NAVEEN BALARAJU

UNIVERSITY AT BUFFALO

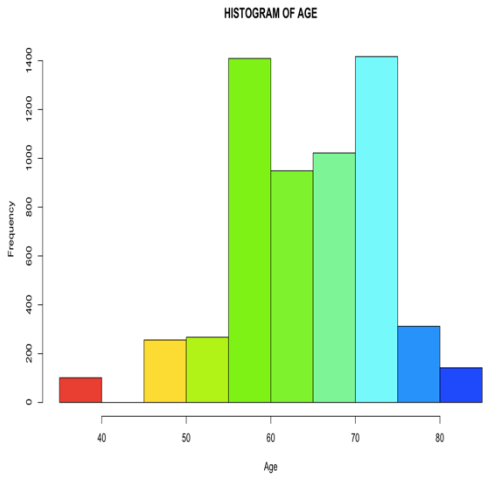
GOAL: To implement Clustering on Parkinson's Telemonitoring dataset and characterize "Jitter" Related variable for a new patient with a High UPDRS score

DATASET DESCRIPTION:

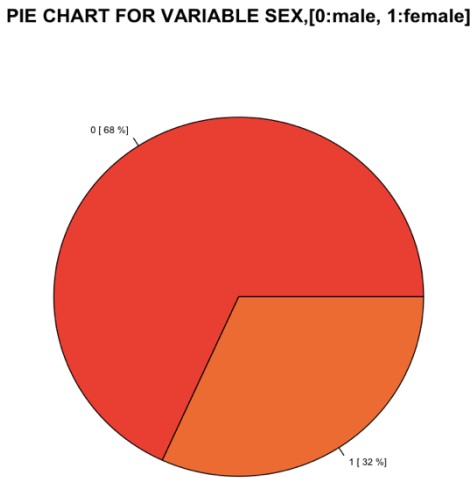
1. Subject # - Integer that uniquely identifies each subject
2. age - Subject age
3. sex - Subject gender '0' - male, '1' - female
4. test_time - Time since recruitment into the trial. The integer part is the number of days since recruitment.
5. motor_UPDRS - Clinician's motor UPDRS score, linearly interpolated
6. total_UPDRS - Clinician's total UPDRS score, linearly interpolated
7. Jitter (%), Jitter (Abs), Jitter: RAP, Jitter: PPQ5, Jitter: DDP - Several measures of variation in fundamental frequency
8. Shimmer, Shimmer(dB), Shimmer:APQ3,Shimmer:APQ5,Shimmer:APQ11,Shimmer:DDA - Several measures of variation in amplitude
9. NHR, HNR - Two measures of ratio of noise to tonal components in the voice
10. RPDE - A nonlinear dynamical complexity measure
11. DFA - Signal fractal scaling exponent
12. PPE - A nonlinear measure of fundamental frequency variation

EXPLORATORY DATA ANALYSIS:

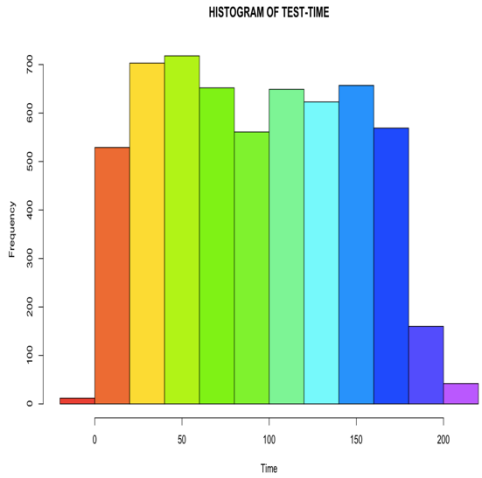
1. Analysis of variable Age - The Histogram shows that approximately 90% of the population is of age 60 and above, while only 10% is below age of 60.



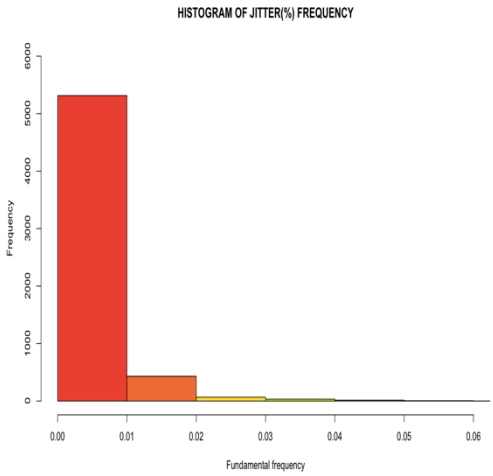
2. Analysis of variable Sex - The dataset contains 68% percent of male population and 32% female.



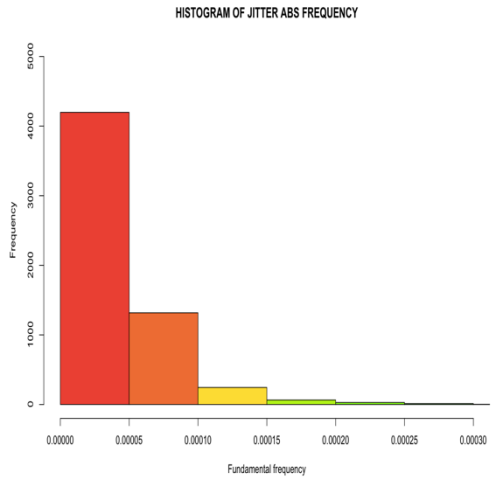
3. Analysis of variable Test_time - There are only 732(12.45% of the population) patients who are under the clinical trial for less than 25 days and majority of the patients are undergoing this trial for more than 25 days.



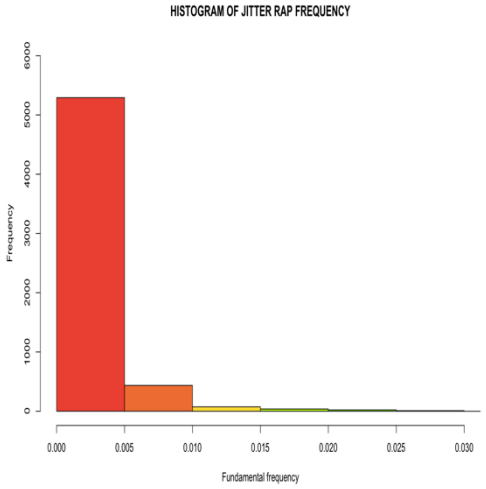
4. Analysis of variable Jitter % frequency-The histogram shows that 93.6% of the patients have the jitter% frequency between 0 - 0.01. Whilst, only a small percentage of patients have the jitter% greater than 0.01.



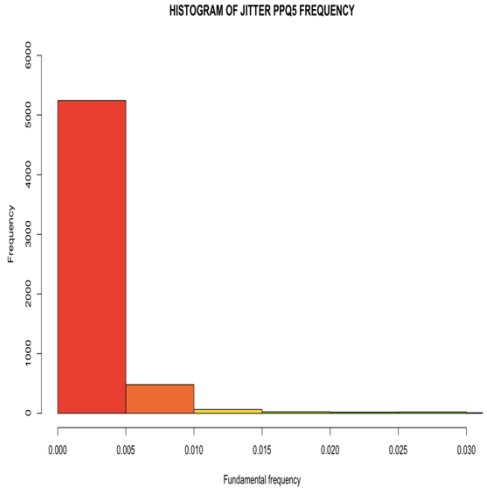
5. Analysis of variable Jitter Abs frequency - The histogram shows that 70% of the patients have Jitter Abs frequency between 0 and 0.00005, while rest 30% have above 0.00005.



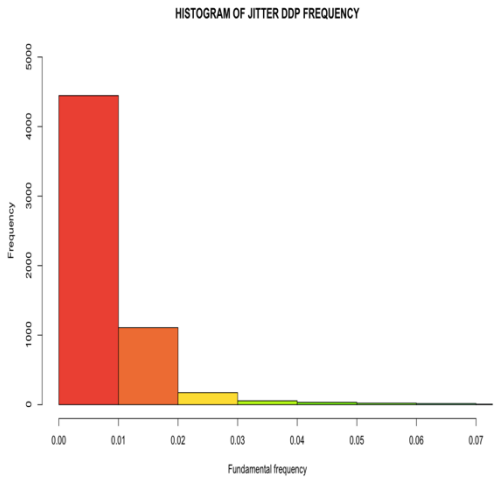
6. Analysis of variable Jitter rap frequency - The histogram shows that 88.5% of the patients have Jitter Rap frequency between 0 and 0.005, while rest have above 0.005.



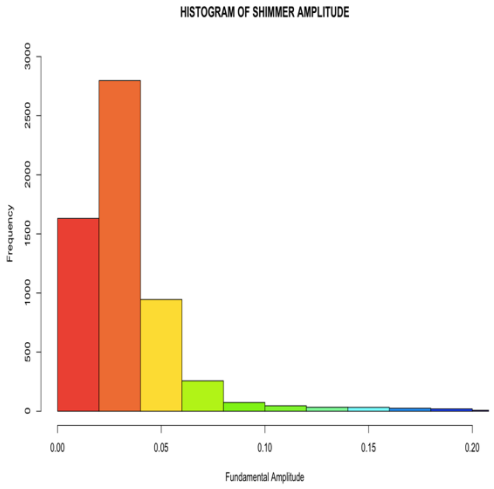
7. Analysis of variable Jitter PPQ5 frequency-The histogram shows that 88.5% of the patients have Jitter PPQ5 frequency between 0 and 0.005, while rest have above 0.005.



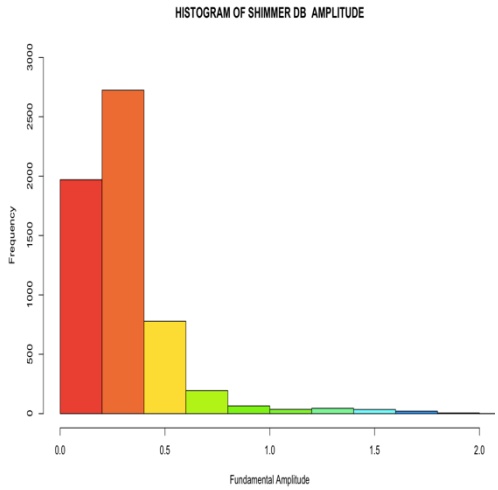
8. Analysis of variable Jitter DDP frequency- The histogram shows that 76% of the patients have Jitter DDP frequency between 0-0.01 and 17% have frequency between 0.01-0.02. while rest 7% have frequency above 0.02.



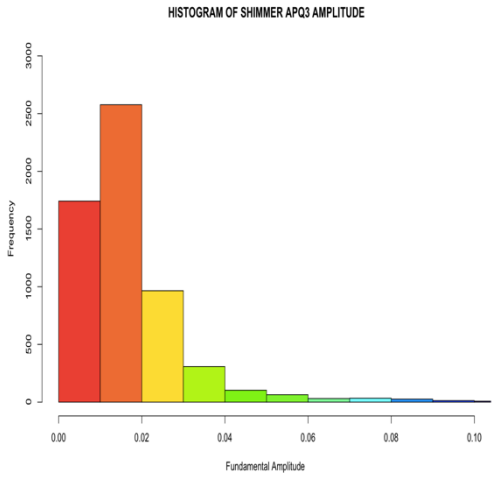
9. Analysis of variable Shimmer - The histogram shows that about 85% of the patients (5032) have the shimmer Amplitude less than 0.05. While, the rest 155 have Amplitude above 0.05.



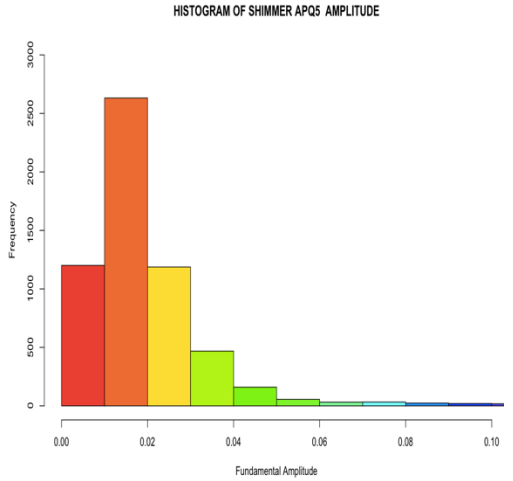
10. Analysis of variable Shimmer Db Amplitude-About 89% of the patients have shimmer DB Amplitude less than 0.5 and the rest 11% have above 0.5.



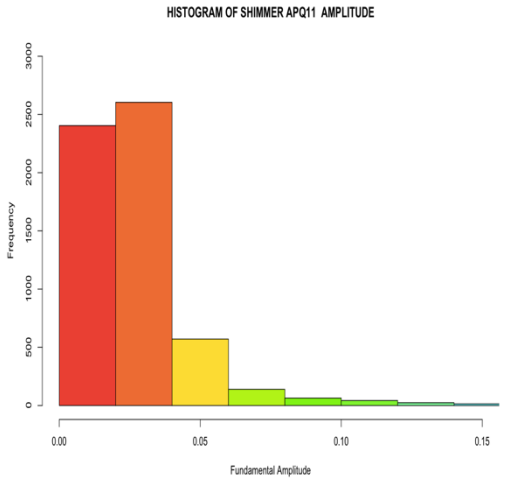
11. Analysis of variable Shimmer APQ3 -The histogram shows that about 73% of the patients have Shimmer APQ3 Amplitude less than 0.02 and the rest 27% have shimmer APQ3 Amplitude above 0.02.



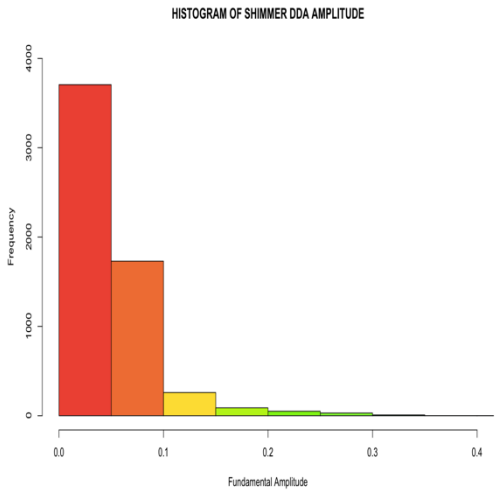
12. Analysis of variable Shimmer APQ5- The histogram shows that 65% of the patients have Shimmer APQ5 Amplitude between 0-0.02 and rest of the patients have above 0.02.



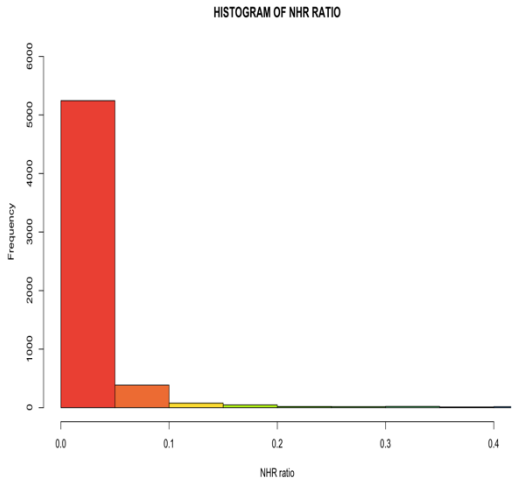
13. Analysis of variable Shimmer APQ11 - The data shows 92% of the patients have the Shimmer APQ11 Amplitude of 0.05 or less.



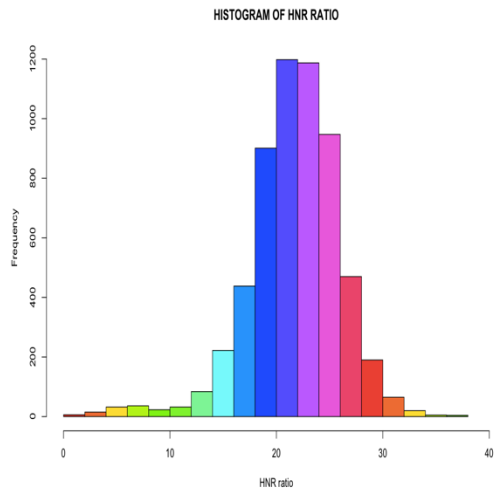
14. Analysis of variable Shimmer DDA- The histogram shows that 92.5% of the patients have the Shimmer DDA Amplitude of 0.1 or less.



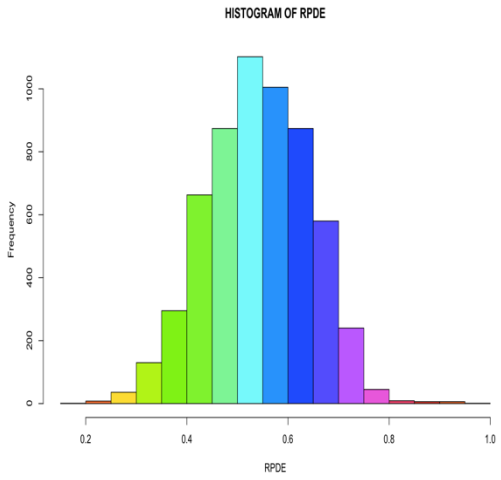
15. Analysis of variable NHR - The histogram shows that about 89% of the patients have NHR ratio of 0.05 or less.



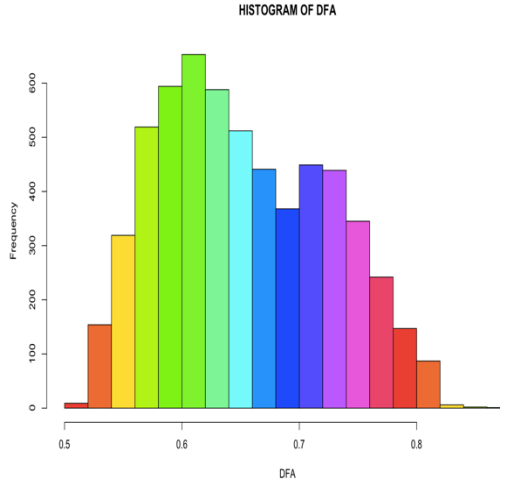
16. Analysis of variable - The histogram for the HNR ratio shows a normal distribution with majority (58%) of the patients have HNR ratio between 19 and 25.



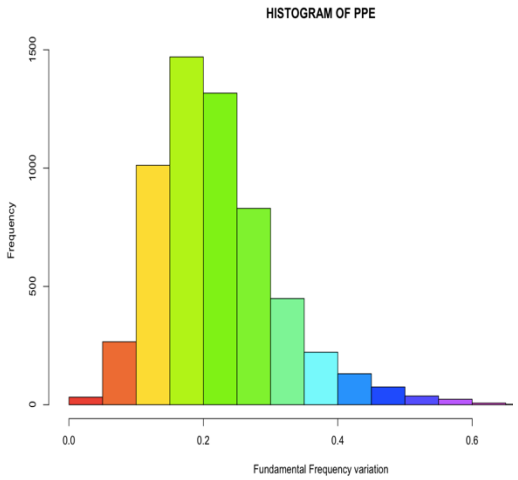
17. Analysis of variable RDPE - The histogram of RPDE show a normal distribution centered at RDPE value of 0.54 and 52% of the patients have the RPDE value between 0.46 and 0.61.



18. Analysis of variable DFA - The Signal fractal scaling exponent shows a normal distribution centered at 0.65.

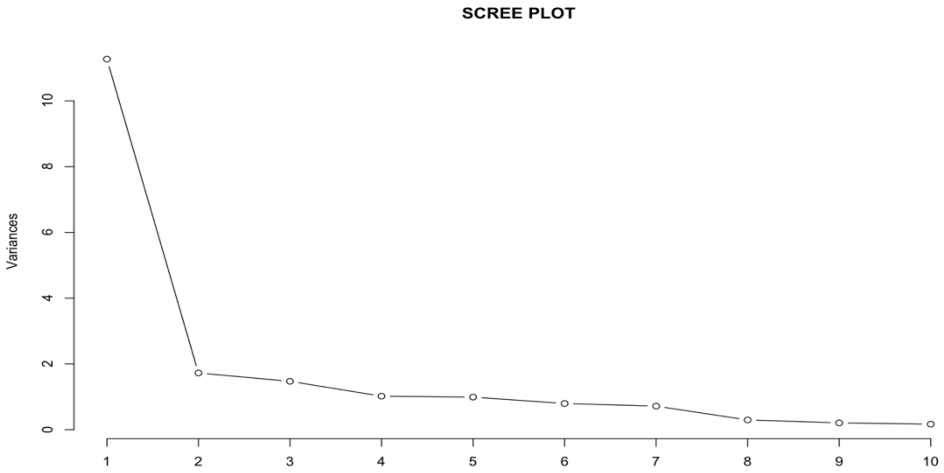


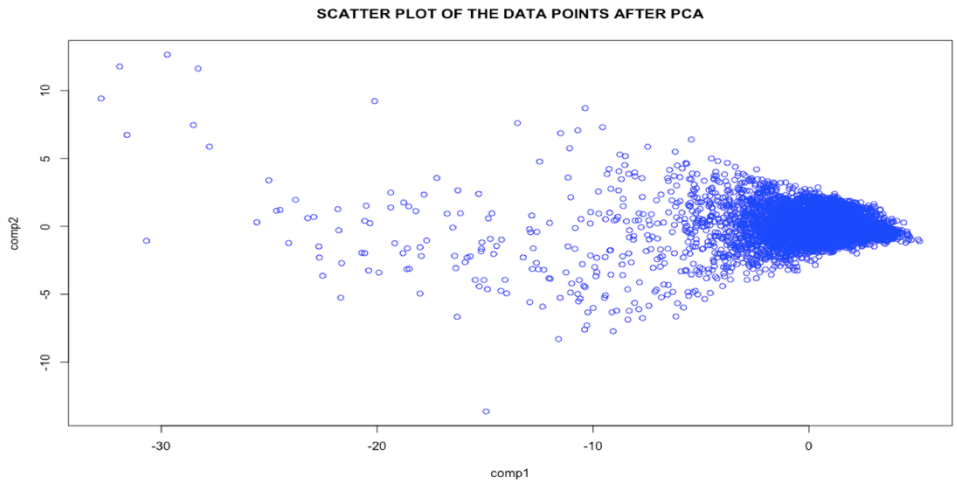
19. Analysis of variable PPE - About 95% of the patients have PPE fundamental variation of 0.1 and above.



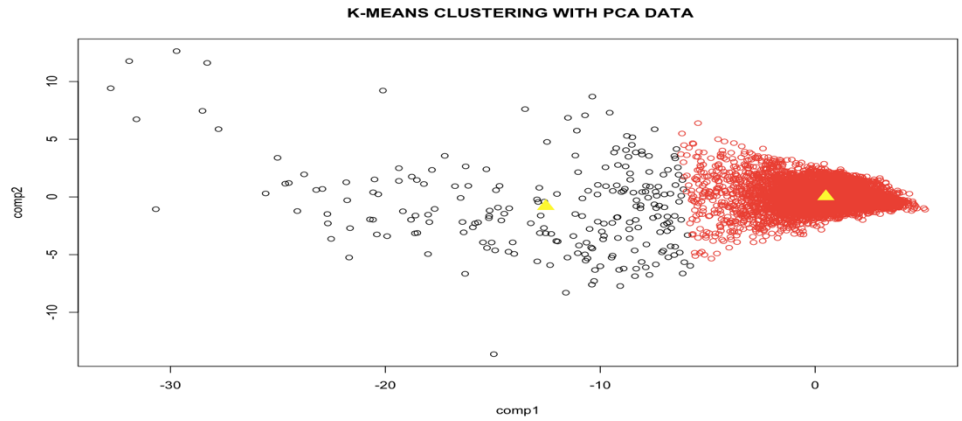
CLUSTERING:

I've performed PCA on the Parkinson's telemonitoring dataset and have used first two principal Components for the further analysis.

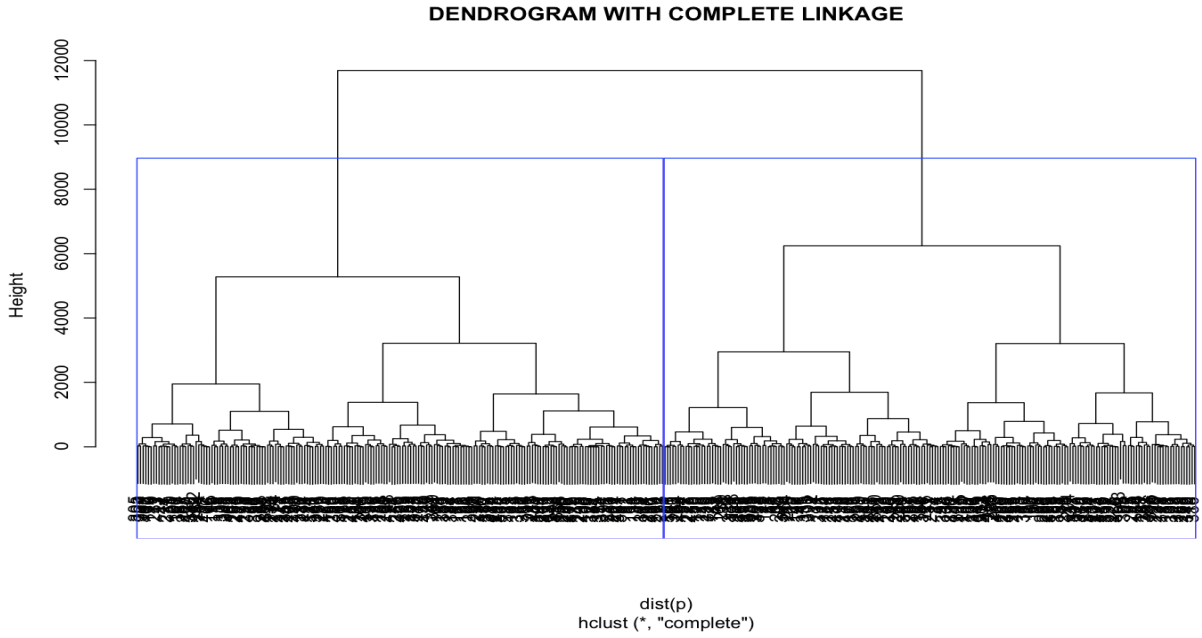




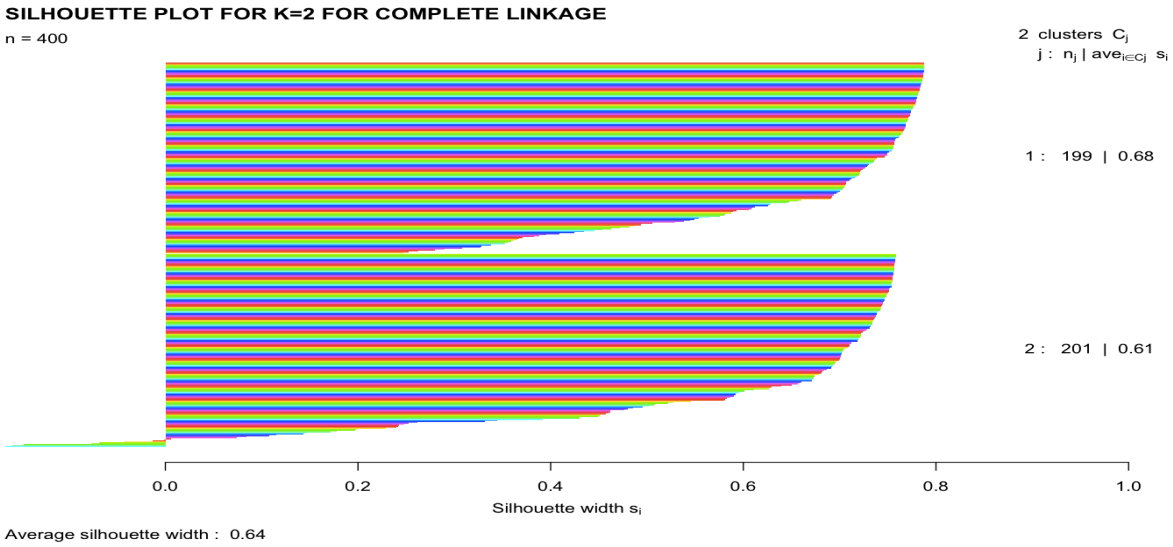
K-Means clustering:



Hierarchical clustering with complete linkage:



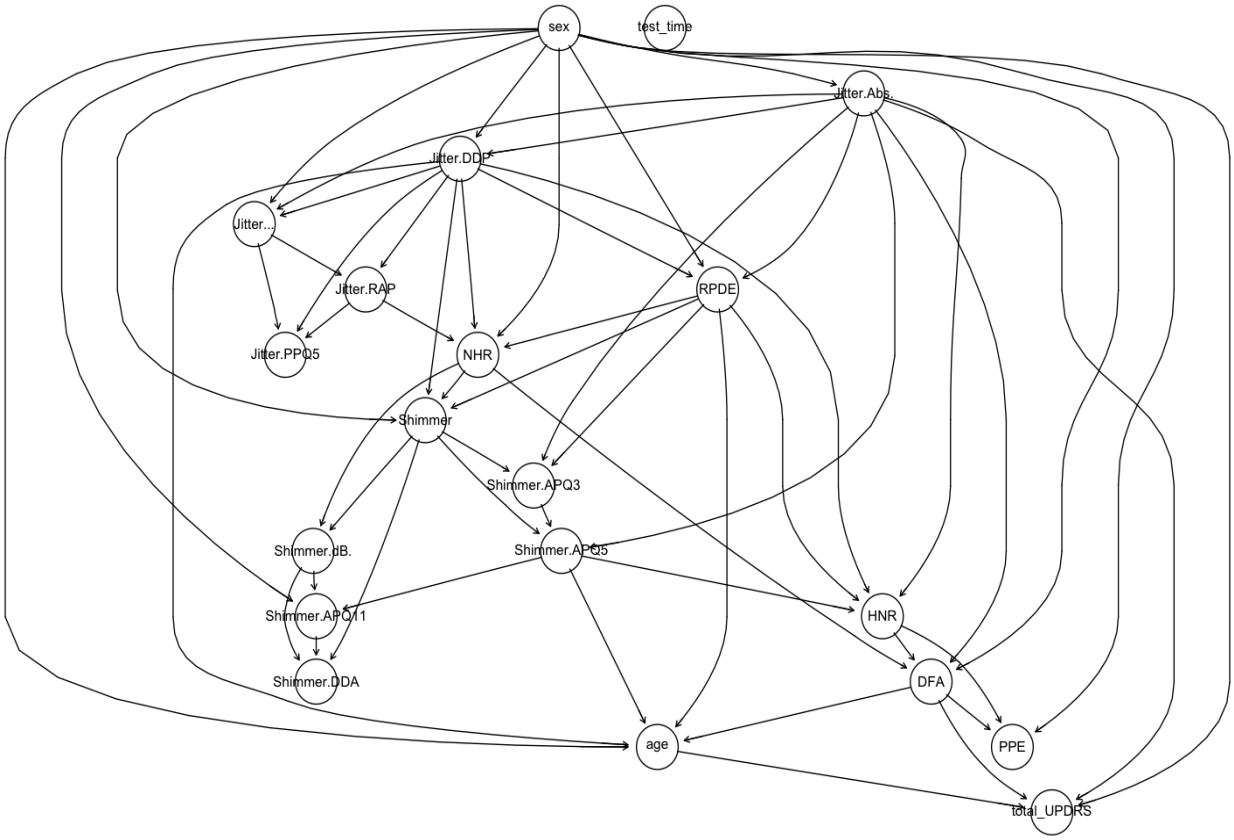
Silhouette plot:



After analyzing the results given by K-means and Hierarchical clustering we get two cluster that captures “motor_UPDRS” and “total_UPDRS. Furthermore, Silhouette plot gives a good average silhouette width of 0.64.

BAYESINA NETWORK:

I've binned the variables into different levels and since our goal is to characterize the jitter related variables for patient with high UPDRS score (above 2 standard deviation above mean). I've binned the total_UPDRS values into 2 levels. All the values>= 50.42 i.e two std above mean are considered a High UPDRS. Furthermore, I've have used Hill climbing algorithm with "BIC" as the score measure to find the optimal DAG.



The Model generates few interesting results.

1.None of the female patients have a High UPDRS scores.

2.For a Male of age >=60, having high jitter abs frequency and high DFA, there is only 3% chance of having high UPDRS score.

3.For a Male of age >=60, having low jitter abs frequency and high DFA, there is 23.4% chances of having high UPDRS score.

4.For a Male of age >=60, having high jitter abs frequency and low DFA, there is only 5% chance of having high UPDRS score.

5.For a Male of age >=60, having low jitter abs frequency and low DFA, there is only 5% chances of having high UPDRS score.