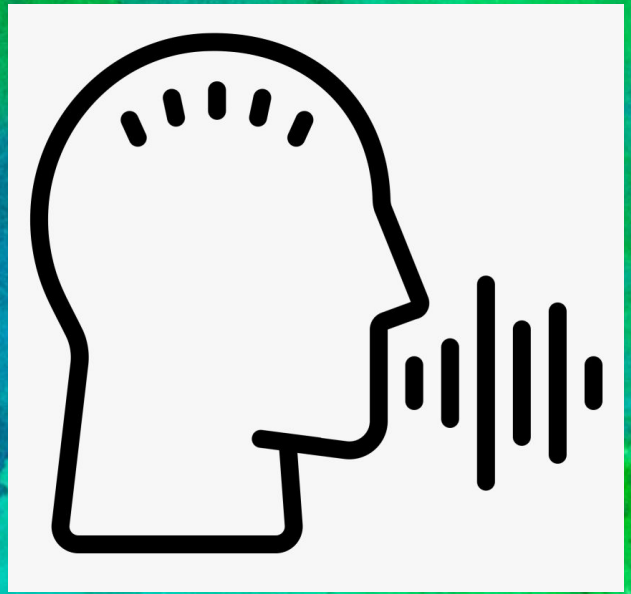
The Relationship between Speech Acoustics and Gender

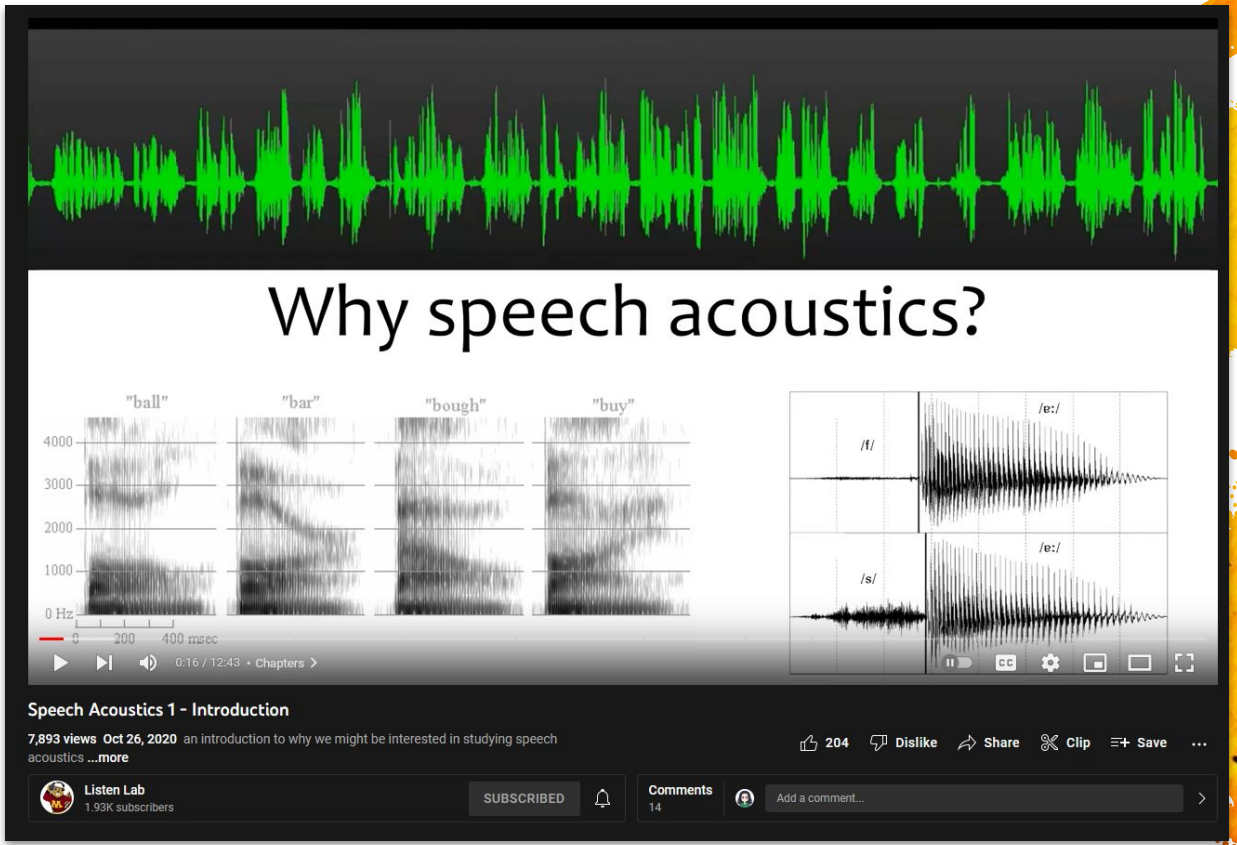


Tech Talent South

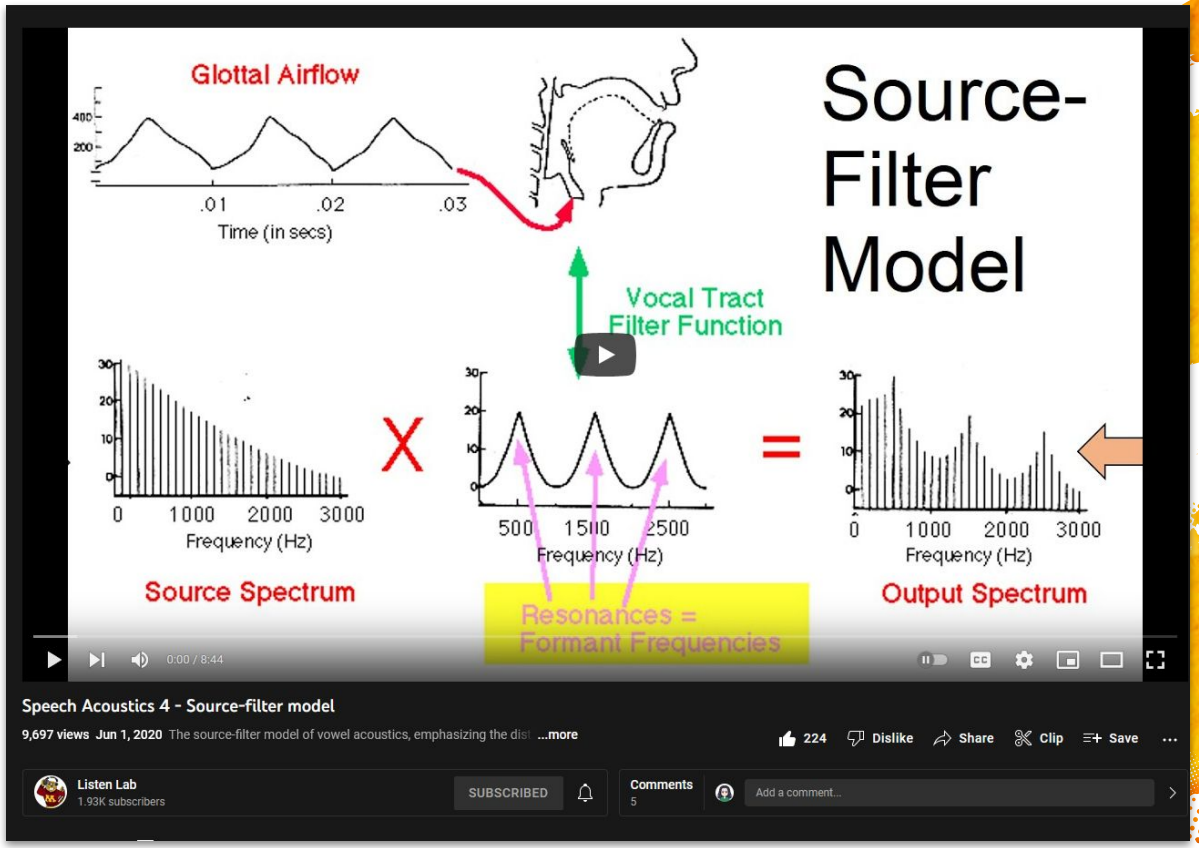
Maria Cordero
Yoon Hwang
Andrew Ramirez
Cherylyn Smith
Ai Yukino

Background

What is Speech Acoustics?



What is Speech Acoustics?



Hypothesis

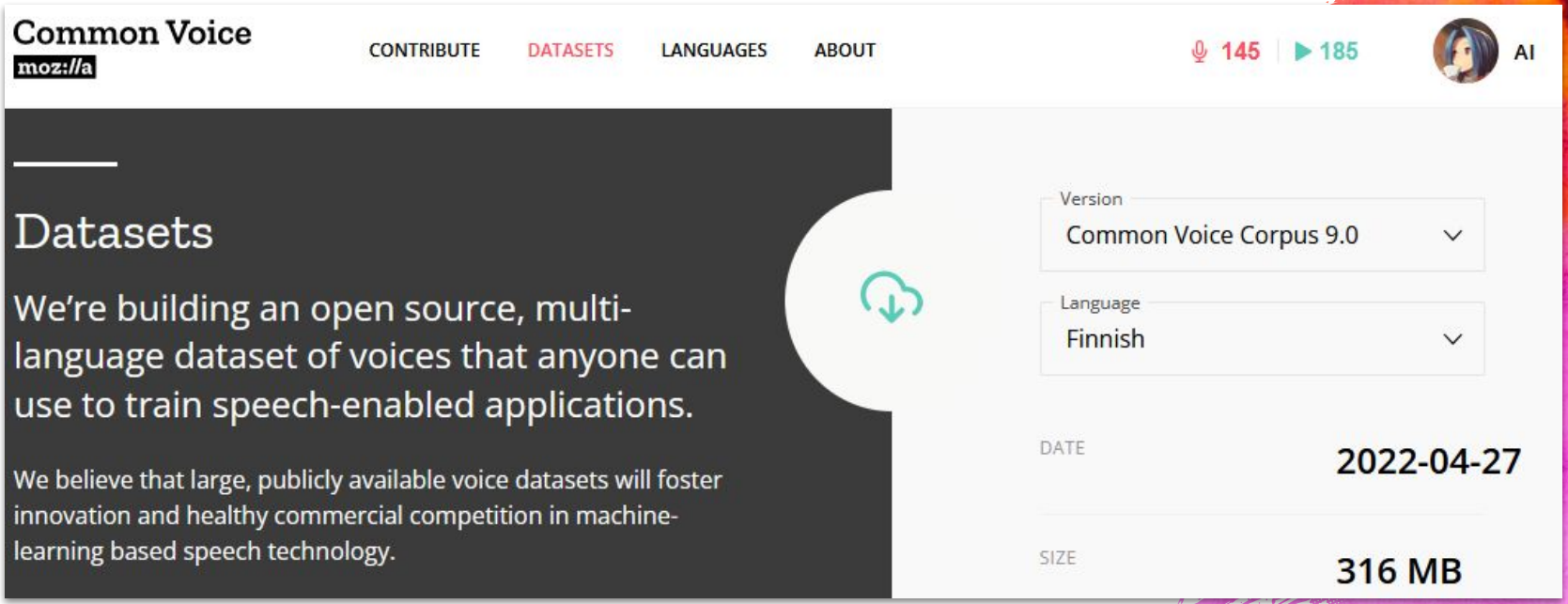
- Sounds where
  - F1 - F0, or
  - F1 / F0

Are larger will be more likely perceived as female than male

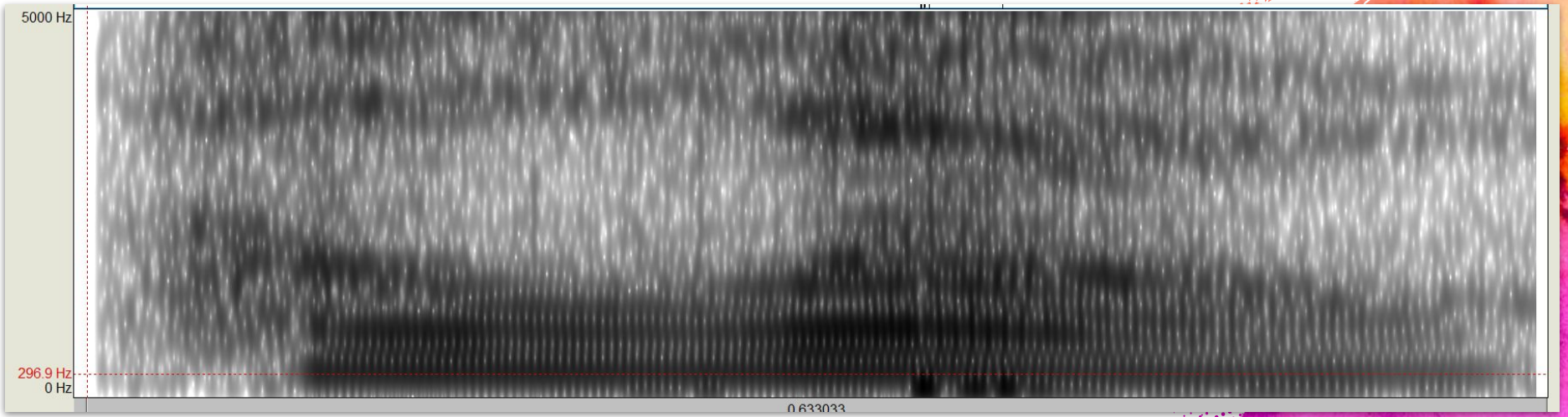
(This hypothesis was proven incorrect)


How exactly is our data extracted?

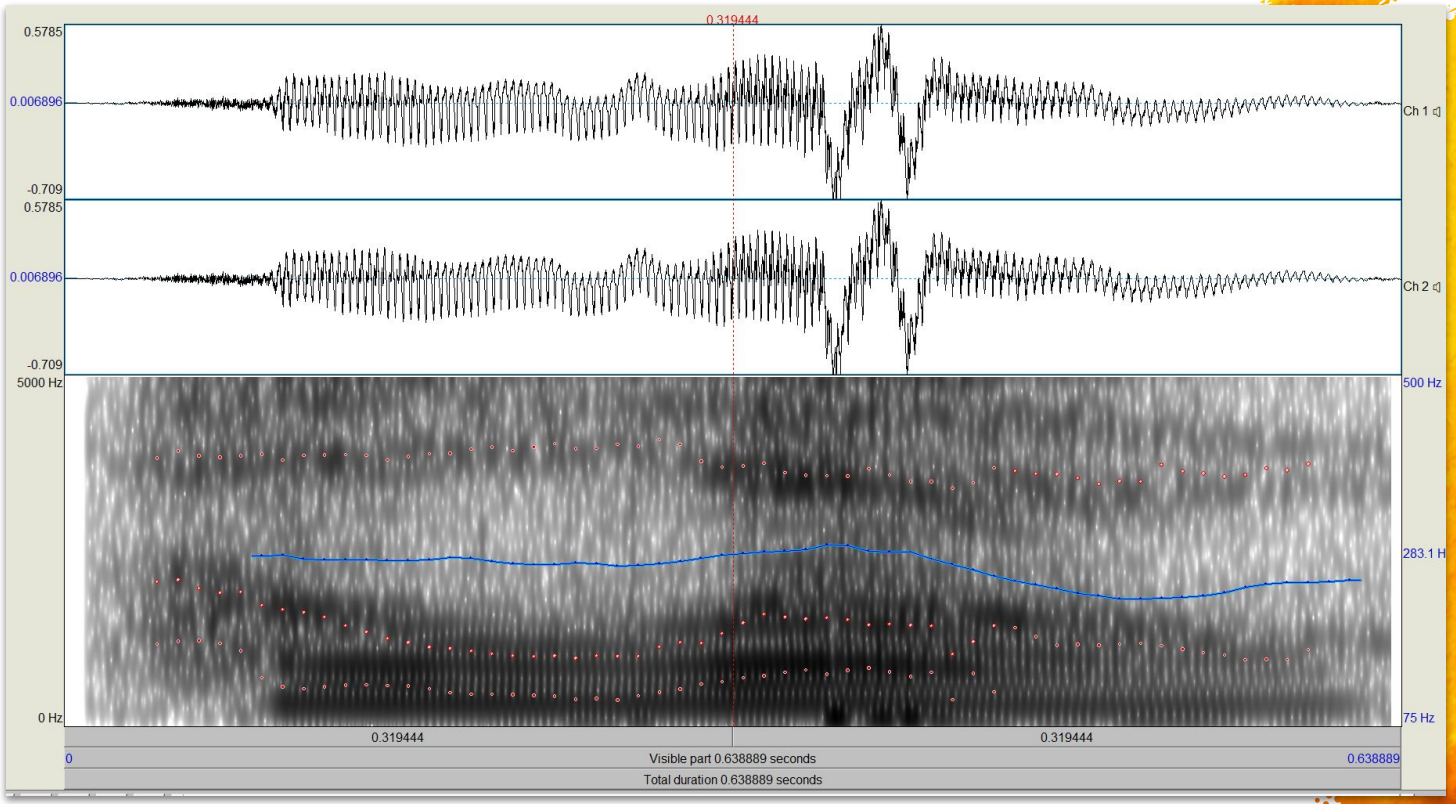
Where does our data come from?



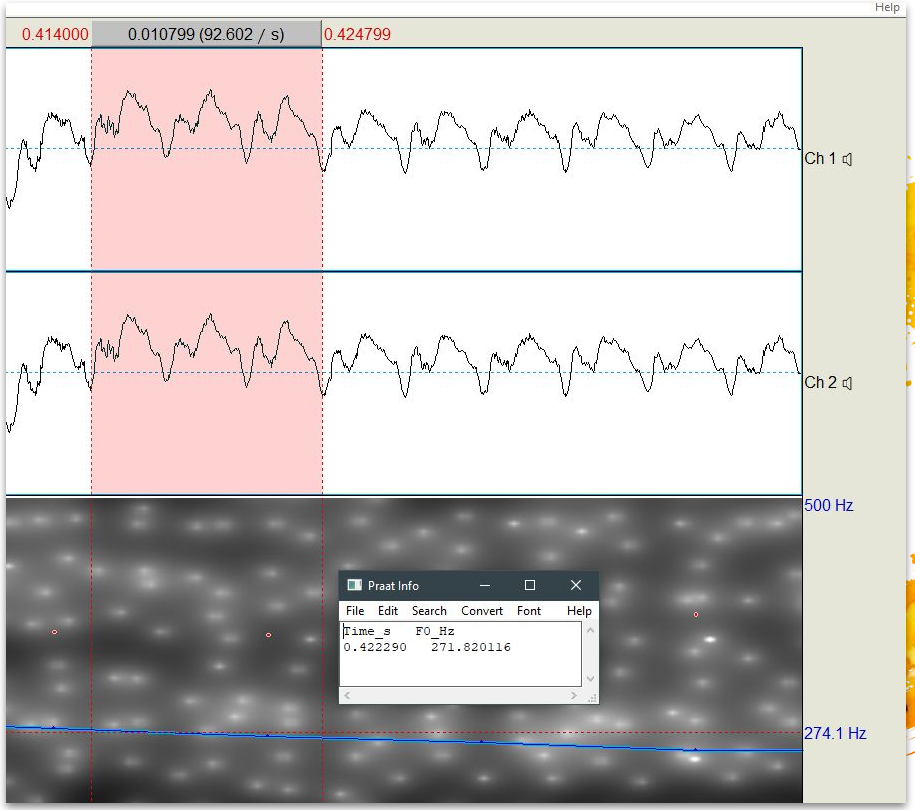
Where does our data come from?



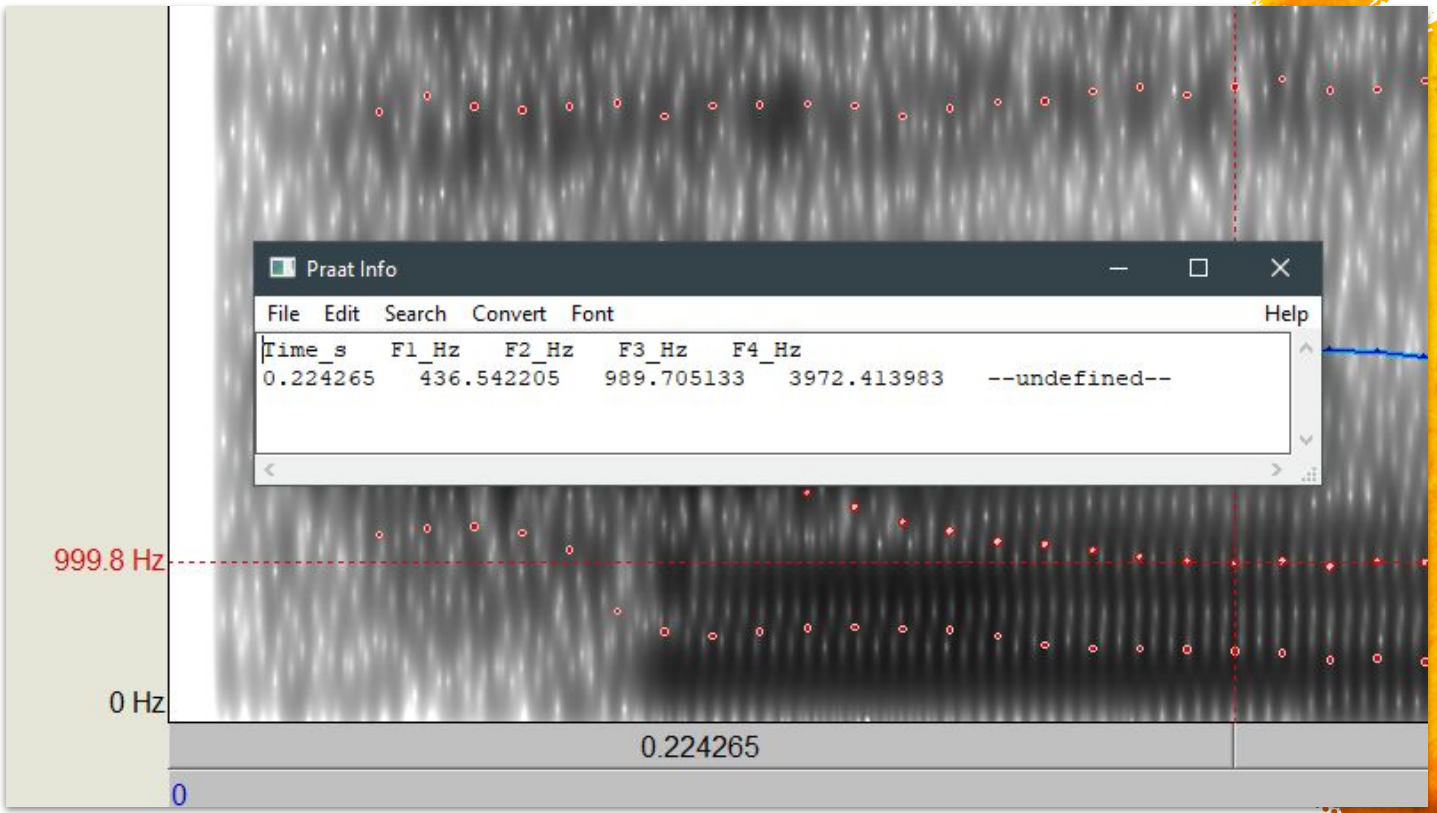
What are we analyzing?



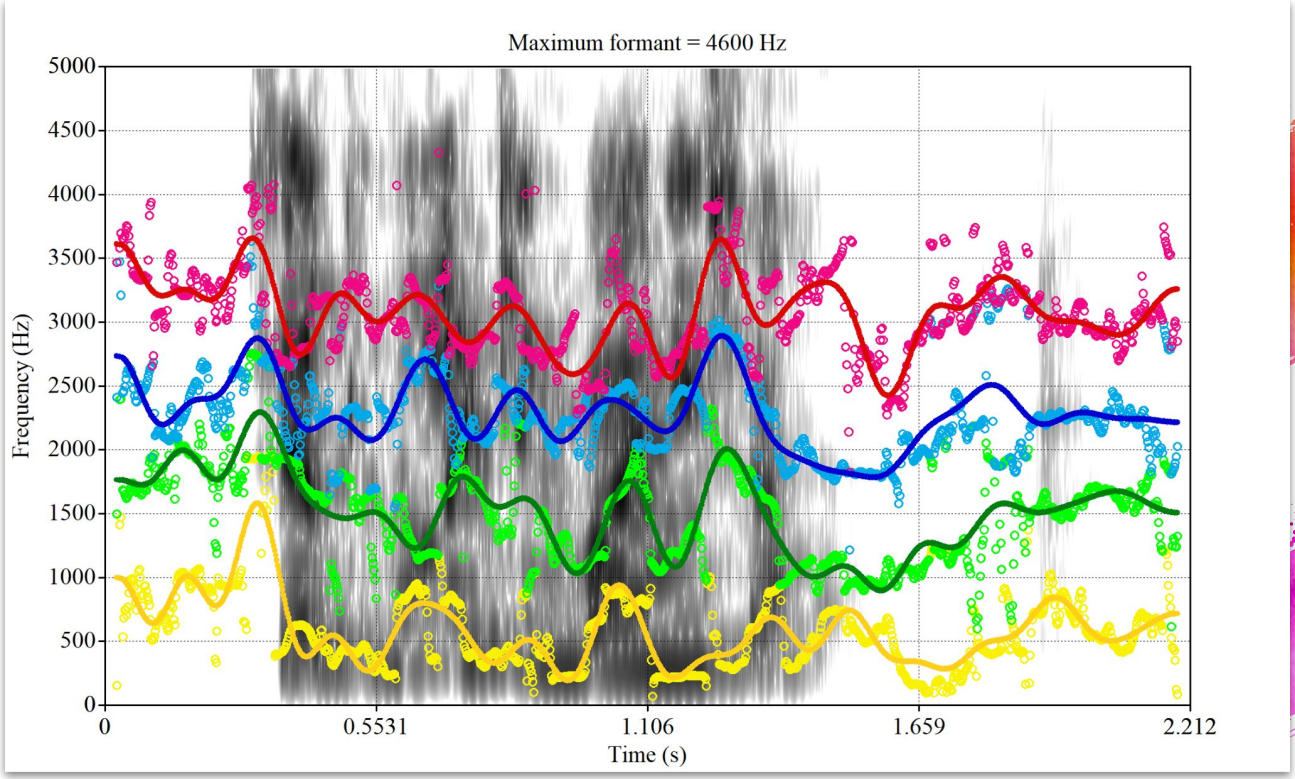
What are we analyzing?



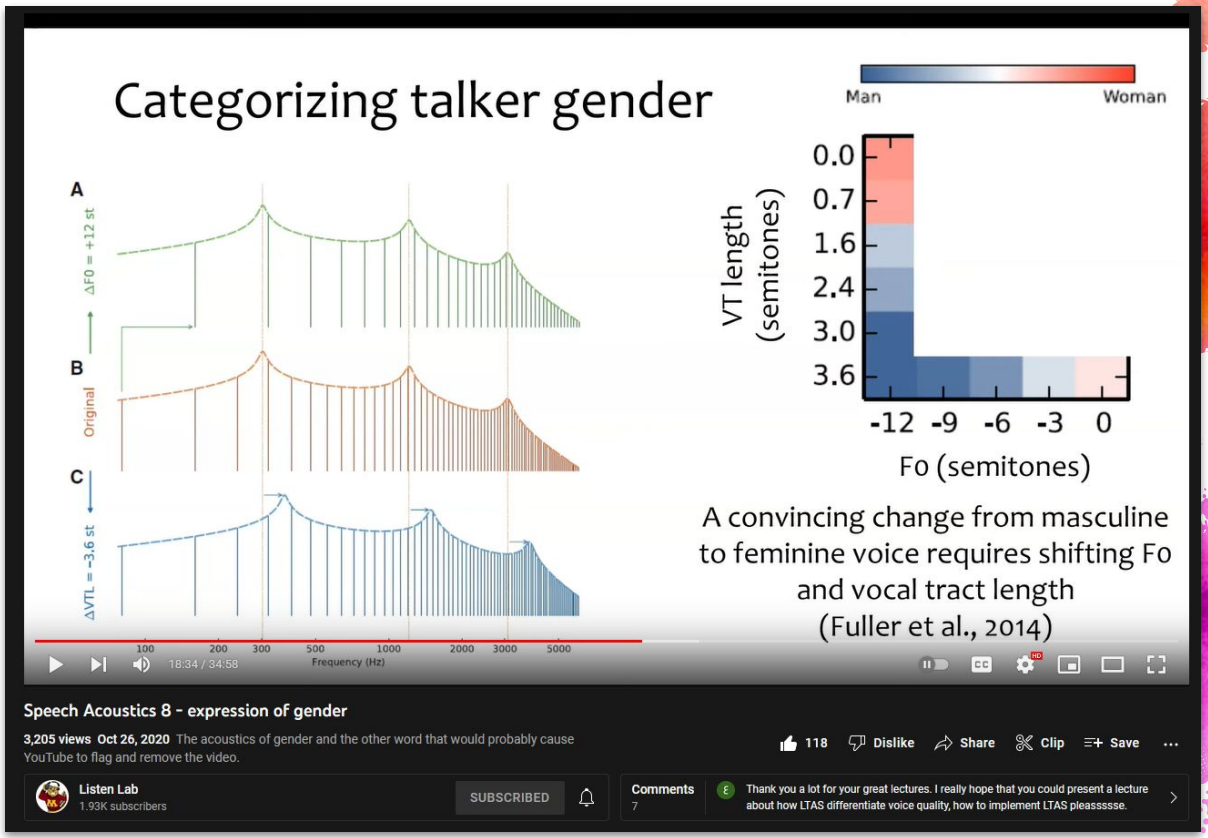
What are we analyzing?



Where does our data come from?

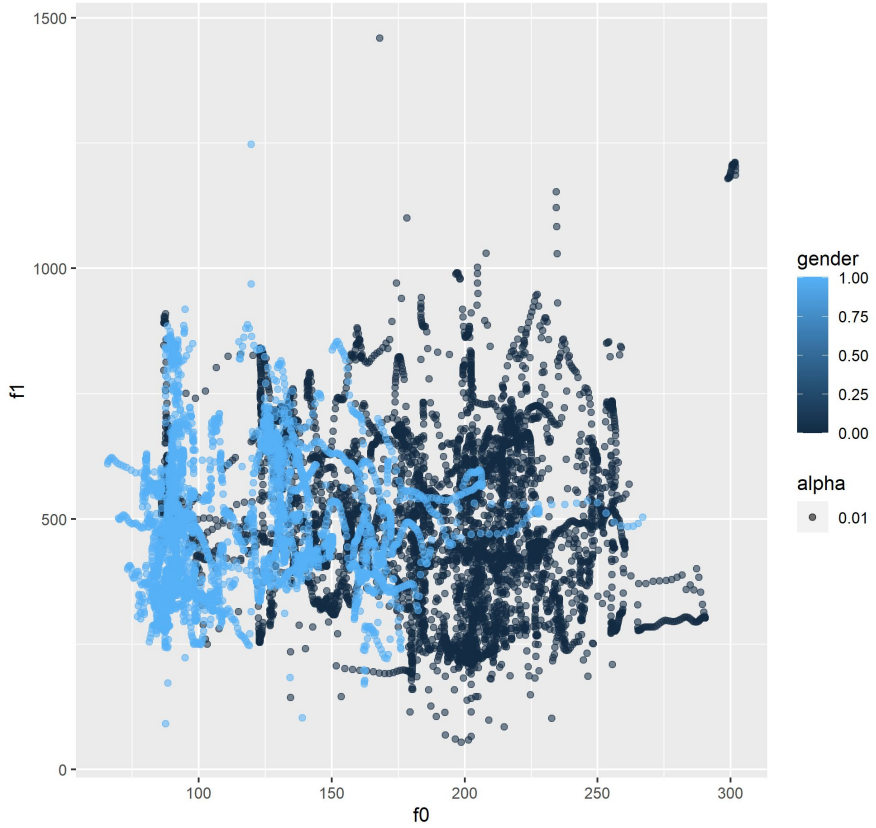


Excluding data:



Our Findings

A graph


1500
1000
500
0
f1
100
150
200
250
300
f0
gender
1.00
0.75
0.50
0.25
0.00
alpha
0.01

Some Linear Regressions

```
Call:
lm(formula = difference ~ gender, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-450.31 -113.88  -16.36  113.09  985.39

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  305.813      2.360  129.56   <2e-16 ***
gender        62.344      3.673   16.97   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 167.4 on 8562 degrees of freedom
Multiple R-squared:  0.03255,   Adjusted R-squared:  0.03244
F-statistic: 288.1 on 1 and 8562 DF,  p-value: < 2.2e-16
```

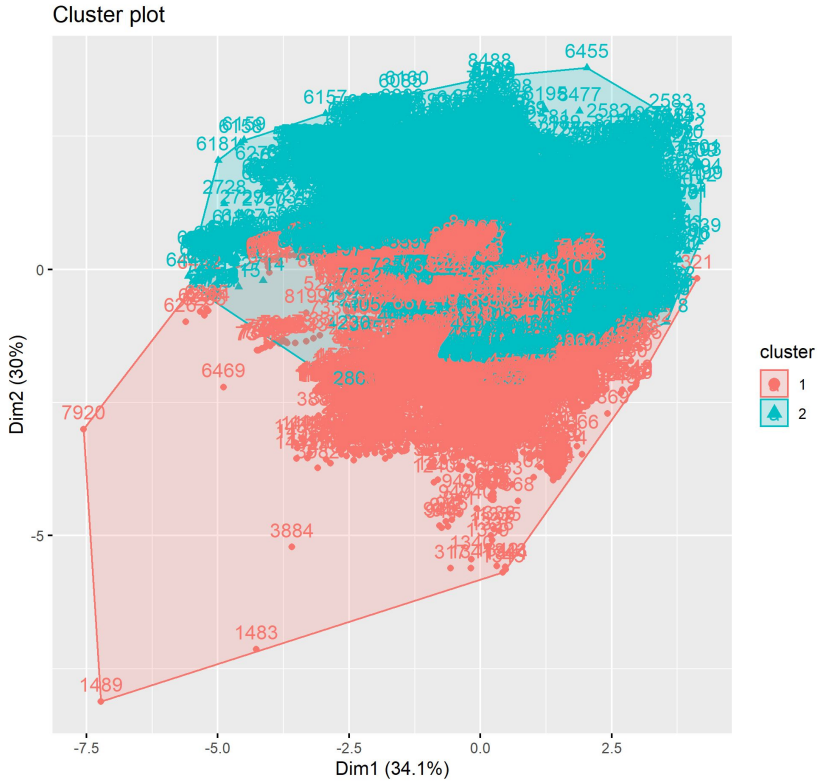
```
Call:
lm(formula = difference ~ ., data = subset(dataMod, select = -c(quotient,
    f0, f1)))

Residuals:
    Min      1Q  Median      3Q     Max
-425.60 -114.10  -27.02  116.59  750.07

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 36.558748  12.059524   3.032  0.00244 **
f2           0.047345   0.004416  10.722  < 2e-16 ***
f3          -0.051975   0.005376  -9.669  < 2e-16 ***
f4           0.104182   0.004649  22.411  < 2e-16 ***
age         -0.635777   0.302496  -2.102  0.03560 *
gender      85.659570   3.827000  22.383  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 159.4 on 8558 degrees of freedom
Multiple R-squared:  0.123,     Adjusted R-squared:  0.1225
F-statistic:   240 on 5 and 8558 DF,  p-value: < 2.2e-16
```

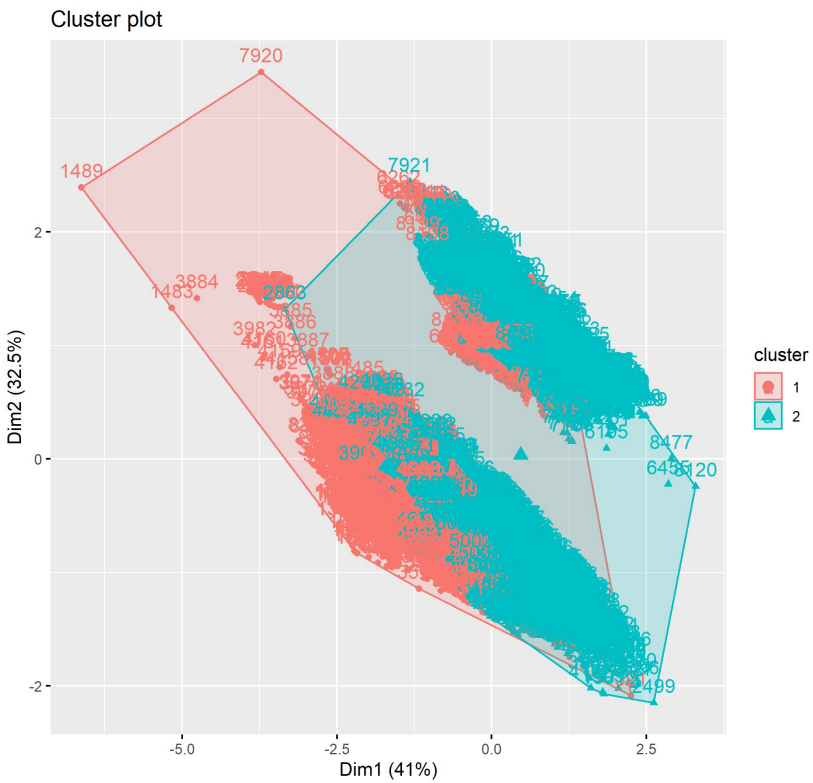
Kmeans clustering

```
K-means clustering with 2 clusters of sizes 3093, 5471

Cluster means:
        f1       f2       f3       f4       f0      age    gender difference quotient
1 538.7707 1665.471 2624.814 3633.948 162.7198 26.53088 0.2683479   376.0509 3.612509
2 464.7688 1105.363 1867.705 2817.508 158.3700 22.55346 0.4946079   306.3988 3.383596
```

```
K-means clustering with 2 clusters of sizes 2854, 5710

Cluster means:
        f1       f2    gender
1 519.7330 1847.494 0.3384723
2 477.3818 1037.827 0.4500876
```

Kmeans graphs



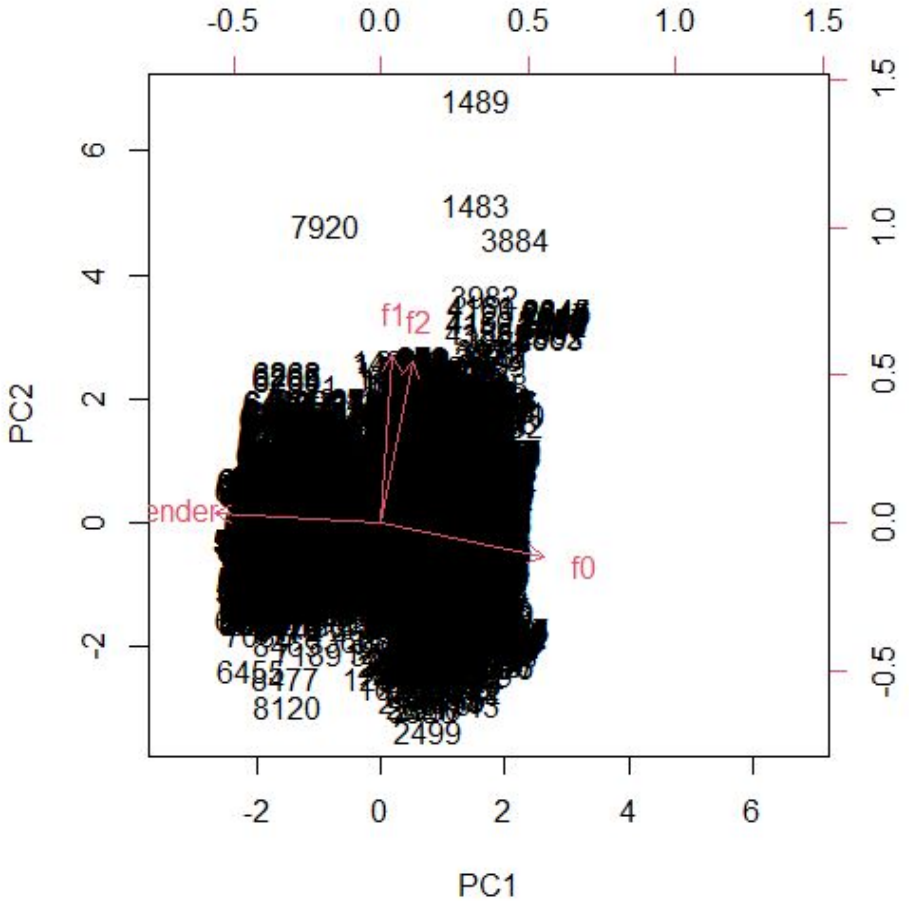


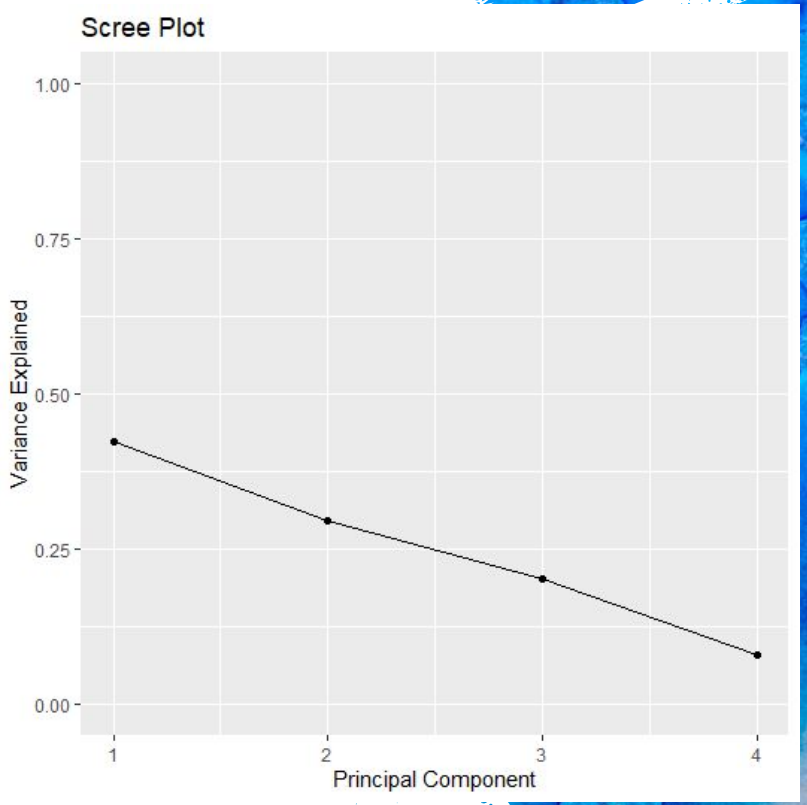
Unsupervised Learning: PCA

- PCA (Principal Component Analysis) is an unsupervised machine learning technique that seeks to find principal components – linear combinations of the original predictors – that explain a large portion of the variation in a dataset
  - explain most of the variability in a dataset with fewer variables than the original dataset

- The relationship between f0, f1, f2, and gender (variance explained by each principal component)
  - PC1: 0.42310653
  - PC2: 0.29636698
  - PC3: 0.20199740
  - PC4: 0.07852909

Unsupervised Learning: PCA Graphs





Supervised Learning: Classification

- Algorithms use labeled data.
- Using classification -> continuous variables

Question:

Can the model predict speaker's sex based on formants?

FORMANTS: harmonics forming picks.

F0 = fundamental frequency (pitch) - laryngeal source.

F1, *F2*, *F3* = filtered frequencies - vocal tract source.

Sound frequency correlates with person's age and sex.

F0: males ~ 130Hz females ~220Hz

Formant frequencies of adult males are 15% > women

Source: Smith, 2010 & Listen Lab, 2020.

Building the Classification Model

```r
# Import library
library(tidyverse)


# Load data
data <- read.csv("/Users/mariacordero/Downloads/formants.csv")

# Explore data
head(data)

# Selecting specific columns
data <- data[, c('f0', 'f1', 'f2', 'gender')]
head(data)
```

	f0	f1	f2	gender
1	232.2	660.5	2377.4	0
2	231.7	504.4	2379.4	0
3	231.2	405.0	2391.1	0
4	230.7	388.5	2439.6	0
5	230.2	438.5	2496.9	0
6	229.6	477.8	2538.8	0

Source: edureka!

Model and Prediction

```
# Splitting data

library(caTools)

sample.split(data$gender,SplitRatio = 0.65)->split_values
subset(data,split_values==T)->train_set
subset(data,split_values==F)->test_set

# Loading RPART to build classification model
library(rpart)

rpart(gender ~ ., data = train_set)->mod_class
predict(mod_class, test_set, type = "matrix" )->result_class
table(test_set$gender, result_class)
```

Source: edureka!

Prediction - Results

```
 result_class
  0.032258064516129 0.0387323943661972 0.0504166666666667 0.0679611650485437
0                13                  116                1247                 49
1                 1                    7                  55                  5
 result_class
  0.301020408163265 0.350282485875706 0.779661016949153 0.818627450980392
0                79                  48                81                  29
1                35                  36               287                  91
 result_class
  0.878048780487805 0.90154711673699
0                 4                94
1                17               704
```

		Predicted	
	Confusion Table	0	1
Actual	0	4	94
	1	17	704

Model Accuracy:

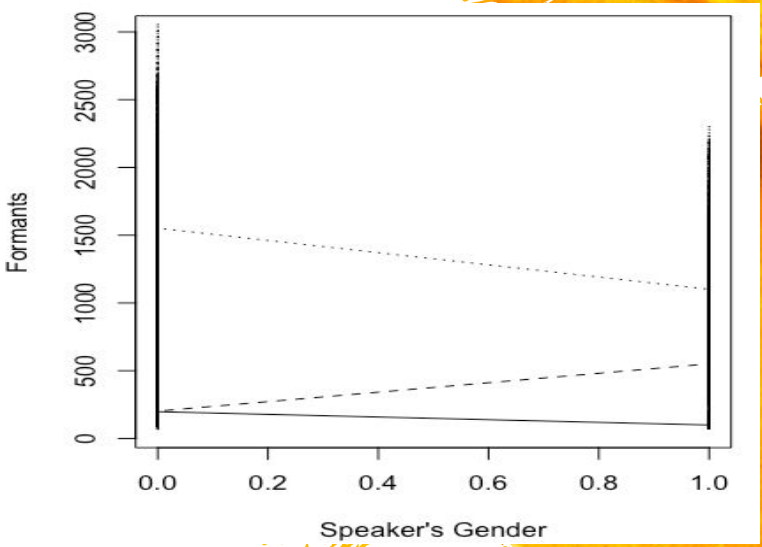
(4+704)/(4+94+17+704)

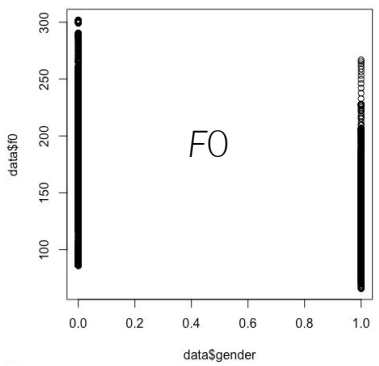
86%

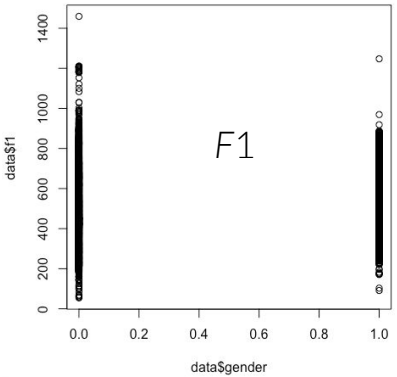
Source: edureka!

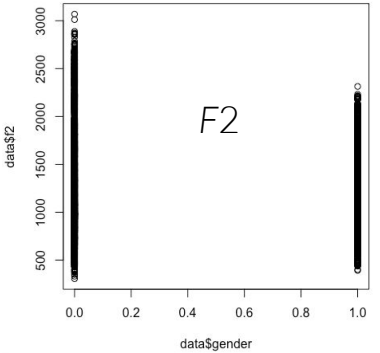
Plotting variables

```
library("tidyverse")
library(ggplot2)

data <- read.csv("/Users/mariacordero/Downloads/formants.csv")

plot(data$gender, data$f0,
     main = "Formants by Speaker's Gender",
     xlab = "Speaker's Gender",
     ylab = "Formants",
     type = "l",
     ylim = c(50, 3000))
lines(data$gender, data$f1,
      lty = "dashed")
lines(data$gender, data$f2,
      lty = "dotted")
```





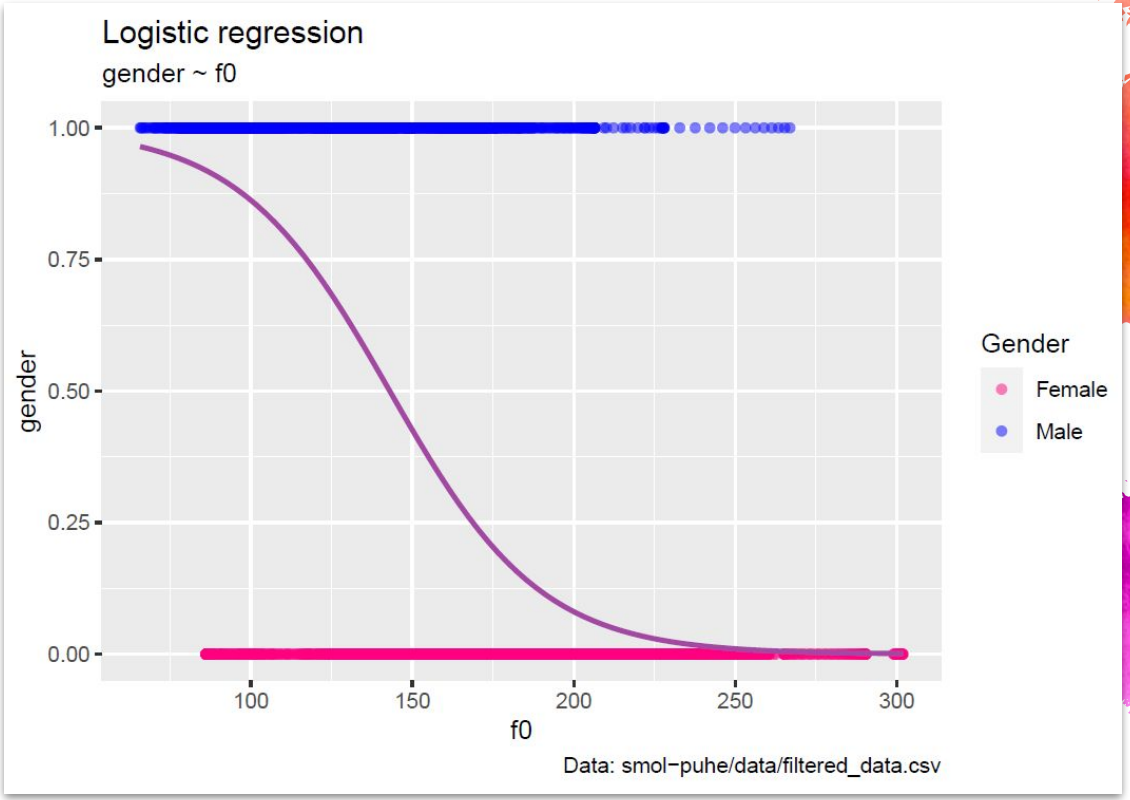




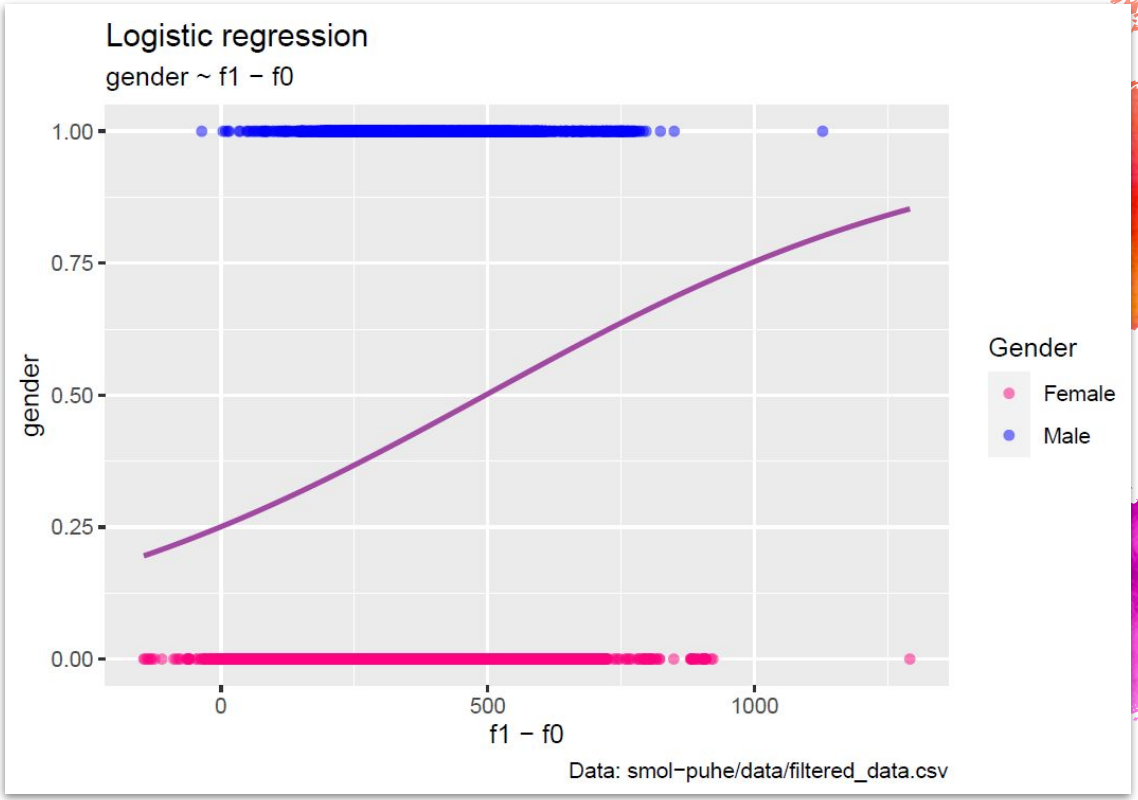
Source: O'Loughin, 2021

General Notes

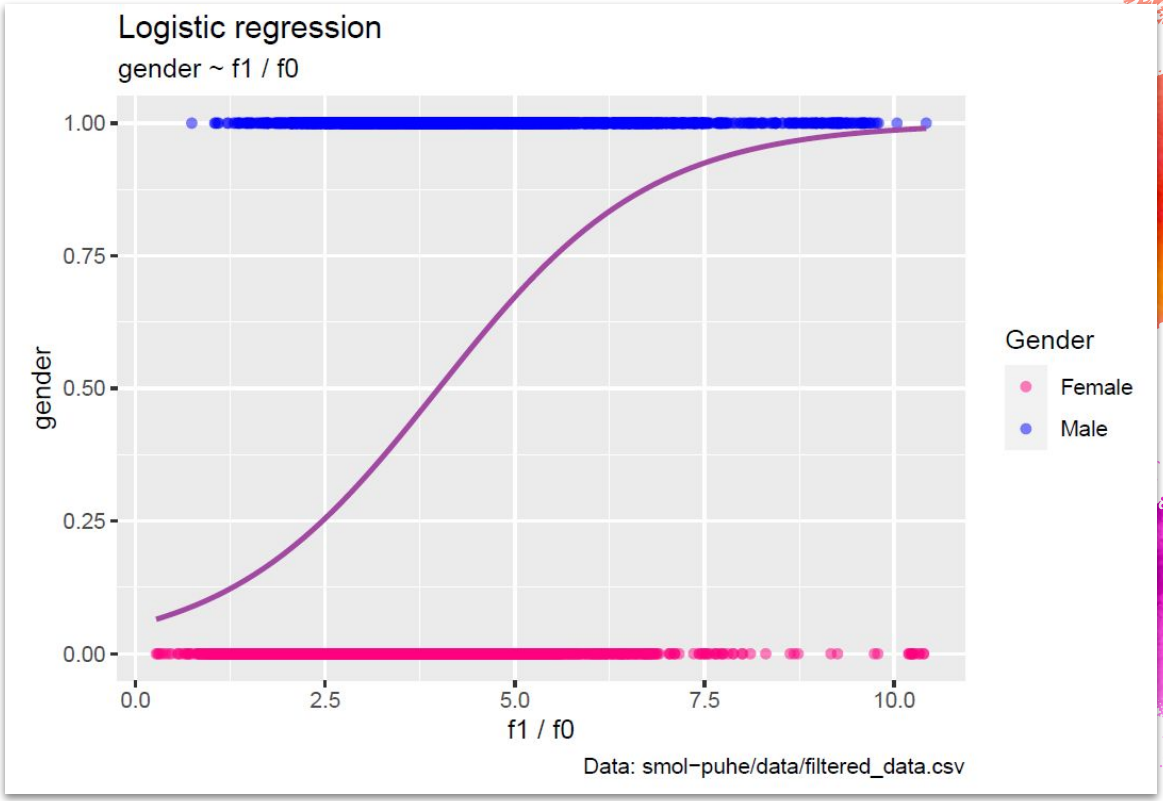
Logistic regression



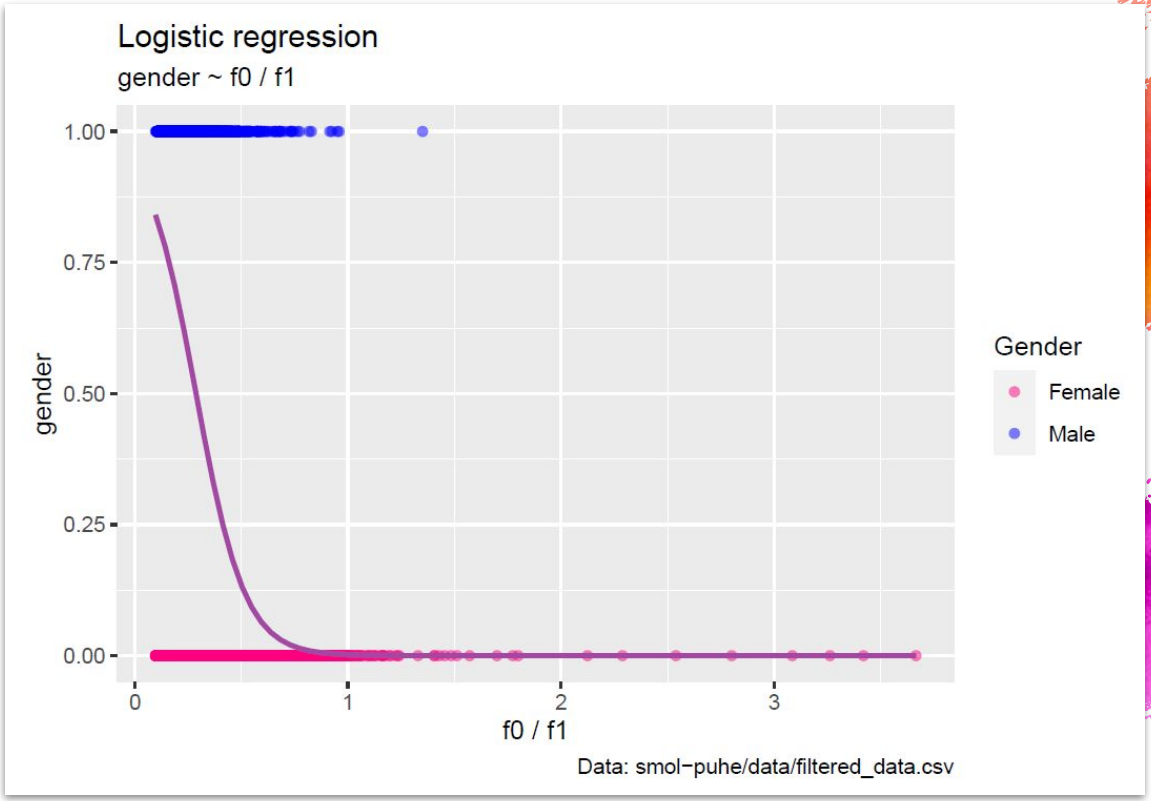
Logistic regression



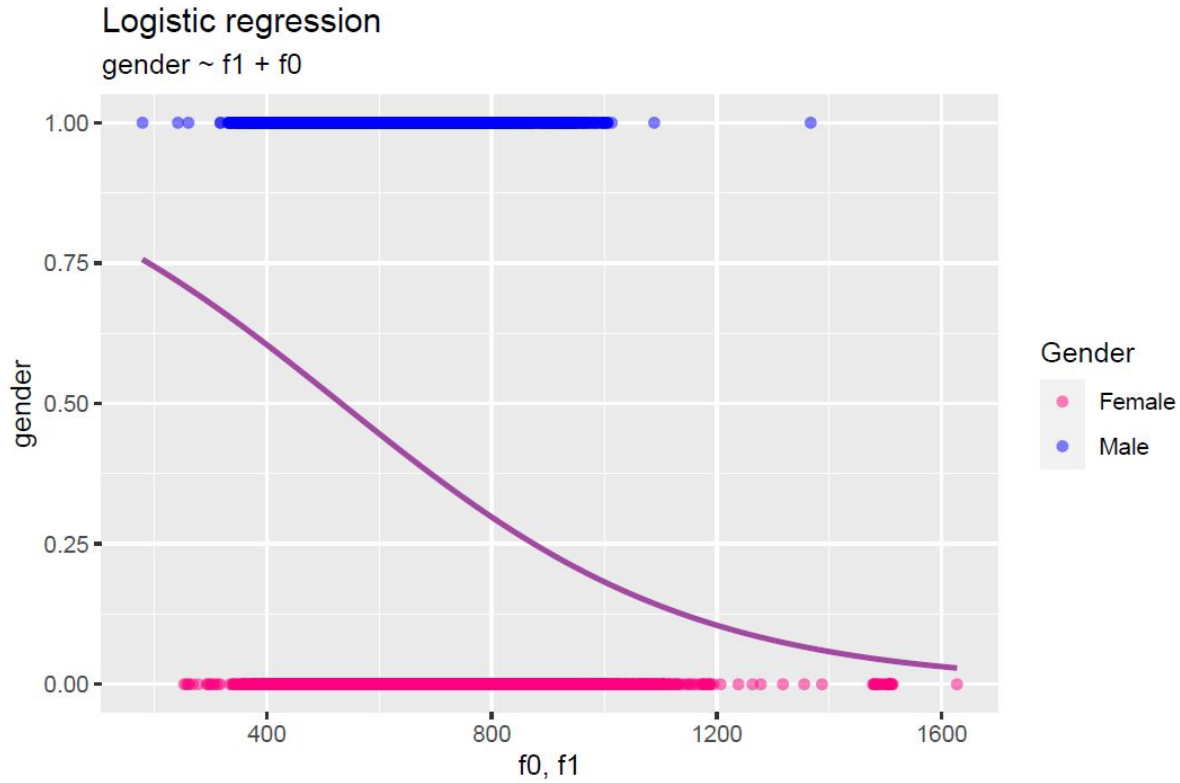
Logistic regression



Logistic regression



Logistic regression



Supervised Learning with Caret

https://topepo.github.io/caret/

“The `caret` package (short for **C**lassification **A**nd **RE**gression **T**raining) is a set of functions that attempt to streamline the process for creating predictive models. The package contains tools for:

- data splitting
- pre-processing
- feature selection
- model tuning using resampling
- variable importance estimation”

The caret package was used to examine multiple processes for isolating the “best” predictive model for a binomial outcome. The output vector (gender) was cast as a factor and used as a category.

Skimr Summary Statistics

"`skimr` is designed to provide summary statistics about variables in data frames, tibbles, data tables and vectors. "

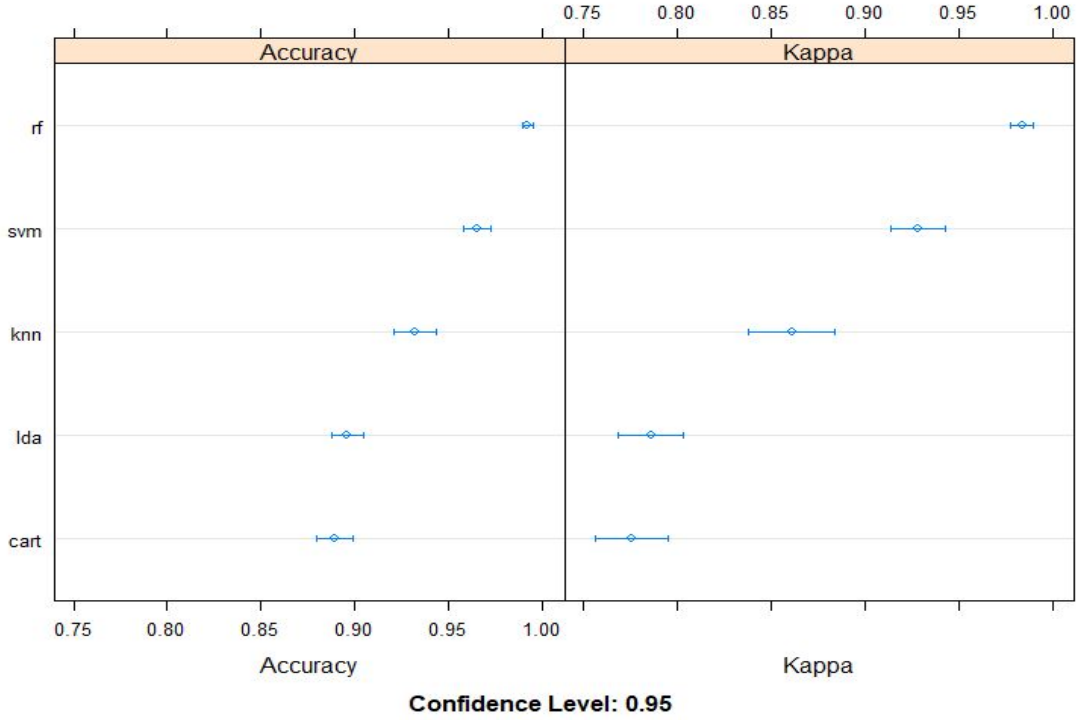
```
-- Data Summary ------------------------
                           Values
Name                       Piped data
Number of rows             5482
Number of columns          9
_____________________
Column type frequency:
  numeric                  8
_____________________
Group variables            gender

-- Variable type: numeric ---------------------------------------------------------------------------------------------------------------
   skim_variable gender n_missing complete_rate  mean    sd    p0   p25   p50   p75  p100 hist
 1 f1            0      0         1             496.  178.   54.2 368.  480.  620.  1211. ▂▇▃▂▁
 2 f1            1      0         1             487.  137.  103.  384.  464.  578.  1248. ▂▇▃▁▁
 3 f2            0      0         1            1339.  530.  332   946.  1237. 1674. 3012. ▂▇▅▂▁
 4 f2            1      0         1            1250.  330.  392.  1037. 1188. 1480. 2314. ▂▇▃▁▁
 5 f3            0      0         1            2221.  535.  913.  1765. 2266  2690. 4248. ▂▇▅▁▁
 6 f3            1      0         1            2010.  376.  940.  1738. 2059. 2298. 3350. ▂▇▃▁▁
 7 f4            0      0         1            3232.  584.  1690. 2796. 3062. 3670. 5313. ▂▇▃▂▁
 8 f4            1      0         1            2928.  398.  1678. 2655. 2985. 3203. 4223. ▂▇▃▂▁
 9 f0            0      0         1             189.   42.1  86.1  159.  198.  217.  302. ▁▂▅▇▂
10 f0            1      0         1             118.   32.9  65.7  90.3  107.  138.  267. ▇▆▂▁▁
11 intensity     0      0         1              72.9   7.01 52.6  68.2  73.4  78.6  85.9 ▁▂▃▇▆
12 intensity     1      0         1              72.6   7.34 52.8  66.8  73.5  79    83.8 ▁▂▃▇▅
13 harmonicity   0      0         1             -19.6  73.6 -226.   3.95 10.8  16.1  34.9 ▁▁▁▁▇
14 harmonicity   1      0         1             -18.5  70.1 -226.   3.6   8.6  12.9  50.3 ▁▁▁▁▇
15 age           0      0         1              26.0   7.34 20     20    20    30    40  ▇▁▃▁▂
16 age           1      0         1              21.3   3.37 20     20    20    20    30  ▇▁▁▁▁
```

Variables showing distinctly different histograms when isolated by gender (f1,f4, f0)

https://cran.r-project.org/web/packages/skimr/vignettes/skimr.html

Model and Prediction

Back to caret:



Examining caret prediction metrics

Models:	lda,	cart,	knn,	svm,	rf				
Number of resamples	10								
Accuracy									
		Min.	1st Qu.	Median	Mean	3rd Qu	Max.		NA's
lda		0.8792711	0.8838269	0.8995434	0.8958124	0.9035947	0.9132420		0
cart		0.8701595	0.8786353	0.8926941	0.8894239	0.9019942	0.9020501		0
knn		0.9111617	0.9178511	0.9327264	0.9320581	0.946347	0.9544419		0
svm		0.9497717	0.9572893	0.9681093	0.9651137	0.9743151	0.9749431		0
rf		0.9840547	0.98918	0.9931507	0.9920227	0.9948669	0.9977169		0
Kappa									
		Min.	1st Qu.	Median	Mean	3rd Qu	Max.		NA's
lda		0.7498791	0.7616346	0.7929919	0.7855969	0.8016922	0.8216689		0
cart		0.7365251	0.7540359	0.7820249	0.7755645	0.800893	0.8018641		0
knn		0.8165404	0.8330466	0.8629379	0.8608631	0.8894888	0.9064445		0
svm		0.8972445	0.9122047	0.9345624	0.9283963	0.9472193	0.9485003		0
rf		0.9672294	0.9776952	0.9858989	0.9835695	0.9894382	0.9953034		0

Predictions with caret

Random Forest

4386 samples
8 predictor
2 classes: '0', '1'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 3947, 3947, 3947, 3947, 3947, 3948, ...
Resampling results across tuning parameters:

mtry	Accuracy	Kappa
2	0.9920227	0.9835695
5	0.9881450	0.9755913
8	0.9870045	0.9732481

Confusion Matrix and Statistics

Prediction	Reference 0	1
0	639	12
1	2	443

Accuracy : 0.9872
95% CI : (0.9787, 0.993)
No Information Rate : 0.5849
P-Value [Acc > NIR] : < 2e-16

Kappa : 0.9736

Mcnemar's Test P-Value : 0.01616

Sensitivity :	0.9969
Specificity :	0.9736
Pos Pred Value :	0.9816
Neg Pred Value :	0.9955
Prevalence :	0.5849
Detection Rate :	0.5830
Detection Prevalence :	0.5940
Balanced Accuracy :	0.9853

'Positive' Class : 0

References

Edureka! (2018). Machine Learning with R | Machine Learning Algorithms | Data Science Training | Edureka. Retrieved from https://www.youtube.com/watch?v=SeyghJ5cdm4&t=505s

Listen Lab. (2020). Speech acoustics 5 - Vowel formants. Retrieved from https://www.youtube.com/watch?v=glnUFa2fLyE&t=2s

O"Loughin, E. (2021). How to plot multiple datasets on the same chart in R #38. Retrieved from https://www.youtube.com/watch?v=aVQXBUCjWas&t=455s

Smith, D. (2010). Does knowing speaker sex facilitate vowel recognition at short durations? Acta Psychol (Amst). 2014 May;148:81-90. doi: 10.1016/j.actpsy.2014.01.010.

Brownlee, Jason. "Your First Machine Learning Project in R Step by Step" https://machinelearningmastery.com/machine-learning-in-r-step-by-step/

Kuhn, The caret Package, https://topepo.github.io/caret/
Using skimr. 2022, https://cran.r-project.org/web/packages/skimr/vignettes/

Prabhakaran, Selva. 2018. "Caret Package – A Practical Guide to Machine Learning in R",https://www.machinelearningplus.com/machine-learning/caret-package/

More references

Fuller, Christina D., et al. "Gender categorization is abnormal in cochlear implant users." Journal of the Association for Research in Otolaryngology 15.6 (2014): 1037-1048.

- Article link: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4389960/
- Video clip summary of relevant results for gender perception: https://www.youtube.com/watch?v=TWRB443YrHI&t=1103s
  - 18:23 -> 21:21 (3 minutes)

Conclusion

Considering only F0 and F1 was insufficient to predict gender – let alone cluster.

Based on analyses such as Yoon's PCA and the various ML models from Cherylyn's and Carat code, considering at least F2 and the higher formants like F3 and F4 could lead to better clustering and predictions.

Conclusion

In addition, Andrew's linear models showed that all of F1, F2, F3, F4, and F0 were statistically significant. There were not clear clusters from his k-means code, so this might suggest that we need to consider variables beyond the formant and fundamental frequency.