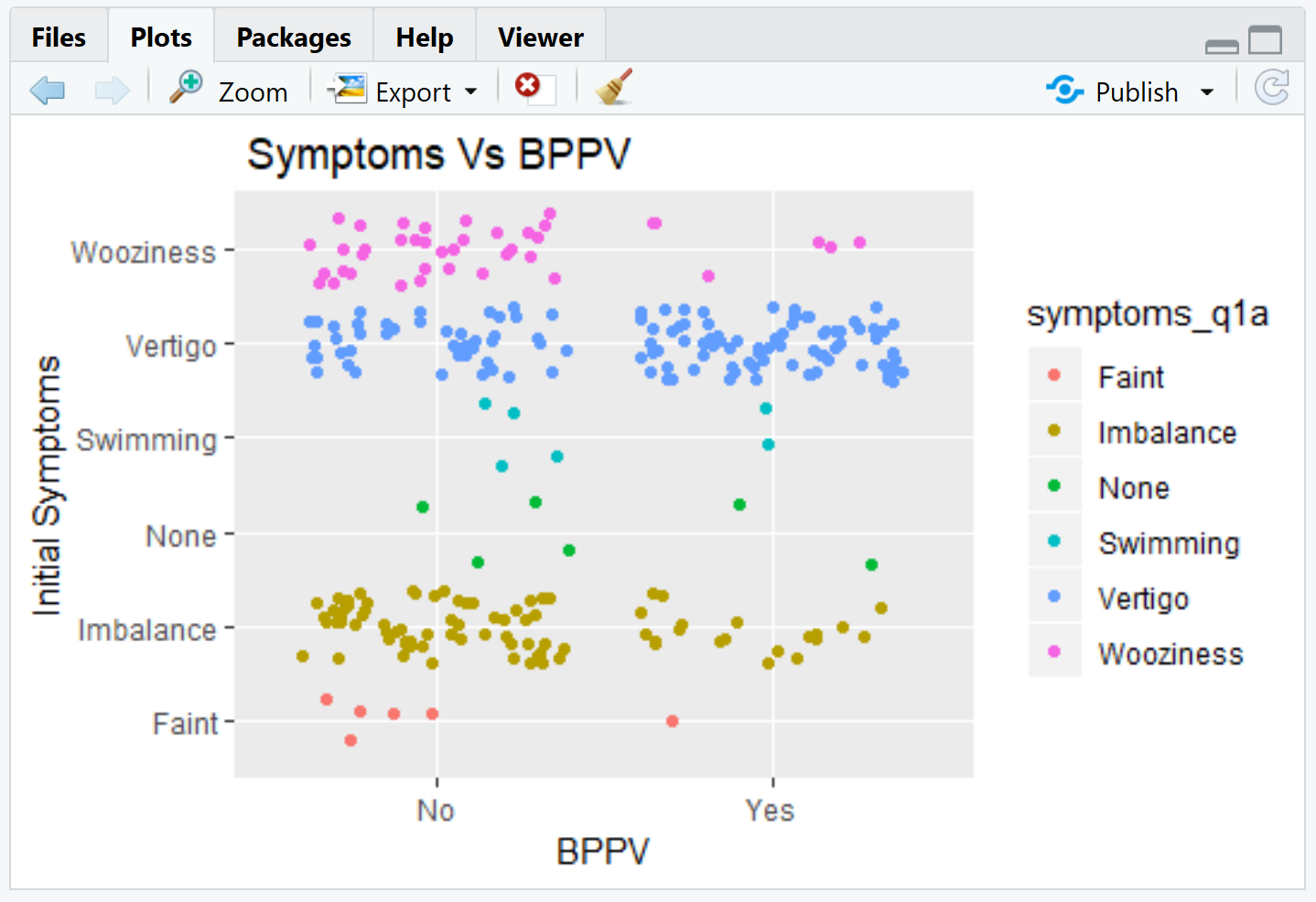
As already noted, the challenge of using K-NN model with our data is the choice of the distance function. Our data was majorly categorical. This means there was no observable gradient between the sets of objects because 0, and 1 numeric value are common between the objects. Aware of the fact that k=K-NN is a lazy classifier (i.e. classification of a selected object is made against a set of the training set[7]). Binary numeric would produce many similar neighbors.

**Feature Selection Jitter plot Analysis**

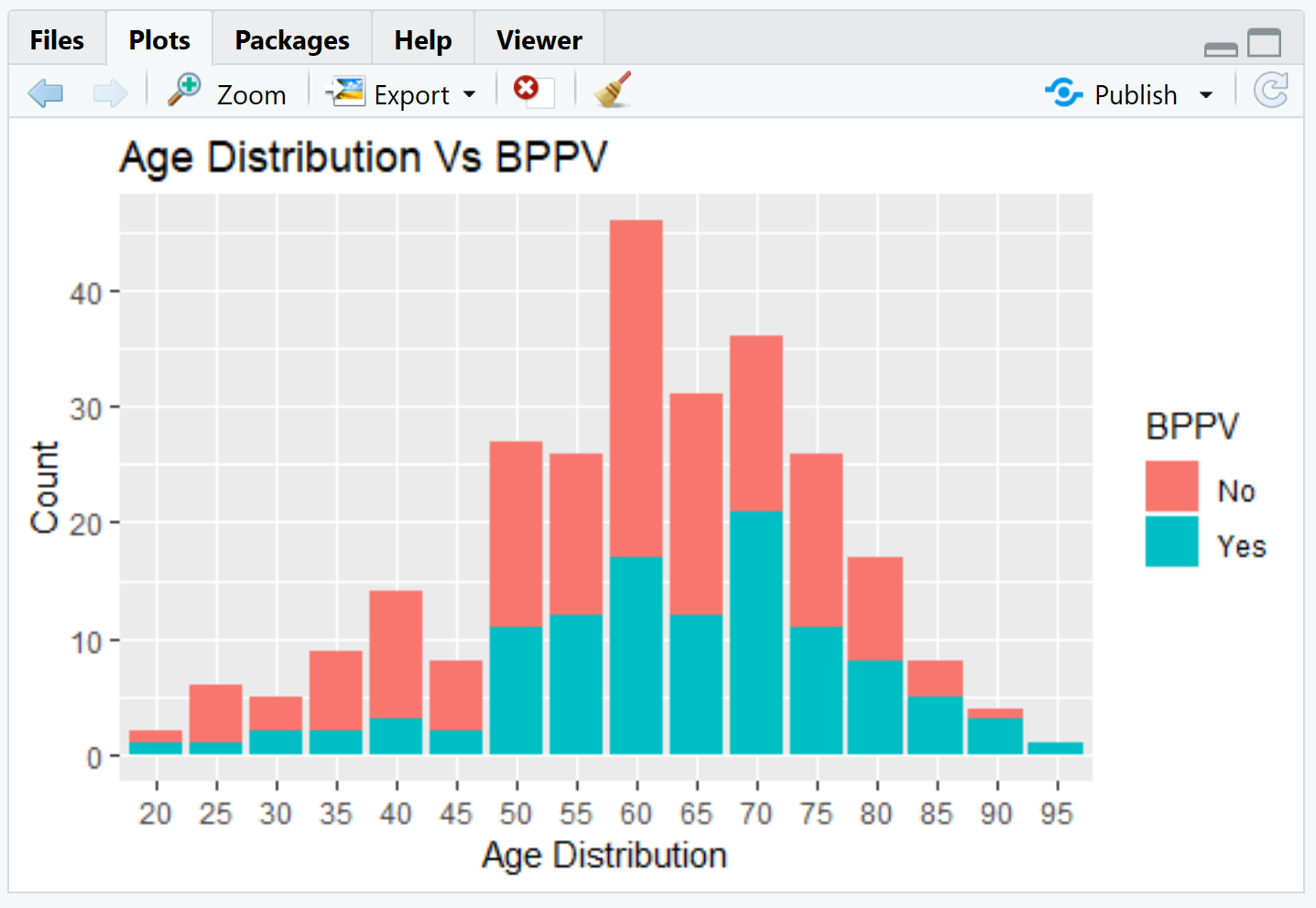
In this phase we identify the features that best distinguish the values of outcome variable.

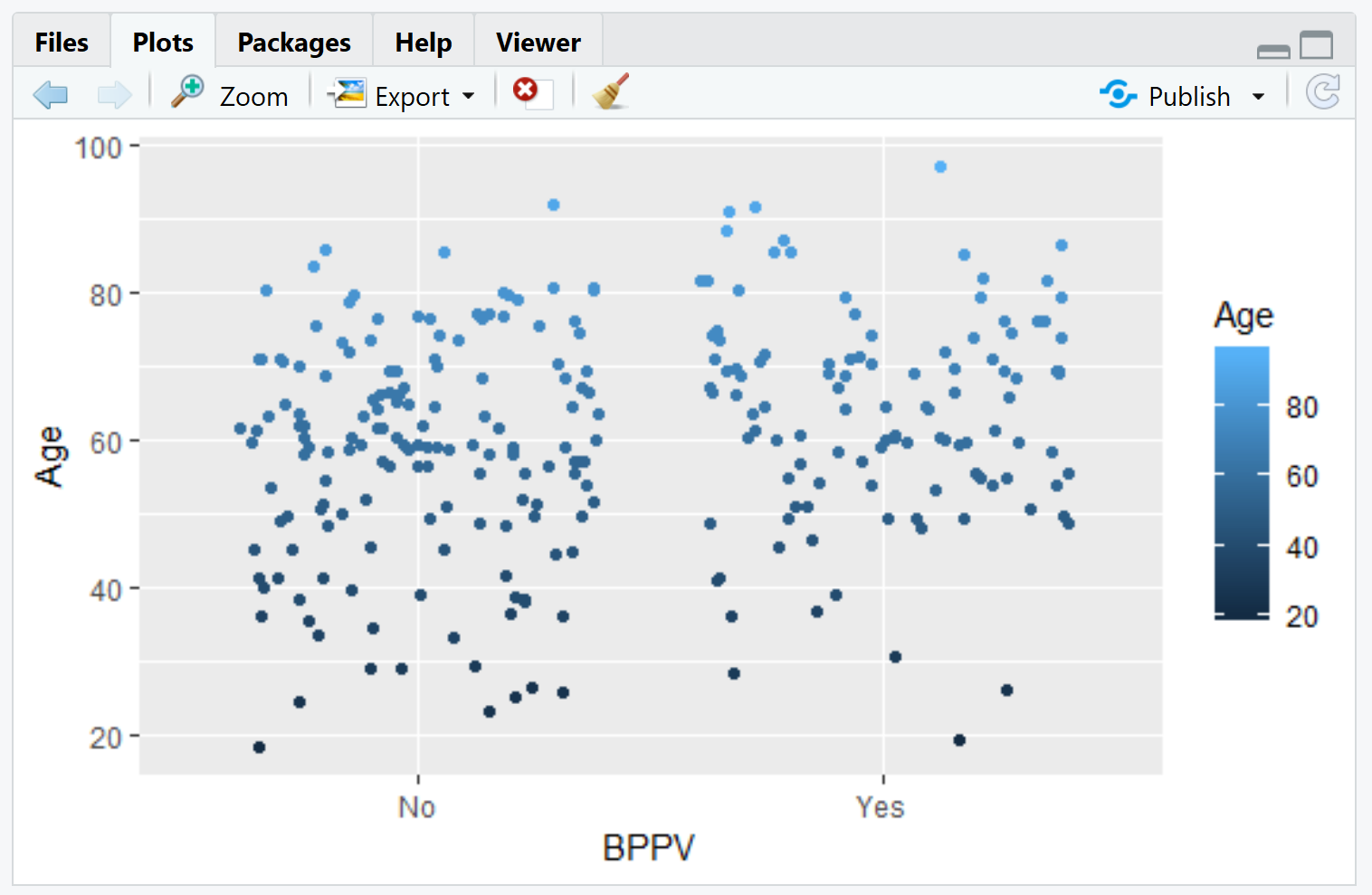
Graph 1 below explains the spread of different symptoms across the dataset. We noticed most people who had vertigo were diagnosed with BPPV.

**Graph 1**

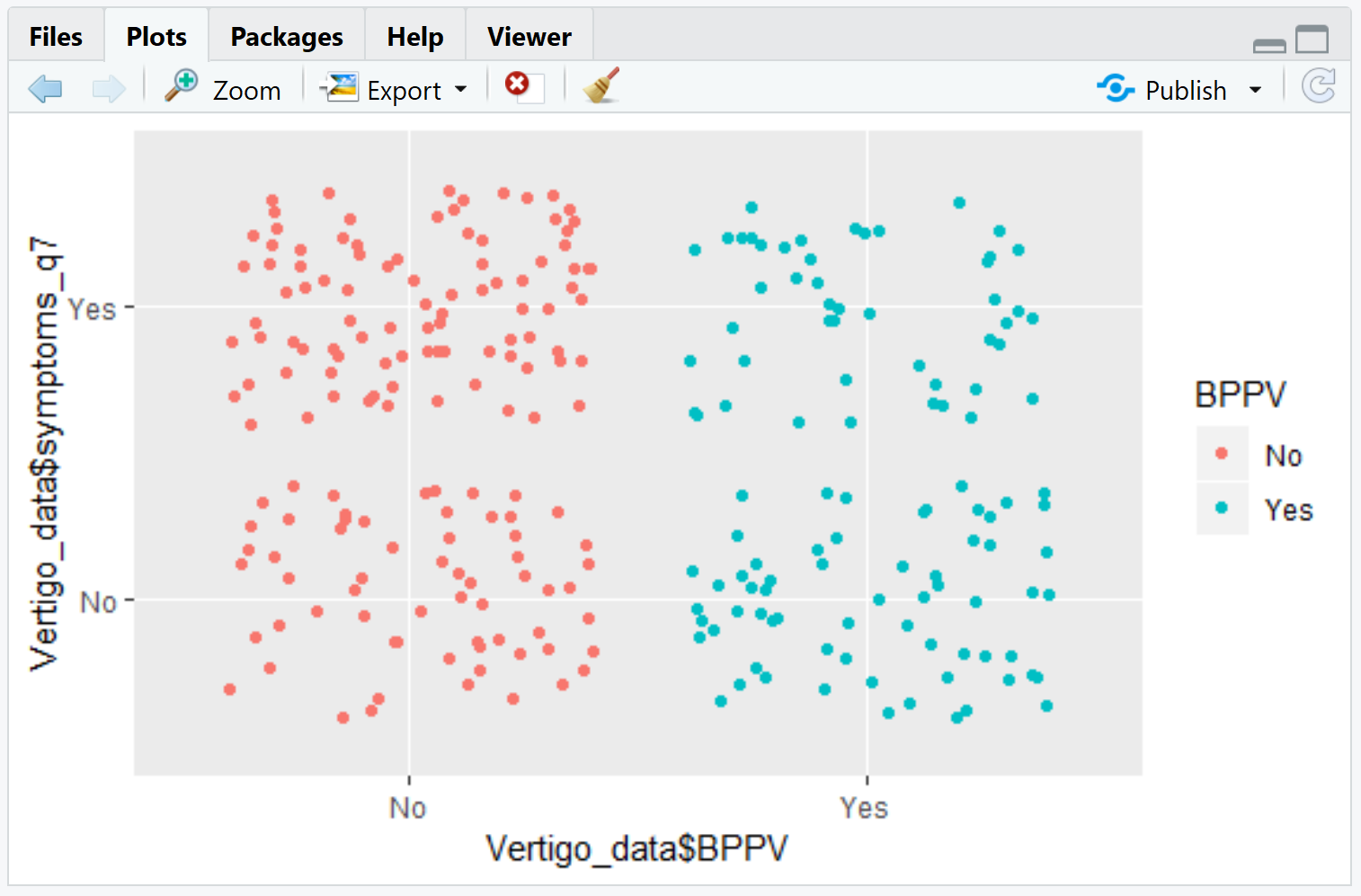


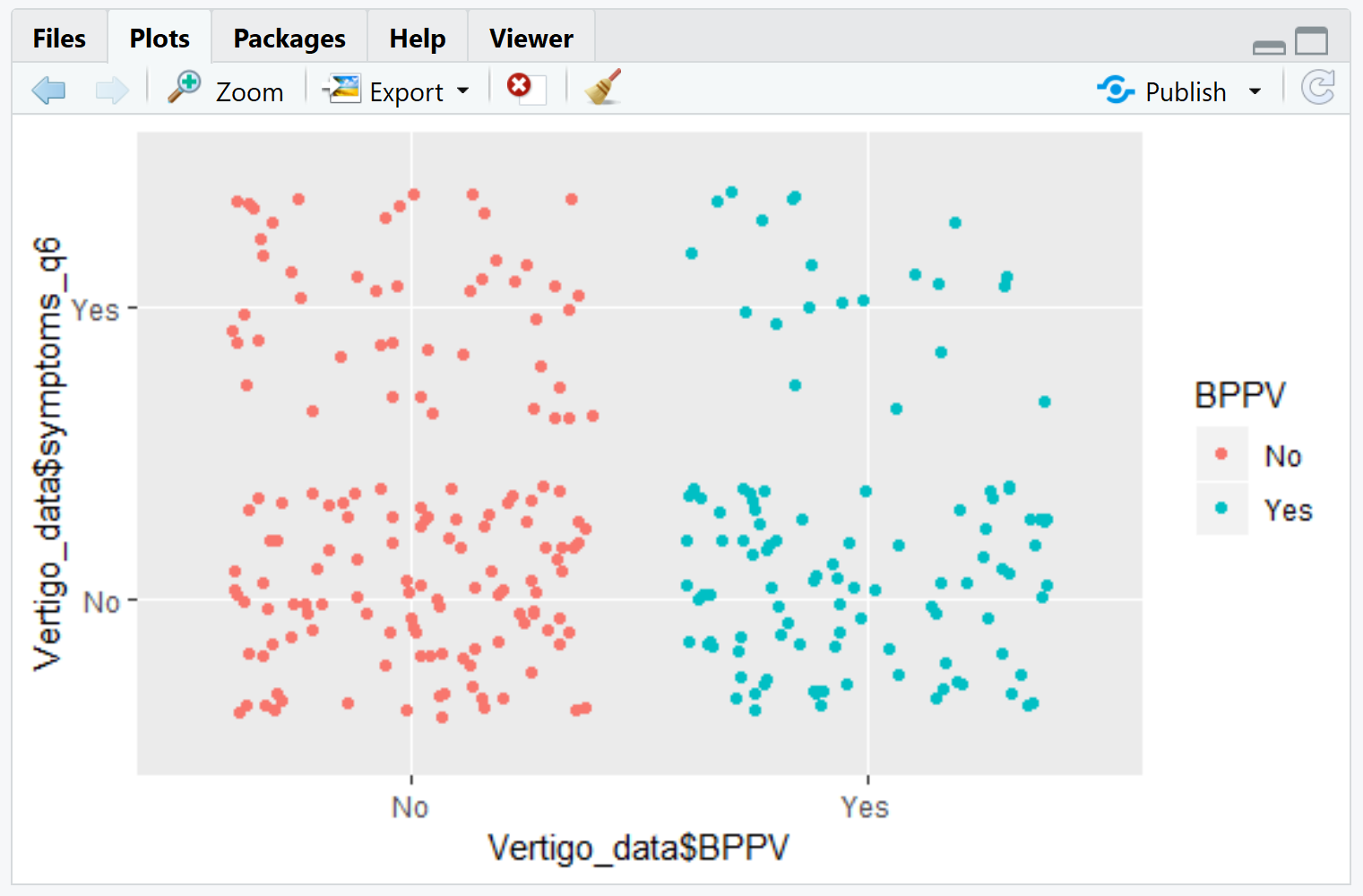
Graphs below indicate the age distribution of the dataset. It seems like over 50 years patients are more likely to be diagnosed with BPPV.

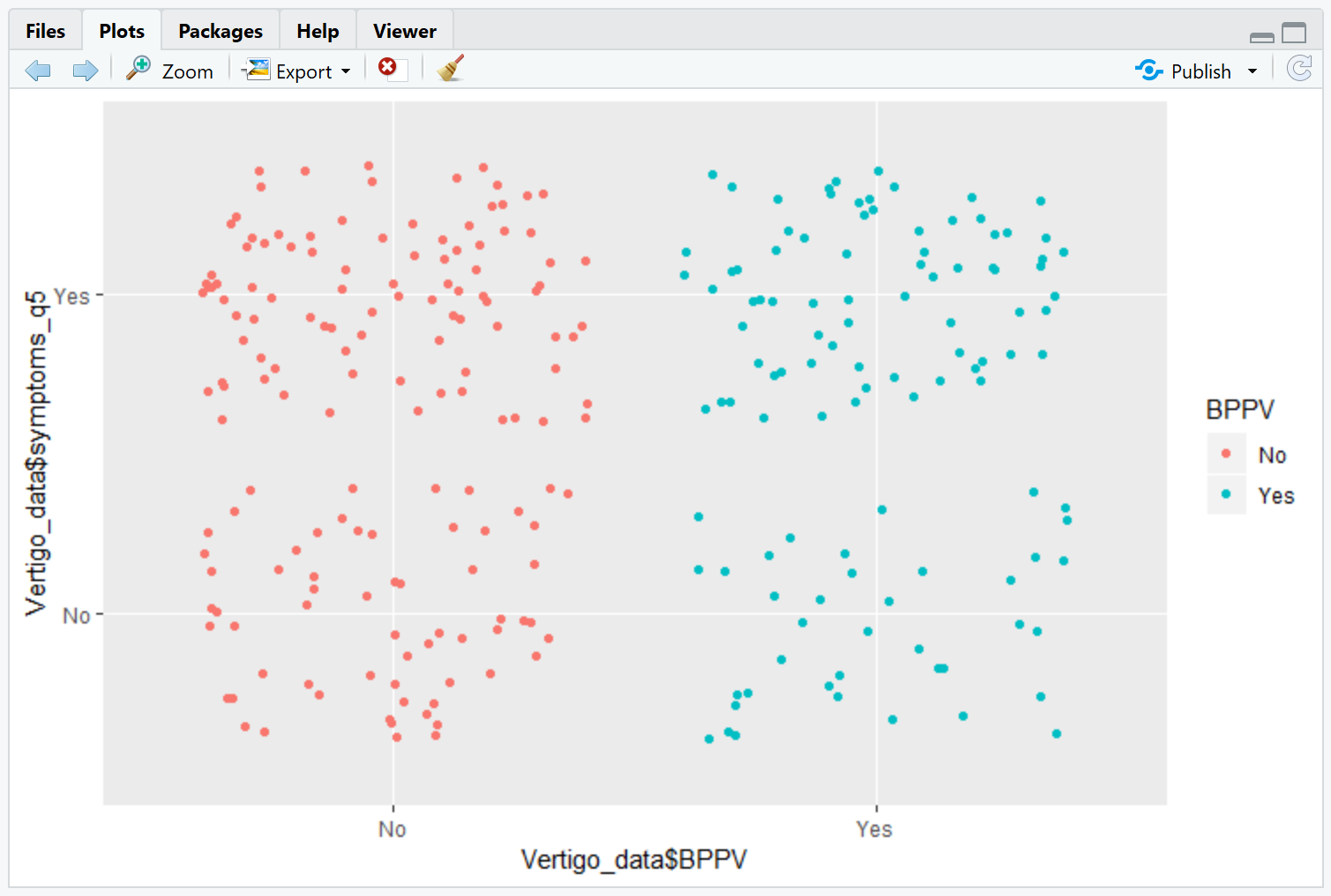


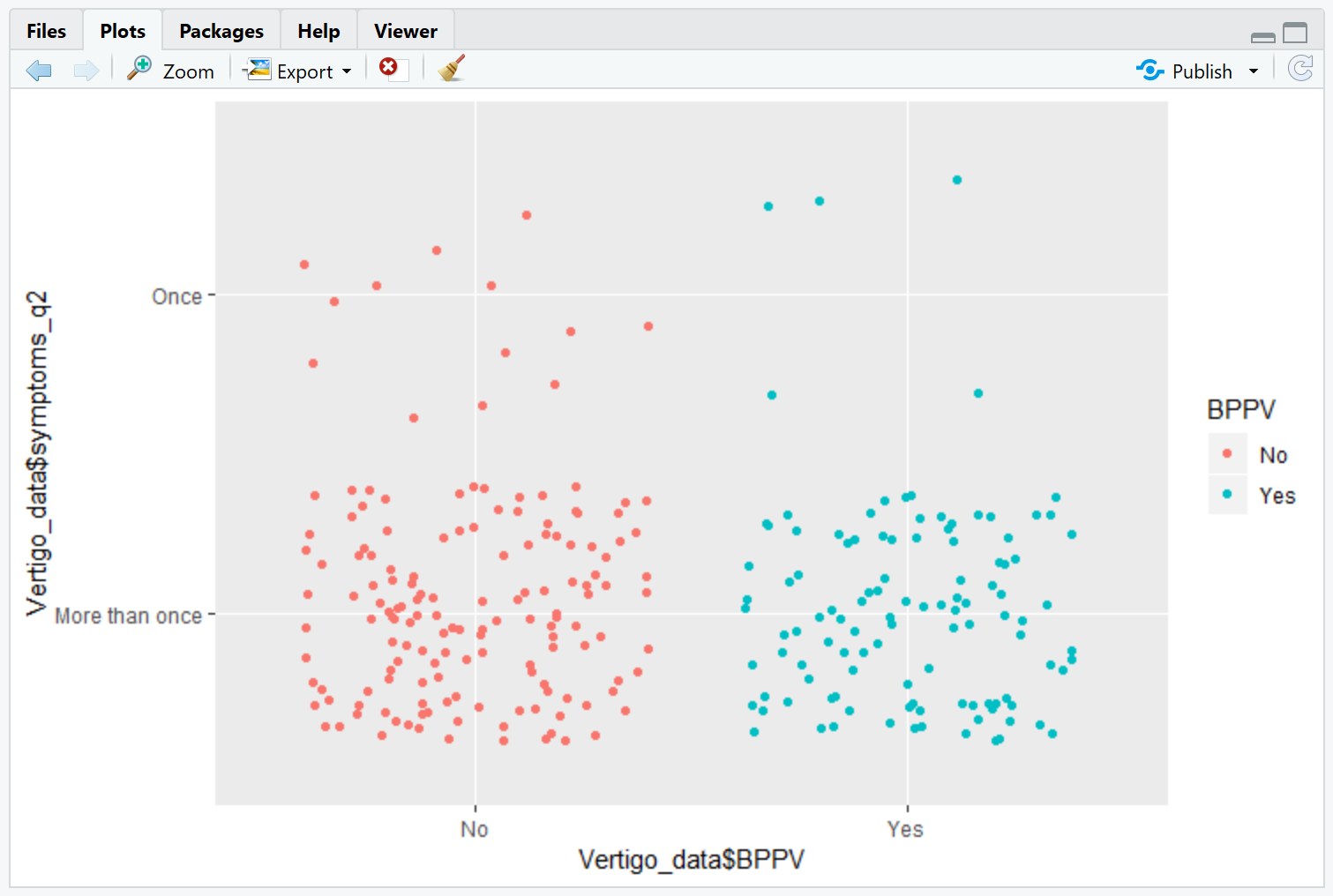


The following four graphs indicate the symptoms responses of Double vision, Blur Vision, dizziness and vomiting sensation. These features seem to have less diagnosis rate.

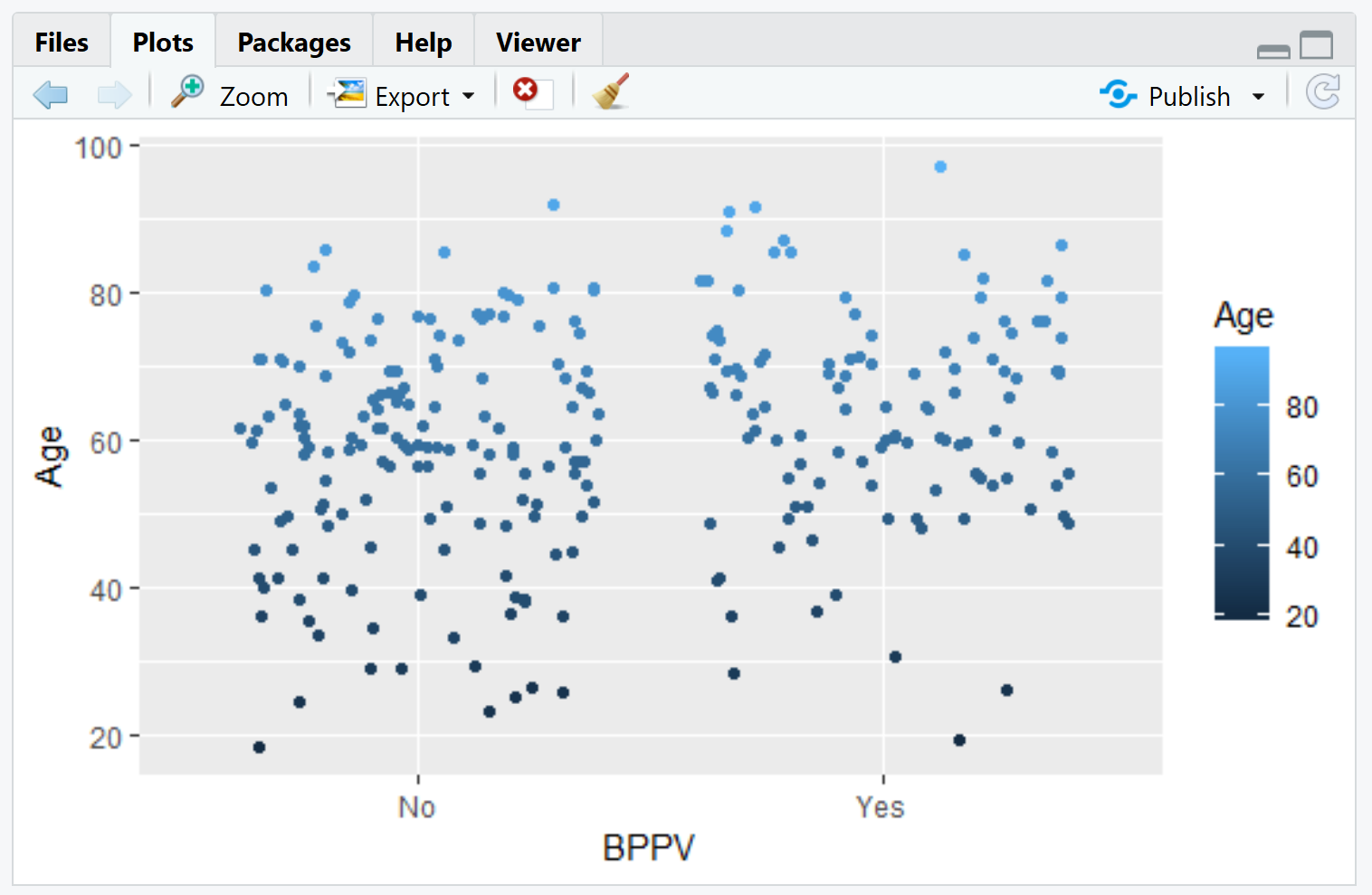


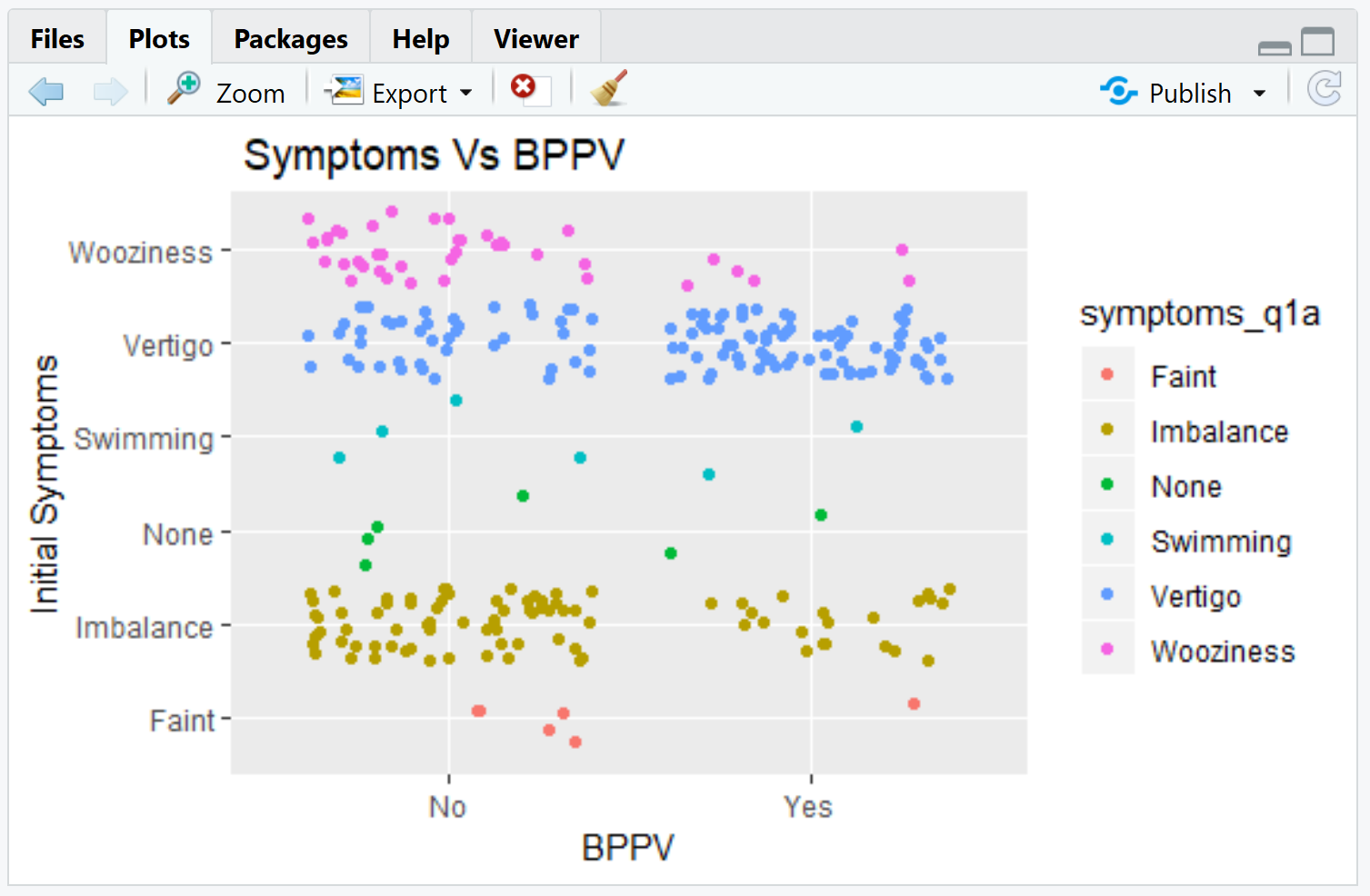


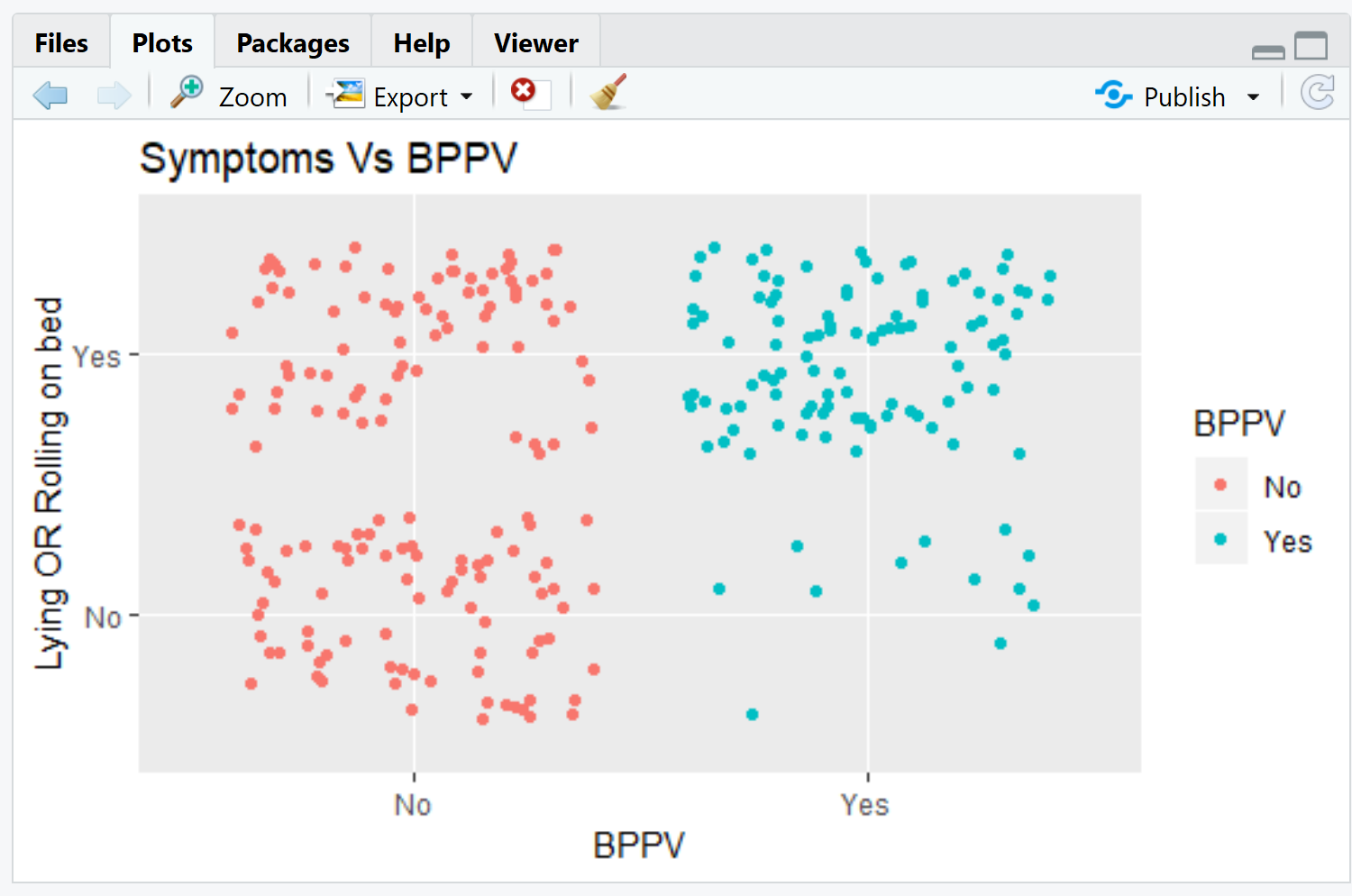


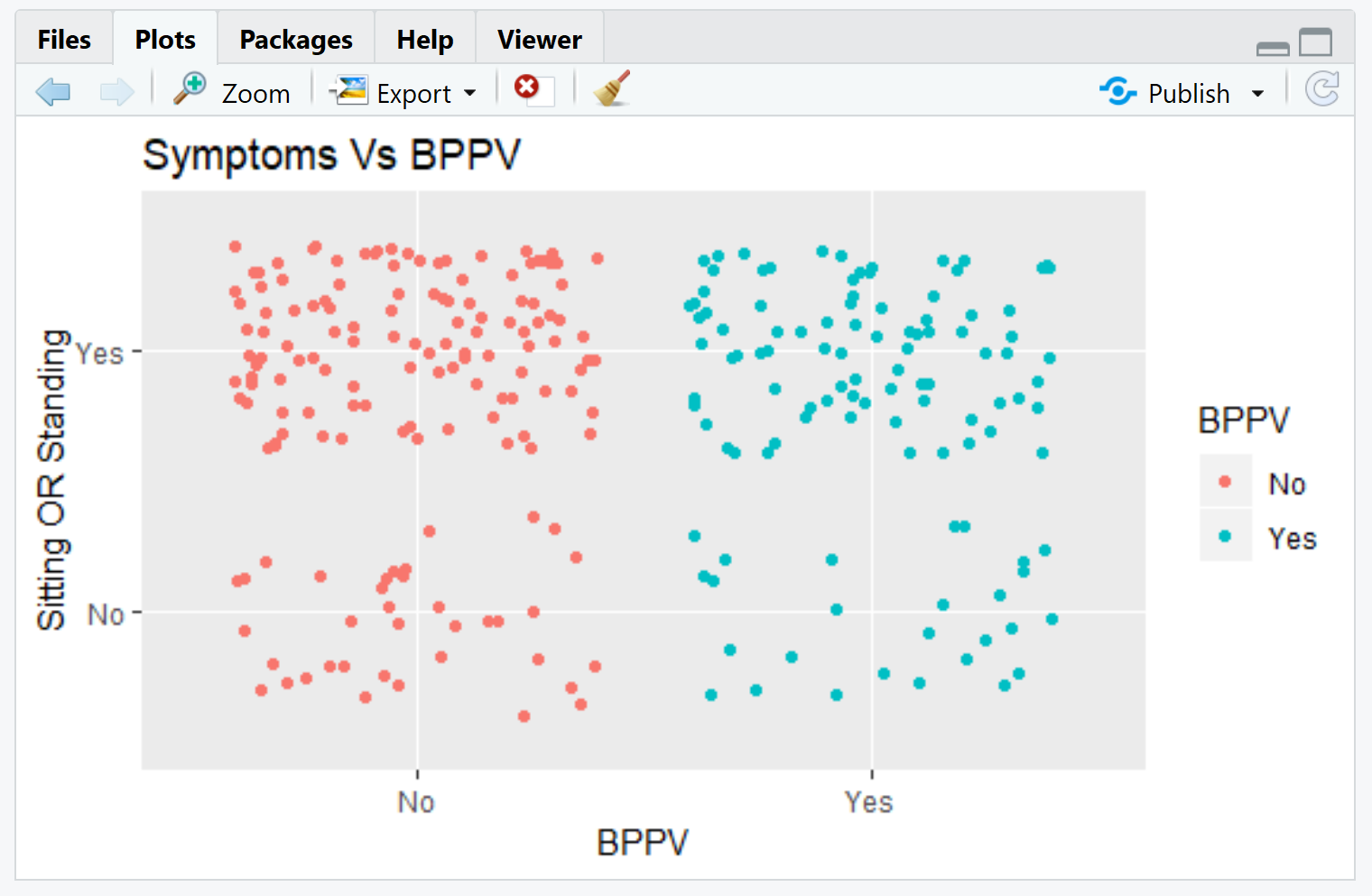


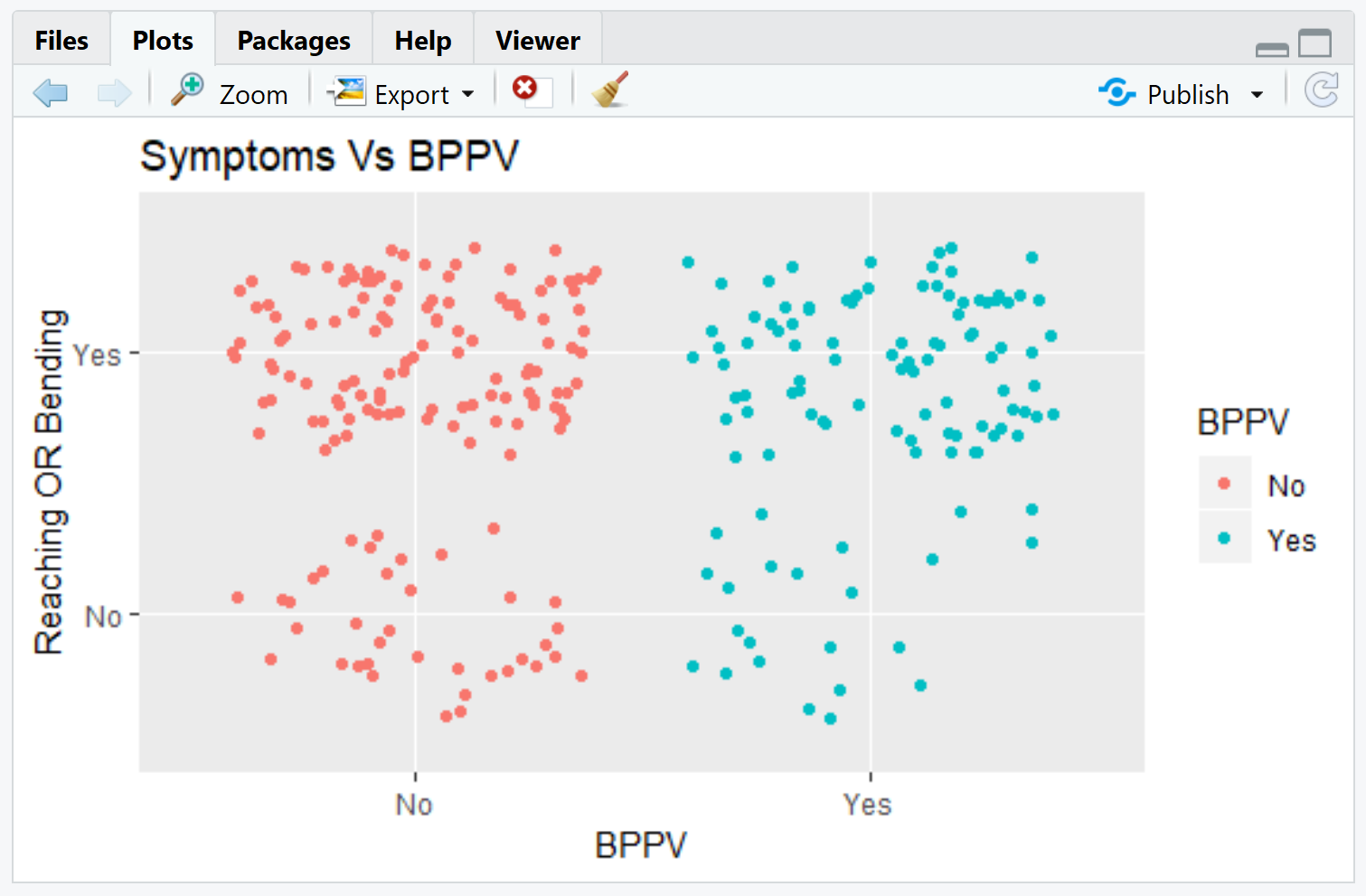
Below listed features seems to significantly affect the diagnosis of BPPV. These features are Age, initial symptoms, rolling over bed, Sitting or standing, Reaching out or Bending, Symptoms last duration less than 1 minute, headache throbs or pulses and sensitivity to light.

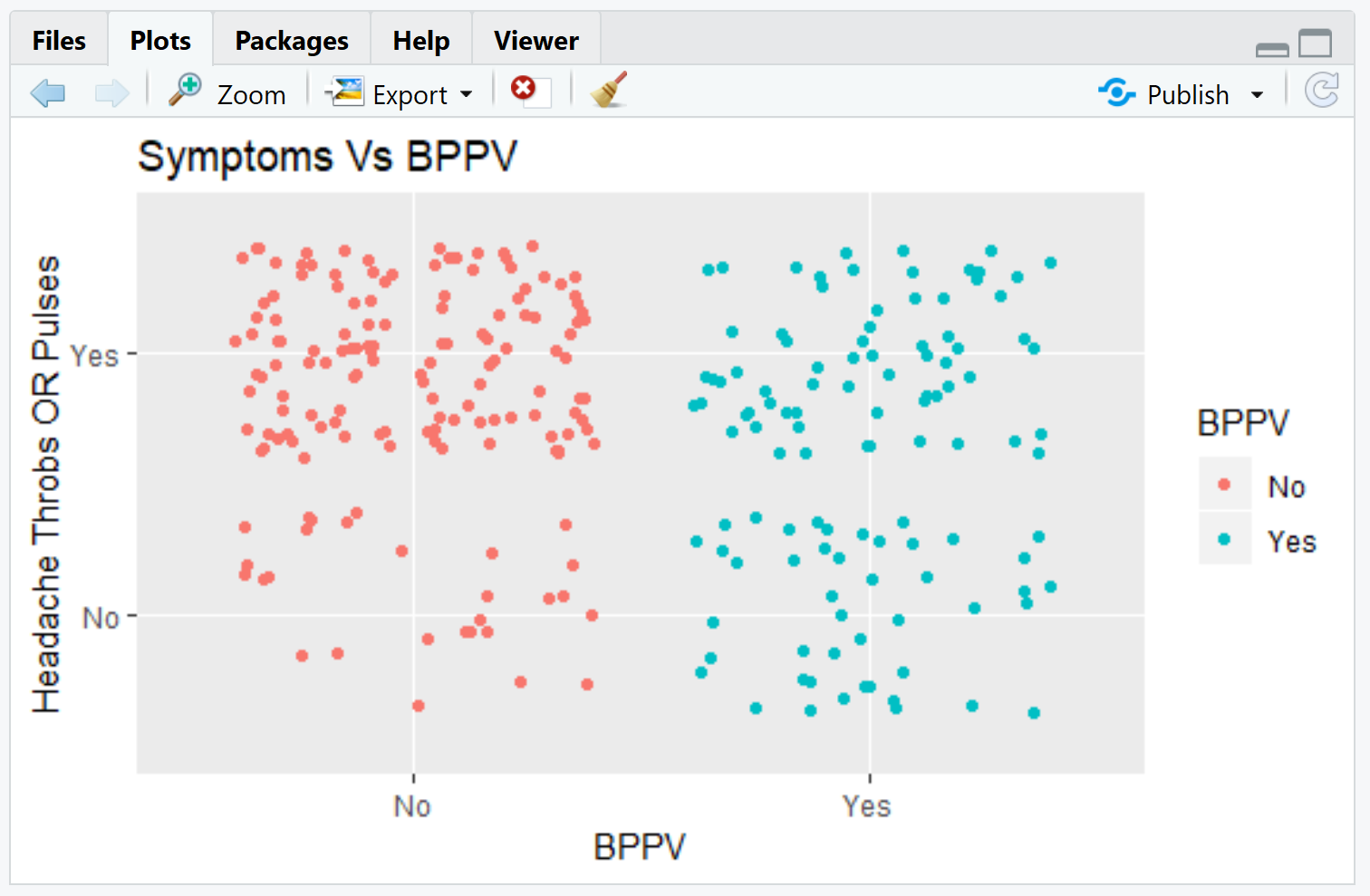


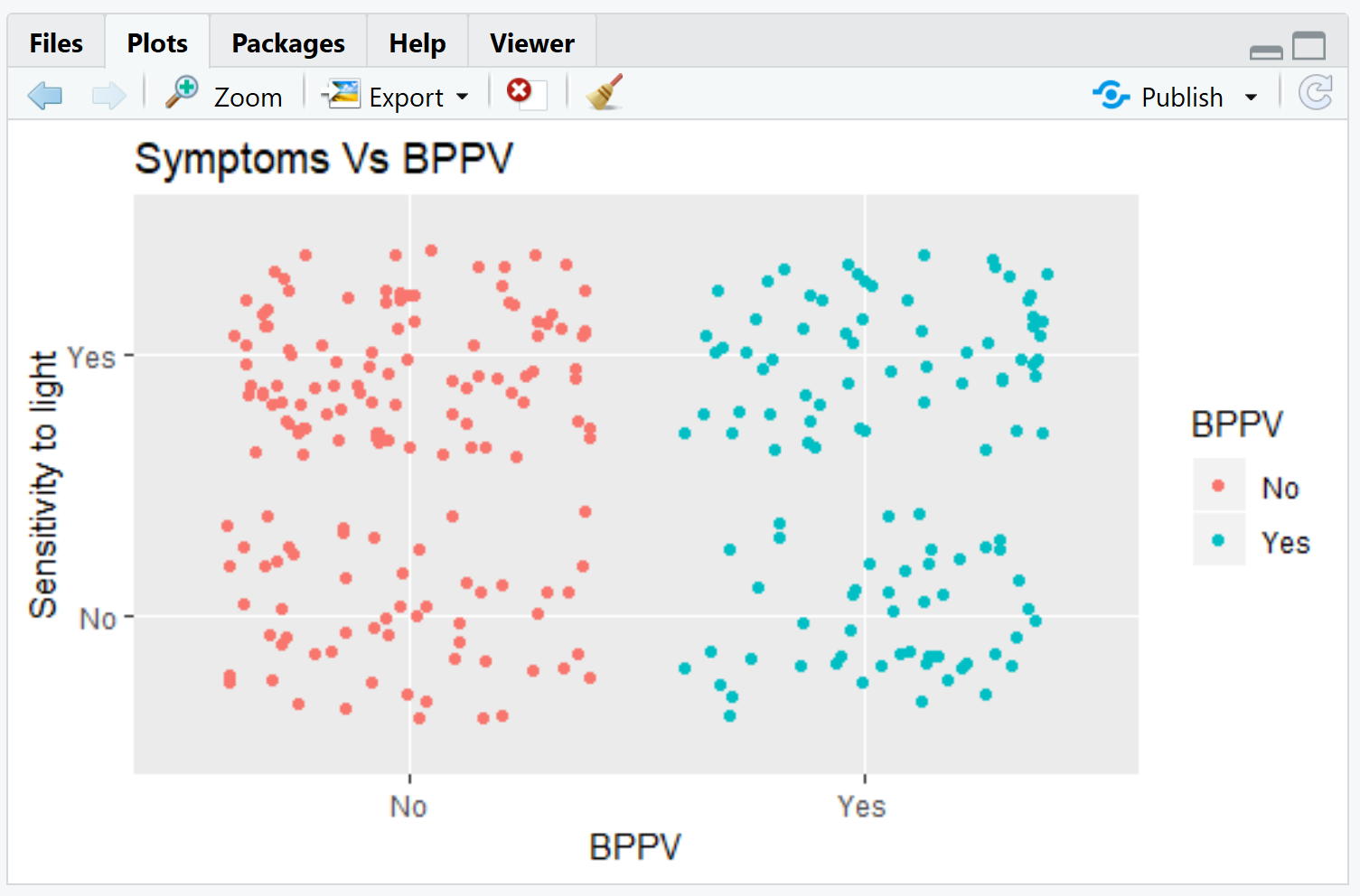


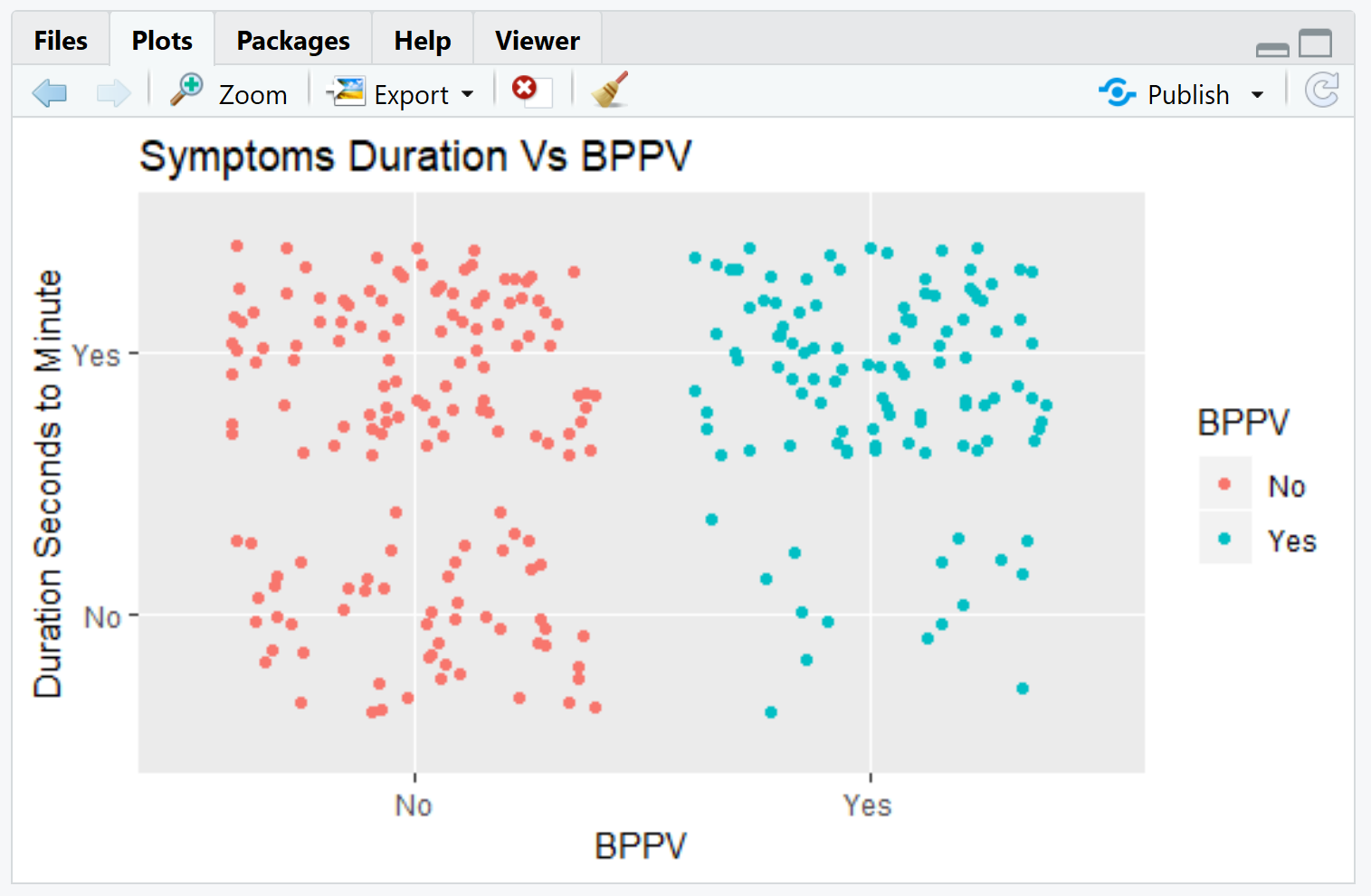






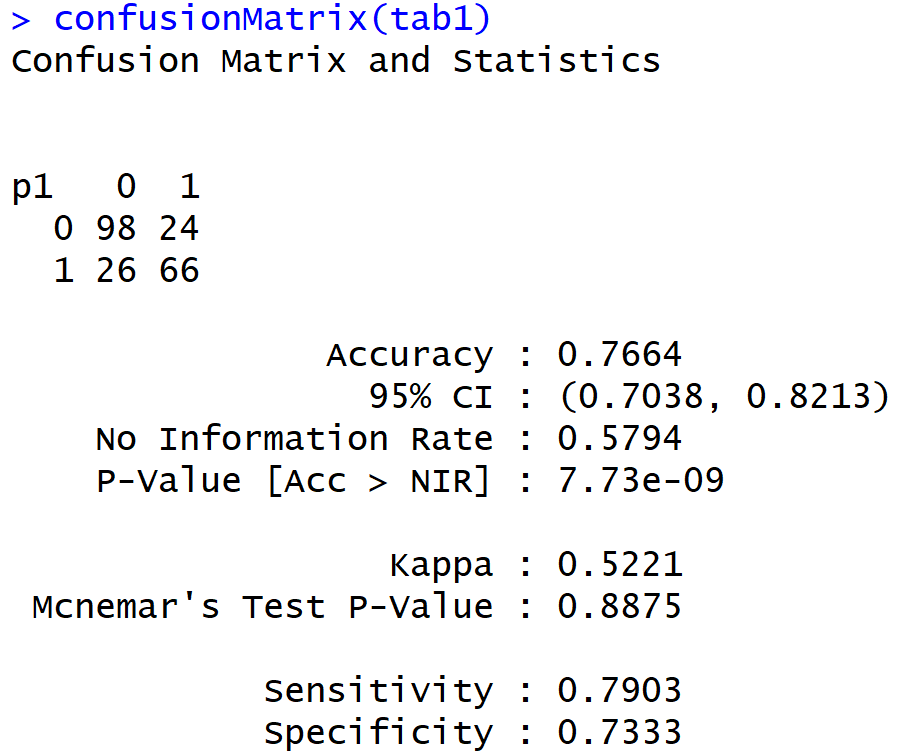


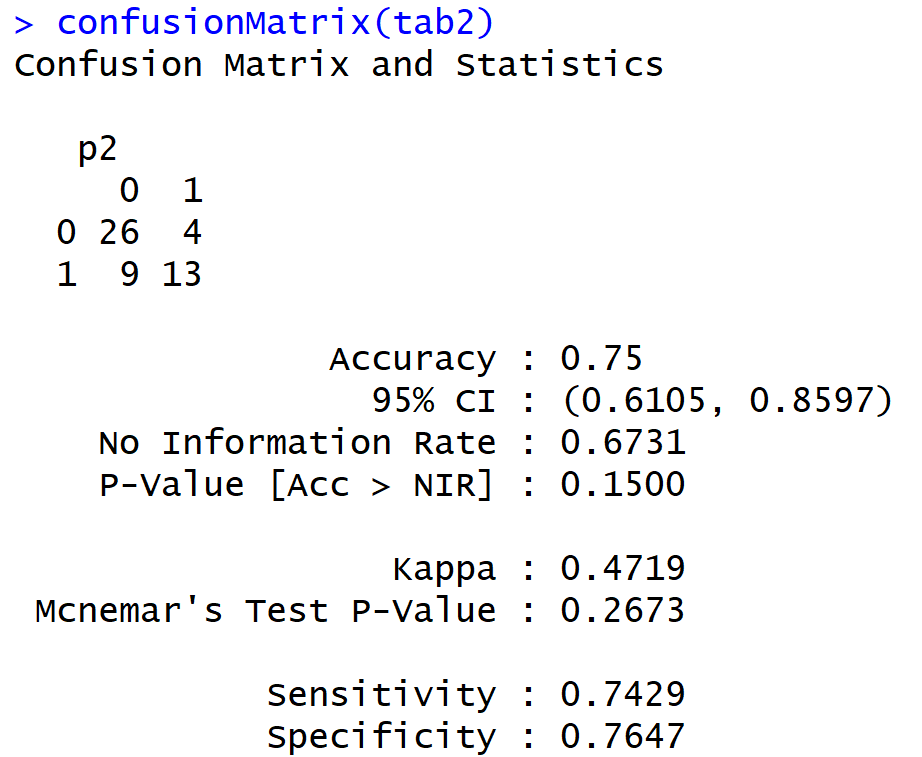




1. **Naive Bayes Results**.

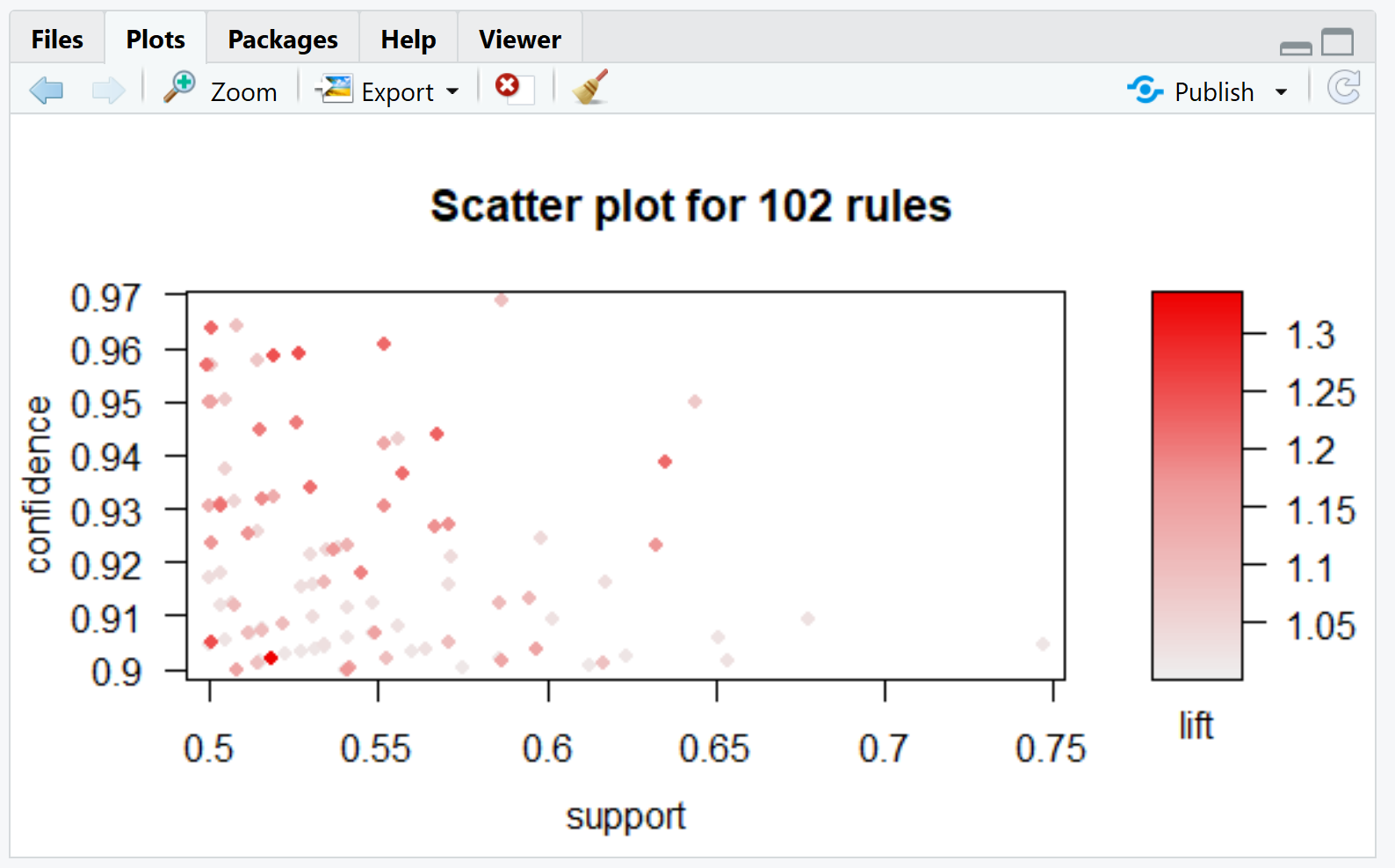
We performed Naïve Bayes evaluation on the selected variables(tab1) and on the whole dataset(tab2) and below were our results. Our selected feature model seems to have performed better than the whole dataset model.

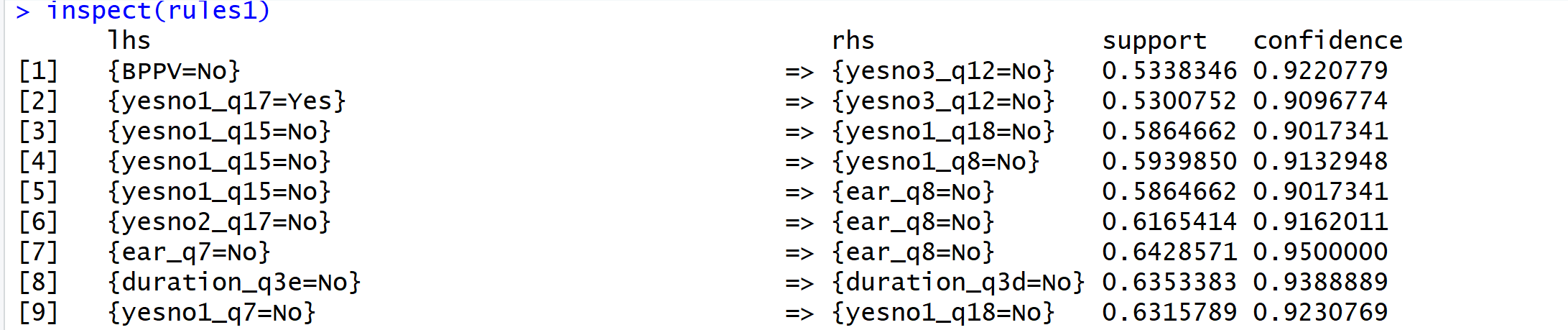




We were also interested in exploring association rules which could help us discover some interesting insight from the dataset. Our apriori analysis only indicated negative association rules like, Hip and Knee Replacement are associated with BPPV, Loud sounds had no association with BPPV etc.

Below is the graph for Apriori association rule graph and list of sample rules.





**VI Conclusions**

In this study, we employed four machine learning models in an attempt to select features to predict BPPV. All implemented models delivered similar results. The decision tree model was the earliest model designed to serve as a reference point for model performance. In the decision tree model, we achieved accuracy, sensitivity, and specificity of 0.73, 0.69 and 0.75 with 10-fold cross validation. In the ANN model, we achieved accuracy, sensitivity, and specificity of 0.71, 0.69 and 0.73. With 10-fold cross validation. In the Naïve Bayes model, with 80%-20% split, we achieved accuracy, sensitivity, and specificity of 0.75, 0.74 and 0.73. And with 10-fold cross validation, the accuracy dropped to 0.73. In the KNN model, with 80%-20% split, we achieved accuracy, sensitivity, and specificity of 0.56, 0.62 and 0.50.

The application of models shows that although KNN did not perform well for categorical data, with “yes” and “no” as the choices for most questions, it did demonstrate a fundamental BPPV characteristic, that being age. BPPV is typically not a condition affecting the young, onset usually occurs later in life. KNN demonstrated just that, confirming that those in their 60’s were significantly more likely to be diagnosed with BPPV. Using 10-fold cross-validation, the accuracy between decision tree, ANN and Naïve Bayes models are pretty close, i.e. in the range of 0.71-0.73. It indicates either a larger dataset is needed, or new ways to combine features is needed to improve predicate accuracy.

Overall, results indicate the need for further research. Although outcomes were not dismal, in the real world of medical diagnosis accuracy, sensitivity and specificity outcomes would need to be significantly higher to constitute a dependable model- values above 90% are typically deemed credible for medical diagnosis.

Future work would look at expanding the repository of current data with more survey results- perhaps by collaborating with another hospital and reducing features by combining prominent predictors into new features. Additional possibilities include expanding the existing dataset by randomly flipping some features; using autoencoder (neural networks) to combine features; and possibly imputing answers for phase three based on phase 1 and phase 2 responses.

**Contributions**

All team members participated in actively discussing project methodology and outcomes. Specific models were generated for project results as follows: John Awebwa generated K-NN models and results; Kunal Singh -Naïve Bayes; Hong Ding-data preprocessing and decision tree analysis; Dianne Williams-ANN analysis. Midway PowerPoint preparation was completed by Kunal Singh. Project report preparation, compilation and editing was completed by Dianne Williams and Hong Ding.

**References**

[1] S. Woodhouse, “BENIGN PAROXYSMAL POSITIONAL VERTIGO (BPPV),” *VEDA: Life Rebalanced*. [Online]. Available: https://vestibular.org/understanding-vestibular-disorders/types-vestibular-disorders/benign-paroxysmal-positional-vertigo. [Accessed: 20-Apr-2019

[2] H. A. Richburg, R. J. Povinelli , and D. R. Friedland , “Direct to Patient Survey for Diagnosis of Benign Paroxysmal Positional Vertigo ,” *2018 17th IEEE International Conference on Machine Learning and Applications*, Dec. 2018.

[3] W. O. Nijeweme-d’Hollosya, L. van Velsen, M. Poel, C. G. M. Groothuis-Oudshoornd, R. Soer, and H. Hermens, “Evaluation of three machine learning models for self-referral decision support on low back pain in primary care,” *International Journal of Medical Informatics*, vol. 110, no. 31, Feb. 2018.

[4] H. Ayeldeen, O. Shaker , G. Ayeldeen, and K. M. Anwar, “Prediction of Liver Fibrosis stages by Machine Learning model: A Decision Tree Approach∗,” *IEEE*, Nov. 2015.

[5] D. R. Friedland, S. Tarima, C. Erbe, and A. Miles, “Development of a Statistical Model for the Prediction of Common Vestibular Diagnoses,” *JAMA Otolaryngology Head Neck Surgery*., Apr. 2016.

[6] A. Bronshtein, “KNN makes predictions just-in-time by calculating the similarity between an input sample and each training instance.,” *Noteworthy- The Journal Blog*, 11-Apr-2017. .

[7] J. Brownlee, “Naive Bayes for Machine Learning,” *Machine Learning Mastery*, 11-Apr-2016. [Online]. Available: https://machinelearningmastery.com/naive-bayes-for-machine-learning/. [Accessed: 04-Apr-2019].

[8] “k-nearest neighbors algorithm,” Wikipedia, 21-Apr-2019. [Online]. Available: https://en.wikipedia.org/wiki/K-nearest\_neighbors\_algorithm. [Accessed: 18-Apr-2019].

[9] J. Han, M. Kamber, and J. Pei, “3 - Data Preprocessing,” in *Data Mining (Third Edition)*, Morgan Kaufmann, 2012, pp. 83–124. ISBN 9780123814791