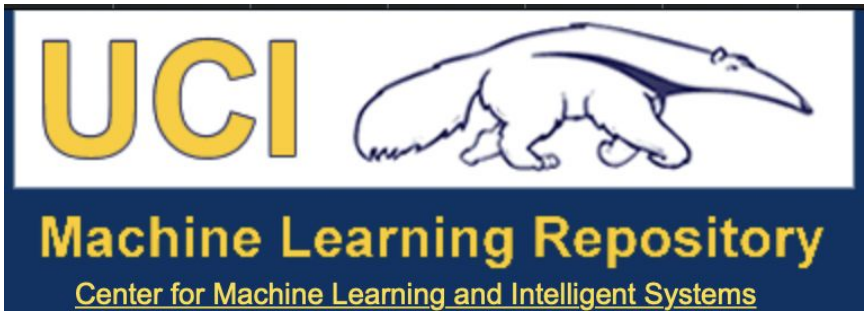
Predicting a Parkinson’s Diagnosis

Data Analyst: Rebecca Lorey

Data Set - General Information

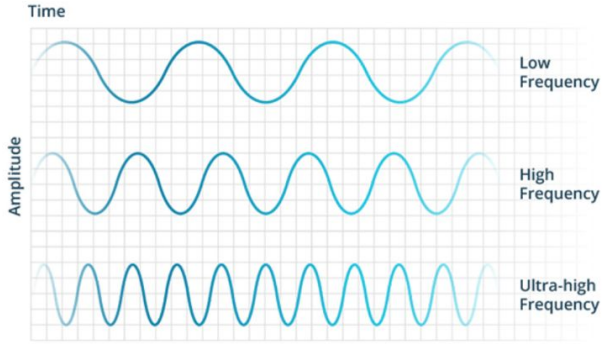
- UCI Machine Learning Repository
- Collected by Max Little at the University of Oxford
- MDVP - Multidimensional Voice Program
- 31 people
- 23 diagnosed with Parkinson’s
- 195 voice recordings



Data Set - Features

Positive Diagnosis of Parkinson's

Biomedical Voice Measurements



- 23 Different Features
- Various Vocal Frequencies
- Measures of Variation in Fundamental Frequency
- Measures of Variation in Amplitude
- Ratio of Noise to Tonal Components in Voice

Data Cleaning

- No Missing Values
- No Duplicates
- No Categorical Data/No discrepancies

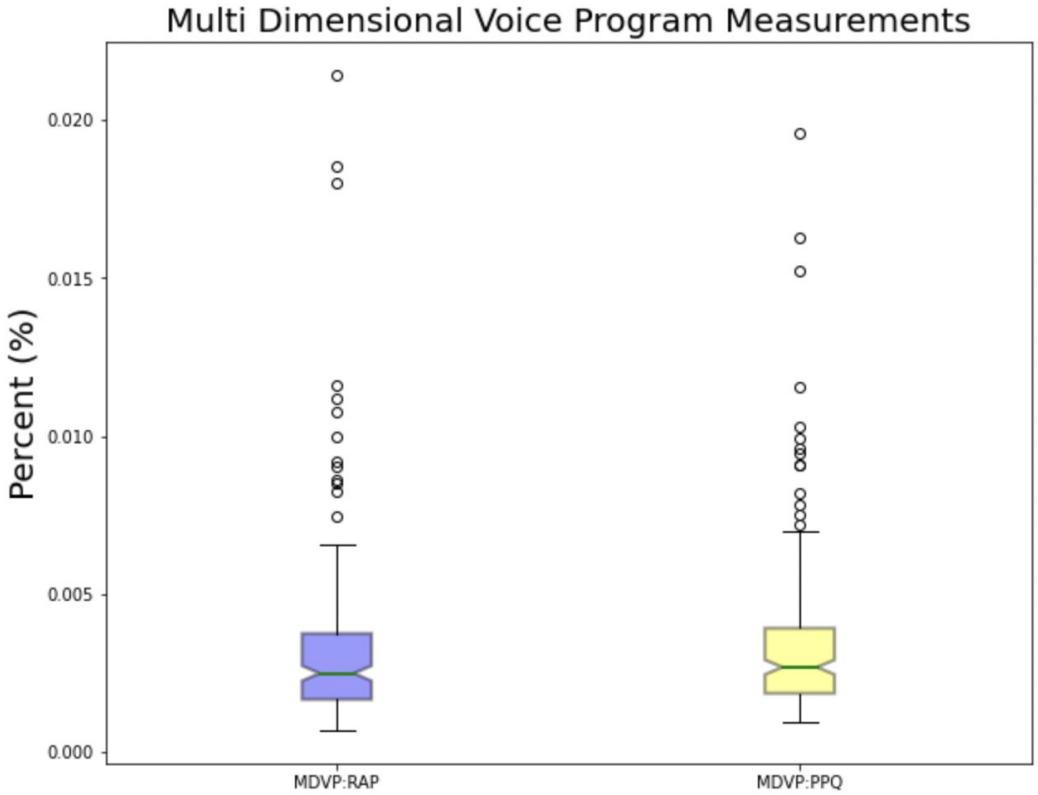
Average Results of Each Subject

Delete extra row of data (_7) for three individuals

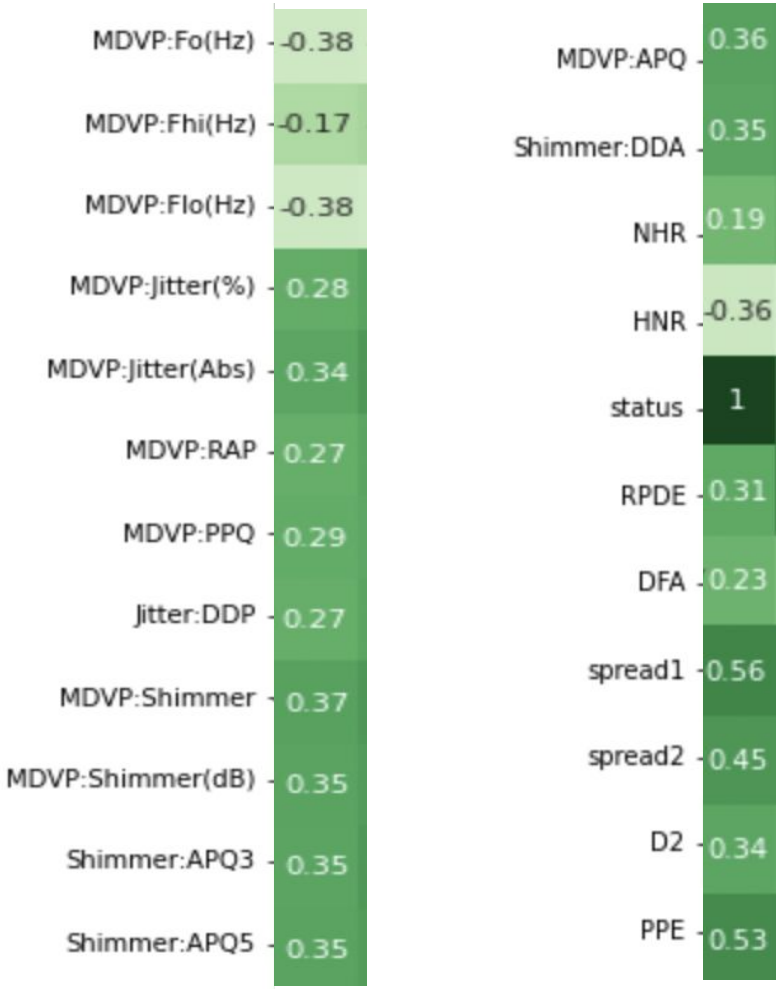
```
phon_R01_S20_1
phon_R01_S20_2
phon_R01_S20_3
phon_R01_S20_4
phon_R01_S20_5
phon_R01_S20_6
phon_R01_S21_1
phon_R01_S21_2
phon_R01_S21_3
phon_R01_S21_4
phon_R01_S21_5
phon_R01_S21_6
phon_R01_S21_7
```

Average the six samples for each feature per individual

Univariate Analysis



Heatmap: Correlations



Next Steps:

- Identifying and removing outliers
- Modeling