Predicting a Parkinson's Diagnosis

Data Analyst: Rebecca Lorey

<u>Data Set - General Information</u>

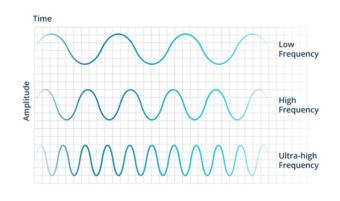
- UCI Machine Learning Repository
- Collected by Max Little at the University of Oxford
- MDVP Multidimensional Voice Program
- 31 people
- 23 diagnosed with Parkinson's
- 195 voice recordings



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

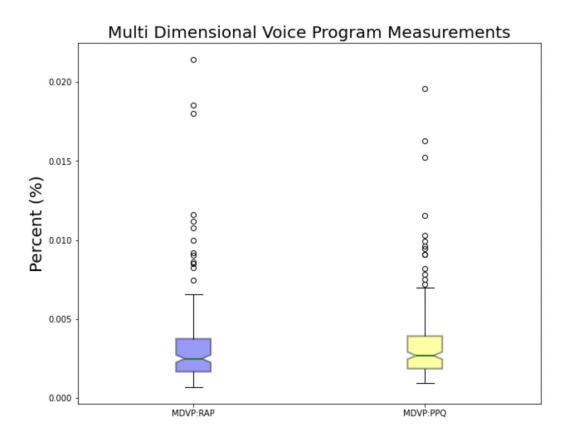
Delete extra row of data (_7) for three individuals

```
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

MDVP:APQ -	-0.30	MDVP:FO(HZ)
Shimmer:DDA - 0.35	-0.17	MDVP:Fhi(Hz)
NHR - ^{0.19}	-0.38	MDVP:Flo(Hz)
HNR -0.3	0.28	MDVP:Jitter(%)
status - 1	0.34	MDVP:Jitter(Abs)
RPDE - 0.31	0.27	MDVP:RAP
DFA -0.23		MDVP:PPQ
spread1 -0.56		Jitter:DDP
spread2 - 0.45		MDVP:Shimmer
D2 - _{0.3} 4		MDVP:Shimmer(dB)
PPE -0.53		Shimmer:APQ3
0.53	0.35	Shimmer:APQ5

Next Steps:

- Identifying and removing outliers
- Modeling

