Early Detection of Parkinson’s Disease Utilizing Vocal Analysis Through a Supervised Machine Learning Approach

Introduction

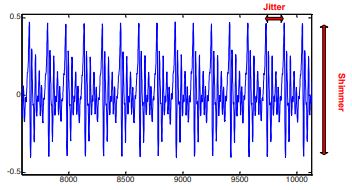
Parkinson’s disease is a progressive neurodegenerative disorder that affects millions worldwide. The deficiency in dopamine in the brain leads to major symptoms such as tremors in the face, legs, arms, and hands. Other major symptoms that result from the dopamine deficiency include difficulty swallowing, muscle rigidity, movement issues, impaired motor skills, and speech problems. This condition typically affects persons aged 45 and higher. Speech problems are a result of the vocal cord contractions resulting in a disturbed voice. The speech issues can often result in a higher mean pitch and greater variations in formants. The fluctuations that are observed Parkinson’s patients can be measured and quantified using jitter, shimmer, and formants because laryngeal muscles have difficulty staying in a consistent position.

In current clinical settings, Parkinson’s disease is often diagnosed in very late stages due to the latent manifestation of symptoms. By the time Parkinson’s disease is diagnosed, the damage to vital cells has already been done resulting in decreased effectiveness of current treatments. Because of the nature of the condition, It is key to detect Parkinson’s disease in earlier stages to develop and implement better treatments for patients. One such way that this challenge is being addressed is through signal analysis on certain vocal parameters.

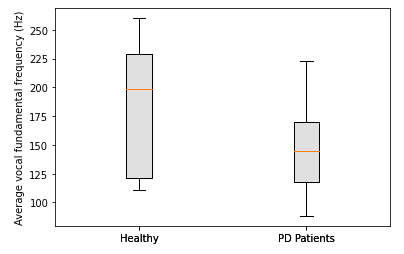
Because Parkinson's disease results in deterioration of vocal muscles, changes in certain vocal parameters can result. Measurement and analysis of some of these vocal parameters to diagnose Parkinson’s disease and other laryngeal pathologies is currently in use clinically and is referred to as “auditory perceptual analysis”. However, auditory perceptual analysis presents many challenges within the clinical realm. The results of the technique vary based on the practitioners’ personal experience when evaluating vocal parameters. Because of the subjectiveness of the method, there is a lack of consensus amongst medical professionals. Therefore, it is critical to find a technique that can evaluate vocal parameters objectively and without bias. Vocal parameters that are obtained in auditory perceptual analysis can describe the voice objectively and utilizing machine learning on these parameters to diagnose Parkinson’s disease can offer a possible solution to the biased interpretations for a more objective approach.

Data analysis and hypothesis

In order to create a model to accurately diagnose Parkinson's disease, many features should be used to help the model’s predictions. Although many features will be utilized for this model, there are certain key features to be aware of when the algorithm is making predictions. Features such as formants, jitter, shimmer and HNR (harmonic to noise ratio) will play a critical role for the algorithm’s predictions and will also provide valuable insights into the data.

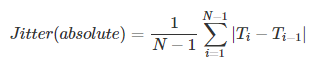


Formants by definition are spectral peaks that are observed in the voice. Formants are pivotal in speech because it helps us distinguish words and vowels through frequency components. Although many formants are applied in the model (Max/min vocal fundamental frequency etc.) average fundamental vocal frequency will be a key feature for the algorithm and gaining insights into the data. Fundamental vocal frequency is defined as the number of times a sound wave that is produced by the vocal cords repeats in a given time period and is measured in Hertz.

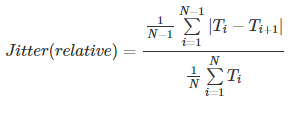


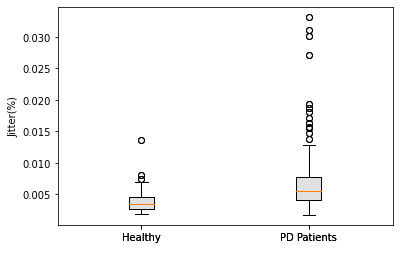
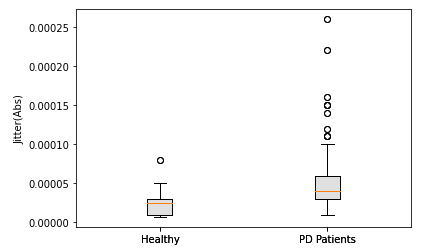
The average fundamental vocal frequency between Parkinson’s disease patients and healthy persons appears to vary greatly. Although the distribution is tighter and the average frequency is lower in Parkinson’s disease patients the distribution of healthy persons varies greatly. The variance observed in healthy persons can possibly be explained by differences in gender, age, state of mind during the recording, and lifestyle.

Jitter is defined as the measure of cycle to cycle variations of fundamental frequency. Jitter is a key feature that can help indicate the presence of Parkinson's disease because Parkinson's disease patients have a difficult time controlling their vocal muscles compared to their healthy counterparts leading to expected higher variance. The 2 primary measures of jitter that will be used as features are jitter(absolute) and jitter(relative). Jitter(absolute) is defined as the average absolute difference between consecutive periods (cycle to cycle variation in fundamental frequency) and is expressed as



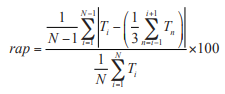
Similar to jitter(absolute), jitter(relative) is defined as the variation in fundamental frequency divided by the average period and is expressed as (usually in a percent form)

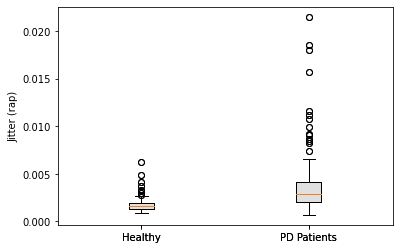




Because Parkinson's disease patients have issues controlling their vocal muscles, the jitter values have a tendency to be higher than in healthy persons. In many cases, there is immensely high outlier jitter values in Parkinson's disease patients. Healthy persons have a very tight distribution of jitter values and mostly fall within a small range due to better control of vocal muscles. The distinction in jitter values outside of the few outliers observed in healthy persons can make the algorithm much more likely to predict the presence of Parkinson's more accurately by observing this tendency.

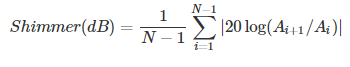
Jitter RAP is another parameter that can provide insight into the data and prove useful for a feature. Jitter RAP is defined as is a measurement of the average disturbance or the average difference between one period and its neighboring periods and is expressed as:

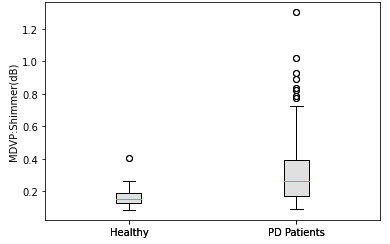




Similar to the other measures of jitter, Parkinson's disease patients tend to have much higher disturbance in their voice most likely due to lack of control over vocal muscles. Many large outliers are present in Parkinson's patients and detection of these abnormally high disturbance values can help aid in the algorithm’s prediction of the presence of Parkinson's disease.

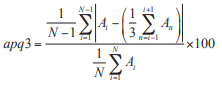
Shimmer is defined as the cycle to cycle variation in amplitude. Similar to jitter, shimmer it can also be a good indicator of the presence of Parkinson's disease because of the lack in control of the vocal cords can result in higher variations of amplitude. Shimmer is reported in many variations that will be used as features for the model. Shimmer (dB), is comparable to jitter(absolute) except shimmer(dB) measures variations in amplitude and is expressed as

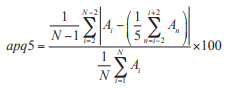


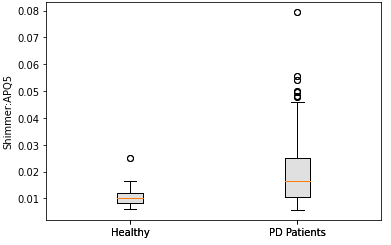
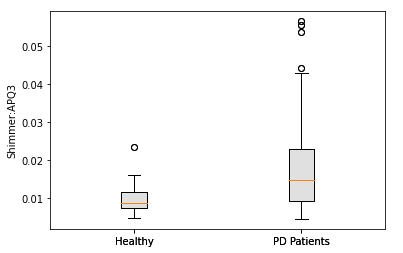


As observed in jitter parameters, shimmer in Parkinson's disease patients display much higher shimmer values with some very large outliers. Healthy persons have a very tight distribution when it comes to shimmer values when compared to Parkinson's patients with a much wider distribution and higher decibel values. These wider and heightened shimmer values can of great use to influence the predictions of the algorithm.

Shimmer APQ3 and APQ5 are defined as measurements of amplitude disturbance within neighboring 2 and 4 periods respectively. They are calculated by taking the average difference in amplitude of a period and the mean amplitude of its closest 2 and 4 neighbors respectively, then divided by the average amplitude. They are expressed as:

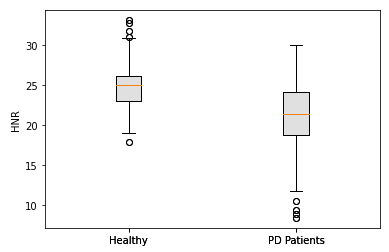






Just like Jitter RAP, the shimmer APQ values are higher for Parkinson's disease patients. Healthy persons have a very small amount of disturbance in their voices with a very tight distribution. A higher disturbance is to be expected for Parkinson's disease patients however, it is more apparent that lack of control over vocal muscles creates a distinguishment between APQ3 and APQ5. Because of the reduced control over laryngeal muscles greater disturbances tend to seem more apparent over time and are reflected with greater occurrences of outliers in APQ5.

HNR or harmonic to noise ratio is a vocal parameter used to help determine the efficiency of the voice. In a general sense, it is a comparison of the amount of air that is pulled from the lungs and converted to energy in the form of vibrations of the vocal cords. HNR can also be defined as an evaluation of the vibration of the vocal cords compared to the glottal noise that is expressed in dB. Typically a higher HNR is associated with health because It denotes the efficiency of the conversion of air from the lungs to vibrational energy from the vocal cords.



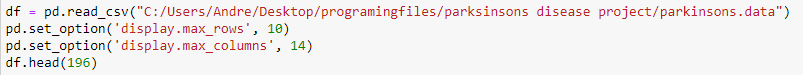
Objectively speaking, healthy persons have a higher average HNR than their Parkinson's disease counterparts denoting their efficiency in producing speech. Many Parkinson's disease patients have extremely low HNR because the vocal cords have a

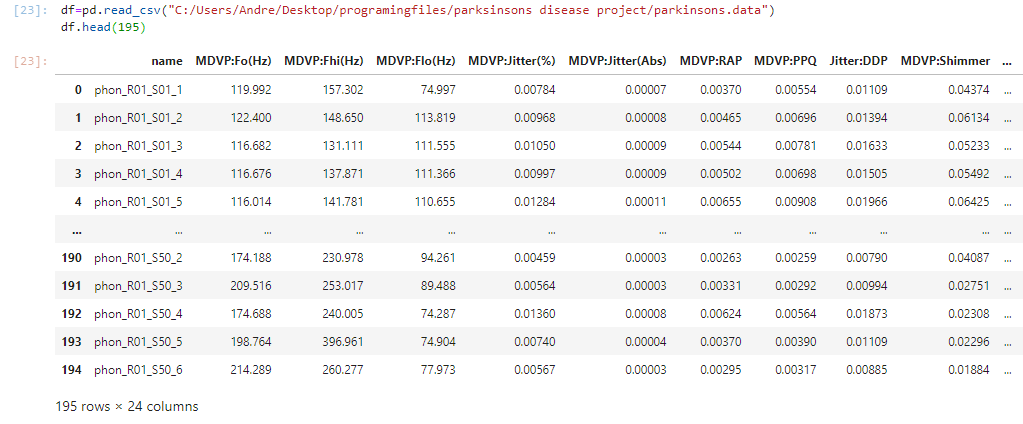
more difficult time generating speech from vibrational energy. Many outliers that are significantly lower than average show that HNR can help boost the predictive capability of the model because of this distinction that exists between healthy persons and Parkinson's disease patients.

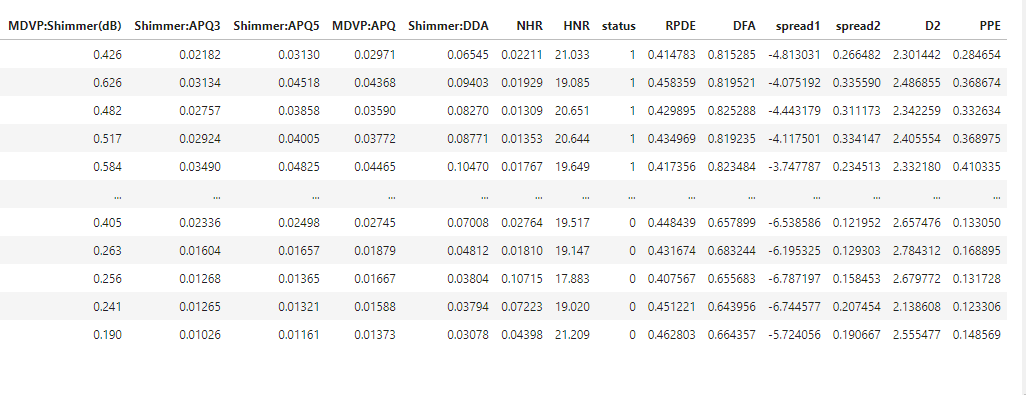
Although there are a relatively small number of occurrences, based on the comparisons and contrasts in the displayed vocal parameters above; the algorithm should still be able to accurately detect the presence of Parkinson's disease. Furthermore, the relatively high number of features in addition to only requiring a single true/false output or label can also contribute significantly to the accuracy of the predictions made by the algorithm.

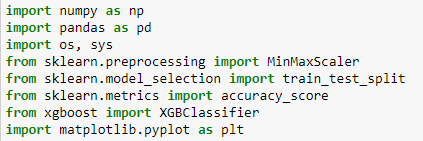
Method and Approach

The dataset that is used to create this model originates from a study done at the University of Oxford in collaboration with the National Centre for Voice and Speech in Denver, Colorado measuring vocal parameters of healthy and Parkinson's disease patients (data and syntax displayed below).



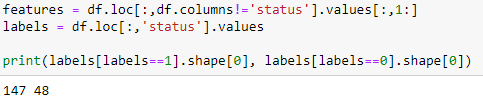






All the libraries that were utilized are listed above with the appropriate syntax displayed. Numpy and pandas were used to convert the data into a data frame that is readable by the algorithm. Scikit learn was utilized to scale the label (in this case status to either a 1 or a 0), split the dataset into training and testing groups and to test the accuracy of the model. Matplotlib was used to create the charts that were previously shown to investigate the data and provide insight.

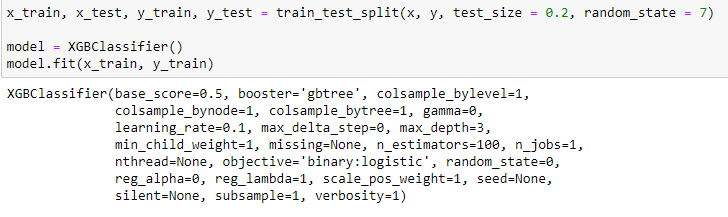
The algorithm that was used to create the model is XGboost. XGboost is an algorithm that works based on decision trees and is designed for speed and performance in applied machine learning. XGboost is free and readily available across many languages such as C++, R, Java, and python. XGboost offers many strengths and is used for regression and classification problems (such as this one). The advantages of this algorithm allow for the handling of small and large datasets, continued training of an already fitted model, and the capability to deal with missing data values. The combined features of availability, flexibility, speed, and performance with classification problems make XGboost an ideal algorithm for early detection of Parkinson's disease through vocal signal parameters.



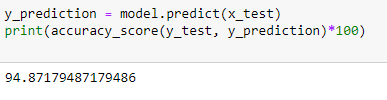
Now that the data has been read as a data frame, features and labels can be defined. Every column except ‘status’ will be used as a feature. Meanwhile, the column labeled ‘status’ will be our label the algorithm will try to predict. (0 for healthy and 1 for a Parkinson's patient). In addition to assigning features and labels the number of occurrences for healthy and Parkinson's disease patients is displayed.



The data frame now needs to be fitted and scaled so that the algorithm can make its predictions. The features need to be normalized and scaled to a range. The fit\_transform() command fits the data and scales it to a given range (in this case -1 to 1 to normalize the data)



Now that the data has been fitted, the data is randomly split into 2 groups. 80% of the data will be used to train the model while the other 20% will be used to test the model’s accuracy in making predictions of either a 1 (Parkinson’s disease patient) or 0 (healthy person). XGboost is then used as a classifier to train the model by using decision trees and ensemble learning.



After the model has been trained, the accuracy was tested with the remaining 20% of the data. The accuracy of roughly 95% is a very great outcome given the number of occurrences in the data. With a higher number of occurrences and the allowed continued training that XGboost offers this model can be improved upon fairly easily and have even stronger outcomes.

Conclusion

The model is very accurate being able to correctly predict the presence of Parkinson's disease with a 95% predictive capability. These results are very impressive given that the dataset had roughly 200 occurrences, was trained off roughly 160 occurrences and predicted roughly 38 of the 40 tested samples correctly. This very high predictive capability is most likely due to the distinction in vocal parameters that were previously displayed in the box and whiskers charts. Many of the vocal parameters that were investigated displayed various outliers and objectively greater differences in magnitude between Parkinson's disease patients and healthy persons. Although the model is very accurate as is, the model can easily be improved with more data. Higher occurrences can definitely help the model to be much more accurate in its predictions as well as be expanded it its utility. Because the model was built utilizing XGboost, continued training of the model with high speed and performance can be accomplished fairly easily and can be widely distributed and implemented because of its open availability to everyone across the internet.

References

Vikas and R. K. Sharma, "Early detection of Parkinson's disease through Voice," *2014 International Conference on Advances in Engineering and Technology (ICAET)*, Nagapattinam, 2014, pp. 1-5, doi: 10.1109/ICAET.2014.7105237.

L. Jeancolas *et al*., "Automatic detection of early stages of Parkinson's disease through acoustic voice analysis with mel-frequency cepstral coefficients," *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, Fez, 2017, pp. 1-6, doi: 10.1109/ATSIP.2017.8075567.

Teixeira, J. P., Oliveira, C., &amp; Lopes, C. (2013). Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters. Procedia Technology, 9, 1112-1122. doi:10.1016/j.protcy.2013.12.124

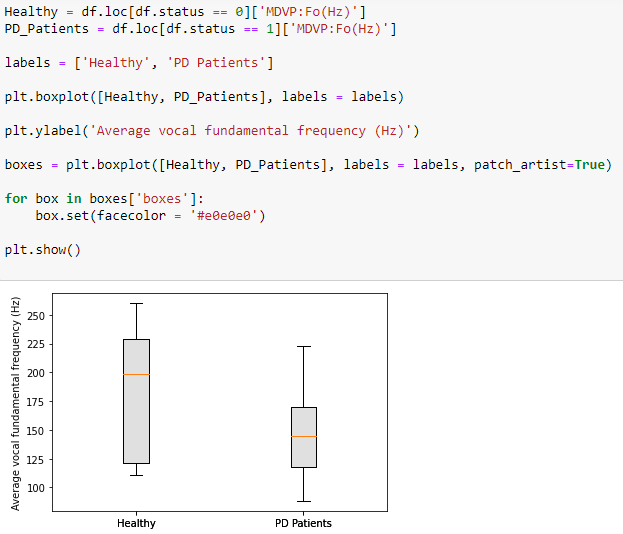
Max A. Little, Patrick E. McSharry, Eric J. Hunter, Lorraine O. Ramig, “*S*uitability of dysphonia measurements for telemonitoring of Parkinson's disease”, 2008 IEEE Transactions on Biomedical Engineering (to appear).

Max A. Little, Patrick E. McSharry, Stephen J. Roberts, Declan A.E. Costello, Irene M. Moroz, “Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection”, 2007 BioMedical Engineering OnLine 2007

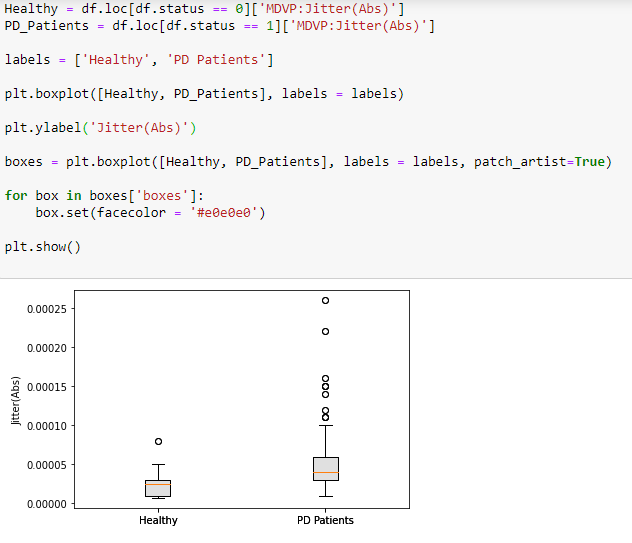
G Fant, "Acoustic Theory of Speech Production. Mouton & Co" in , Netherlands:The Hague, 1960.

Appendix

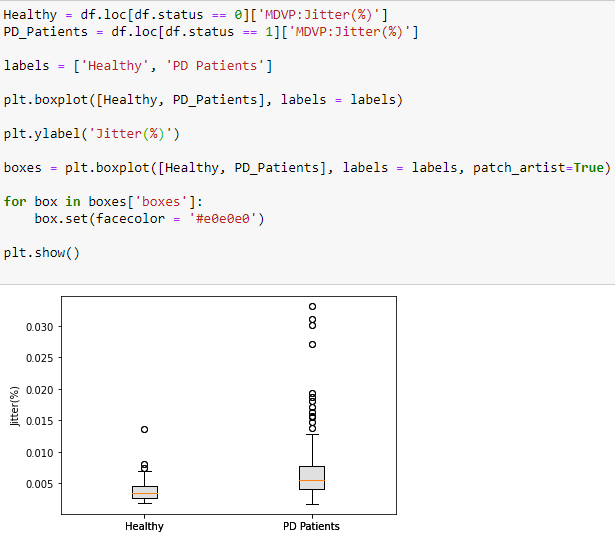
Syntax for the Average fundamental vocal frequency chart



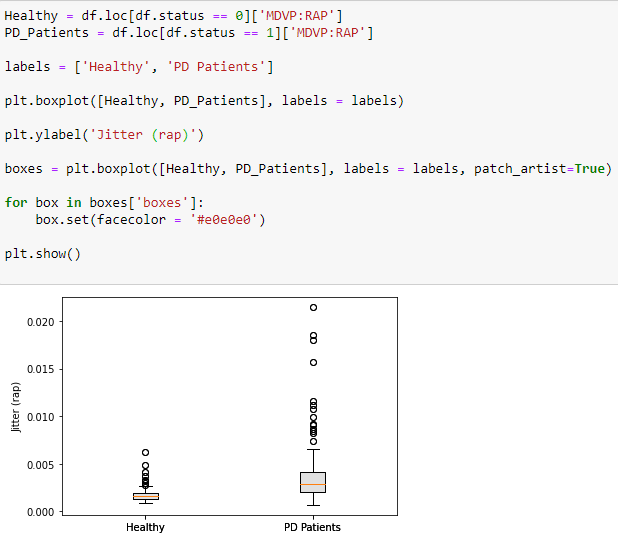
Syntax for the Jitter(abs) chart



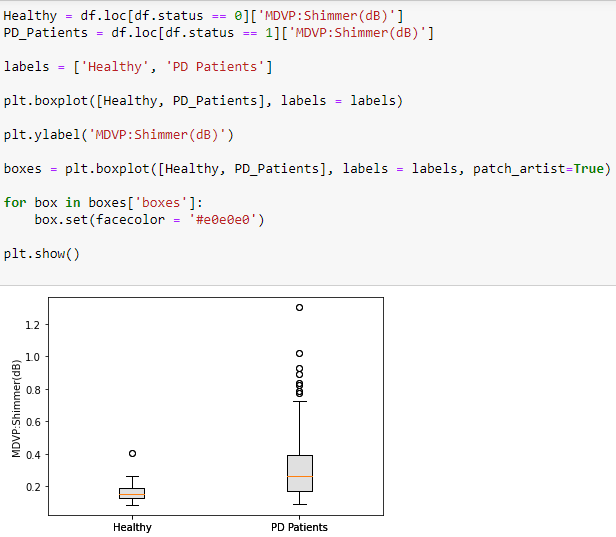
Syntax for the Jitter(%) chart



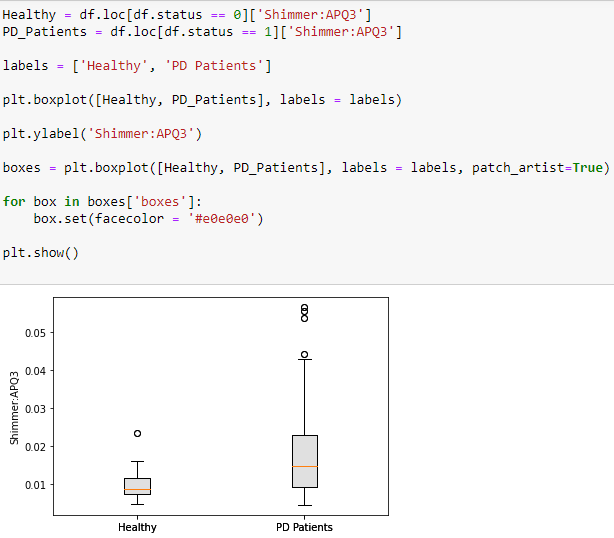
Syntax for the Jitter RAP chart



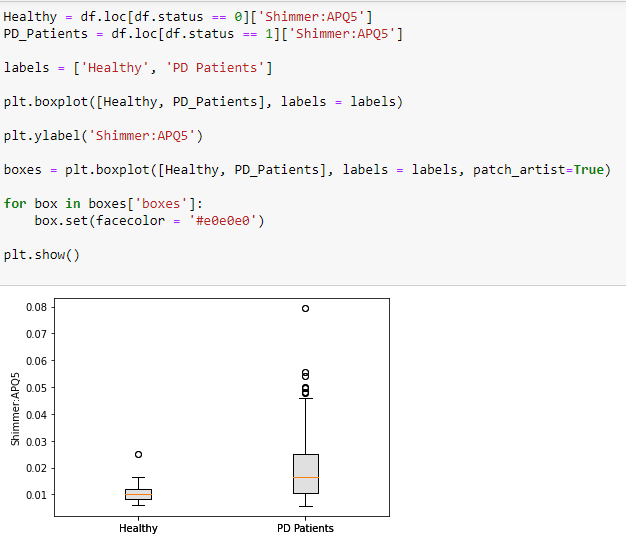
Syntax for the Shimmer(dB) chart



Syntax for the Shimmer:APQ3 chart



Syntax for the Shimmer:APQ5 chart



Syntax for the HNR chart

