Hossain, Tomal
COMPSCI-590V
Spring 2017

HW4: Final Project Design Document

Selected Dataset

Voice Gender Recognition through the application of deep learning and a variety of classification algorithms. My web page will walk a user through classifying at least 2 audio file samples (ideally 1 female and 1 male). The user will have the ability to visualize all the pertinent acoustical properties of the uploaded sound file through various visualizations. The concluding section of the web page story will display key metrics relating to the reasoning for the given classification as well as a simple "playlist" that allowing for the vocal samples to be played conveniently.

http://www.primaryobjects.com/2016/06/22/identifying-the-gender-of-a-voice-using-machine-learning/

Questions/Hypotheses

1.
    a. **Question**: Which variables regarding acoustic properties are most responsible for complicating the accuracy of classification of vocal samples as either male or female from the baseline of simply classifying by some threshold frequency related to the mean fundamental frequency of each vocal sample?
    b. **Hypothesis:** It seems to me that metrics relating to the pitch variability of vocal samples may have a considerable influence on the classification of a vocal sample as either male or female.
2.
    a. **Question:** How do the distributions of acoustic properties related to inflection and tonal quality differ between male and female vocal samples? What does such variation, if at all present, say about gender norms and roles in society regarding language and communication?
    b. **Hypothesis:** It seems to me that speech patterns that show up in the vocal samples are more likely the result of cultural exposure and upbringing rather than biological differences. This observation would be in contrast to mean fundamental frequency, which is certainly much more biologically rooted in the fact that men's voices on average are roughly an octave deeper than female voices.
3.
    a. **Question:** How does the accuracy of the most accurate model utilized in the study that utilized this dataset compare with the accuracy of humans in classifying vocal samples as either male or female? What does the divide

between these two accuracies, if present, say about the ability of humans to discern nuanced acoustic properties from solo voice recordings?

b. **Hypothesis:** In the article that utilized the dataset, the most sophisticated classification algorithm obtains an accuracy of nearly 100%. It seems to me that human beings would average closer to the 80-90% range in this regard.

Processing/Analytics

1. If possible to gain access the original wave files of the 3000+ vocal samples used in the original research project, it would be ideal to perform spectrogram related calculations to reveal information about formants and timbre which may have some bearing on male/female classification.
2. It may be helpful to convert the kilohertz values in the original dataset to approximate 12 tone equal temperament note values as used in Western standard notation. Thus, I may create 20 new columns that correspond with the existing 20 output features except relate note values.
3. It may be helpful to perform some subsetting or dimensional reduction of the original dataset given that some data features seem to be relatively extraneous to others such as entropy, flatness, and skew, and kurtosis.

Visualizations

1. Parallel Coordinates
2. Histograms
3. Heatmap
4. Scatterplot Matrix

Interactions

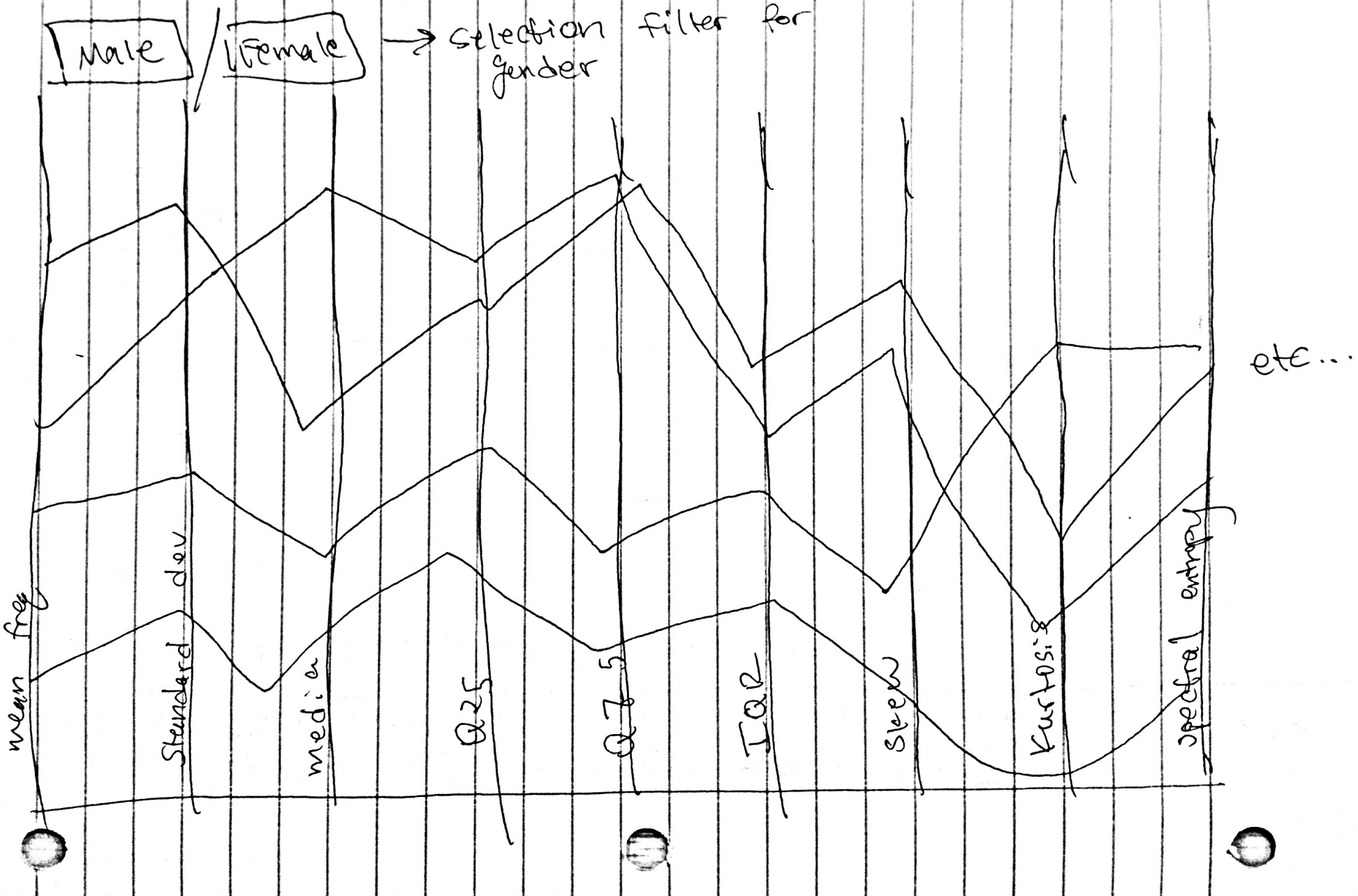
1. Buttons/tabs -> hiding/showing visualizations
2. Selection filters for axes
3. probing/selection for coordinated visualizations
4. Audio components -> ideally include ability to play the vocal sample for a given selected track.

Voice Gender Recognition ⇒ a journey into classifying several vocal samples as male and elders as female.

☆ Background Information. . . . . .

[Table view of dataset]

Question #1 : Complicating features : parallel Coordinates



Question ~~#1~~ #2 : Distributions

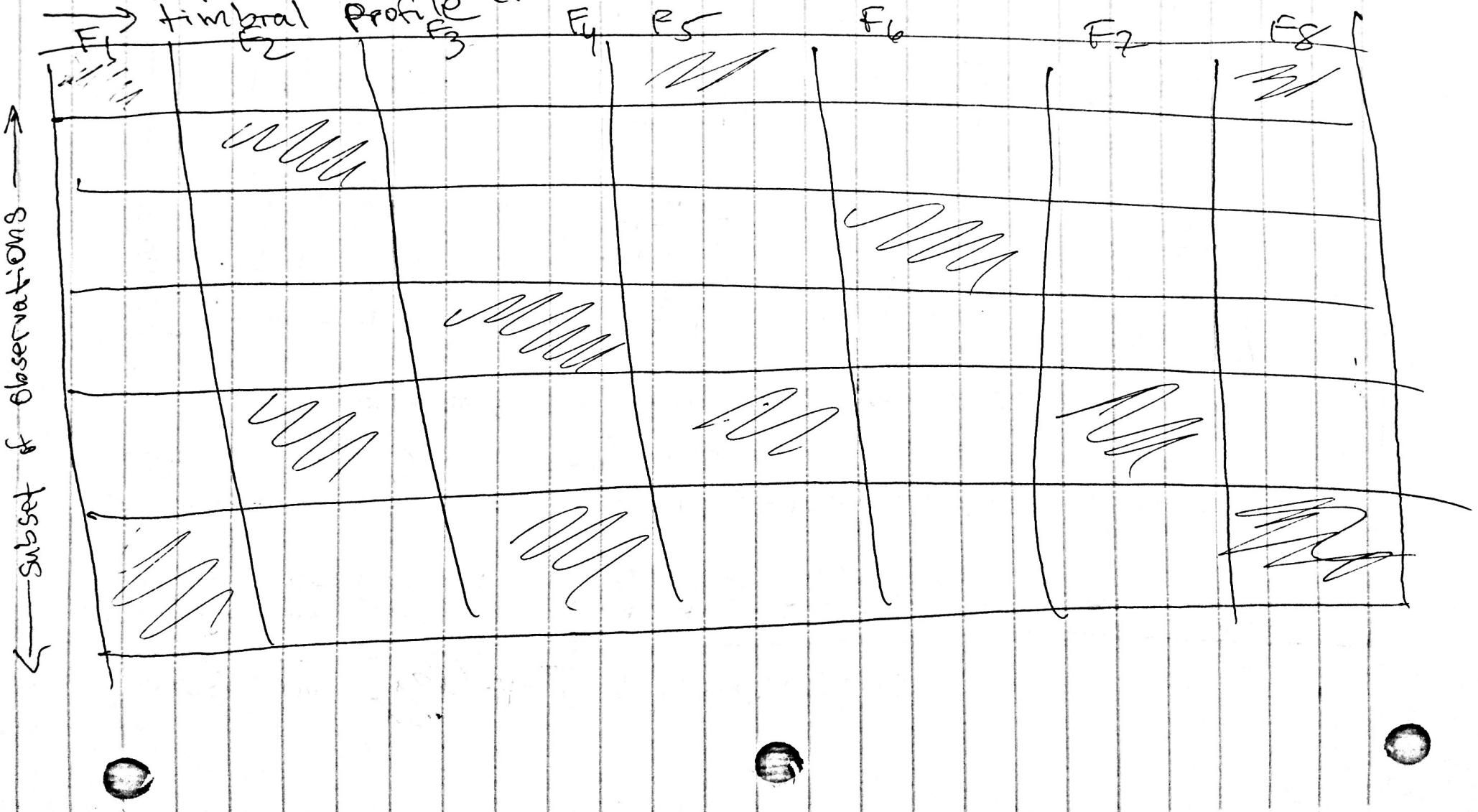
(b) ~~Question: why is it so difficult to visualize/determine which acoustic properties whether a vocal sample is male or female?~~

★ visualization of an observation(s) (ability to select 1 at a time) in histograms.

Distributions of acoustic properties of dataset

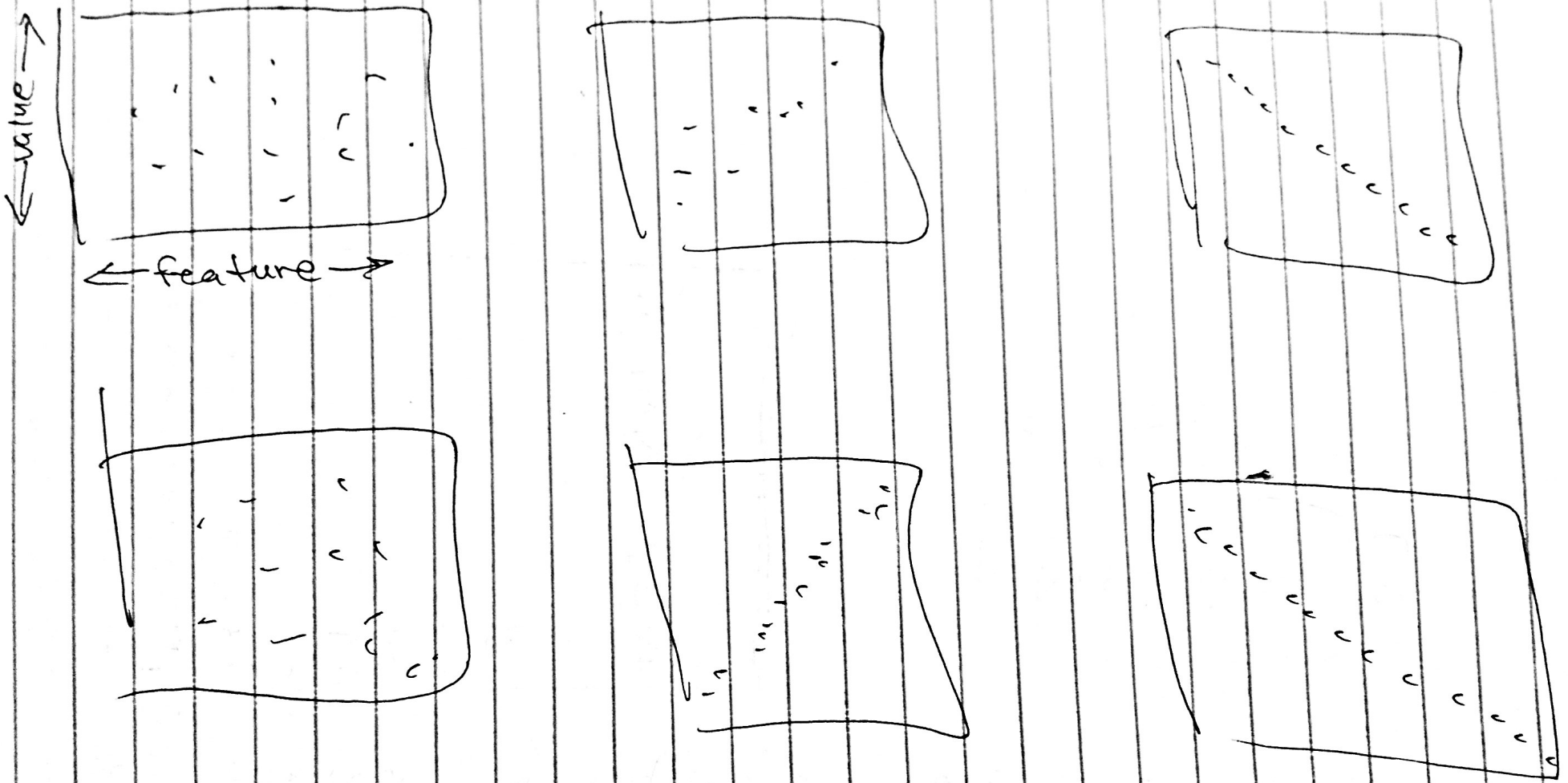
Question 3 : Human vs Computer

Heatmap of Vocal Timbre (subset of observations)



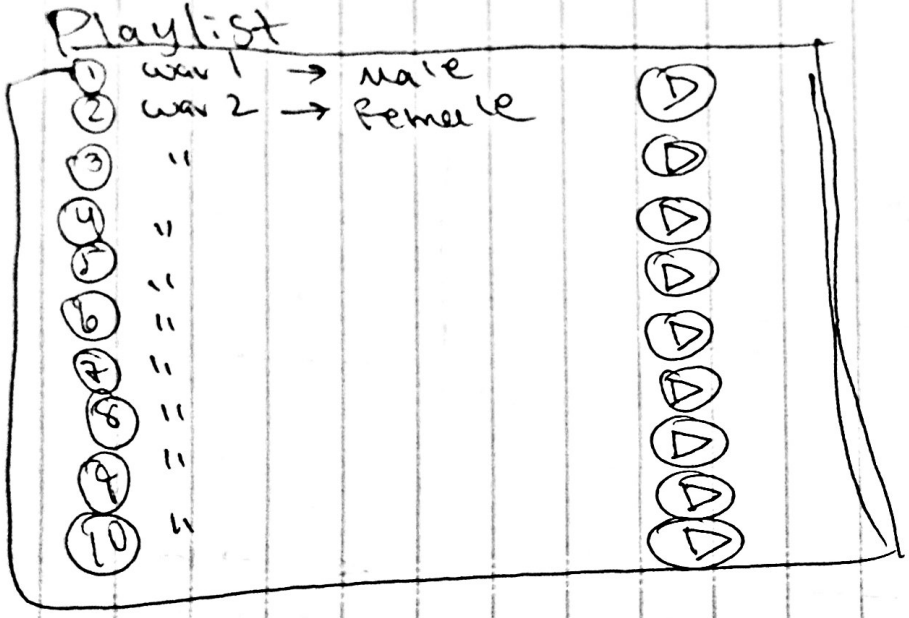
(potential)

Question 4: scatterplot Matrix for most important features



Conclusion

Classification of uploaded wav files



Summary of Statistics for your playlist