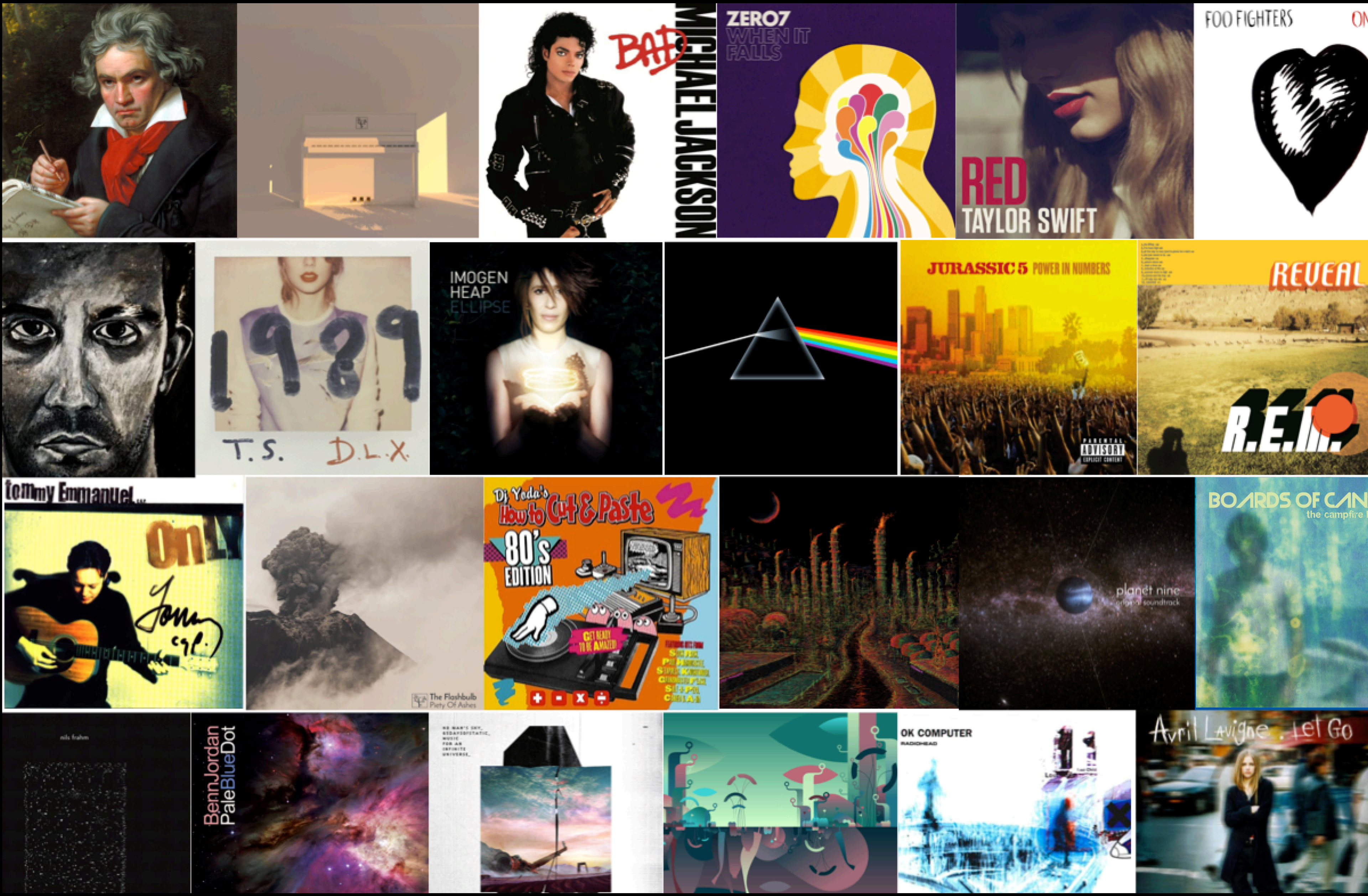
# Machine learning with music




Alex Clarke : informationcake.com

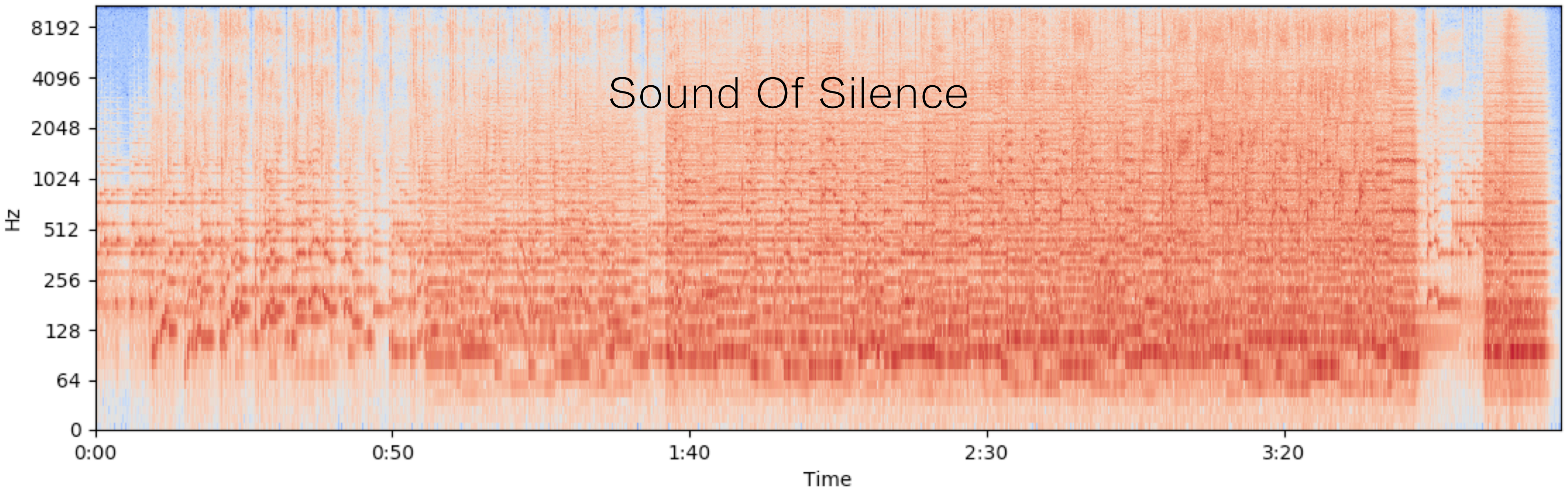
# Machine learning with music

- How good are you at identifying sounds/music?
- Unsupervised clustering: linear & non-linear
- Supervised learning: better accuracy than humans?
- Selecting music for advertising themes (Summer student Naim Sen)
- Man/Woman voice recognition (MPhys students Chloe Hutton & Ana Zardoshti)

Scripts available on github:

https://github.com/informationcake/music-machine-learning

# Audio classification: Context? Biased?




Song originally written in 1964 | by Simon and Garfunkel

Spectrogram is of a cover from 2015 | by a heavy metal band called Disturbed

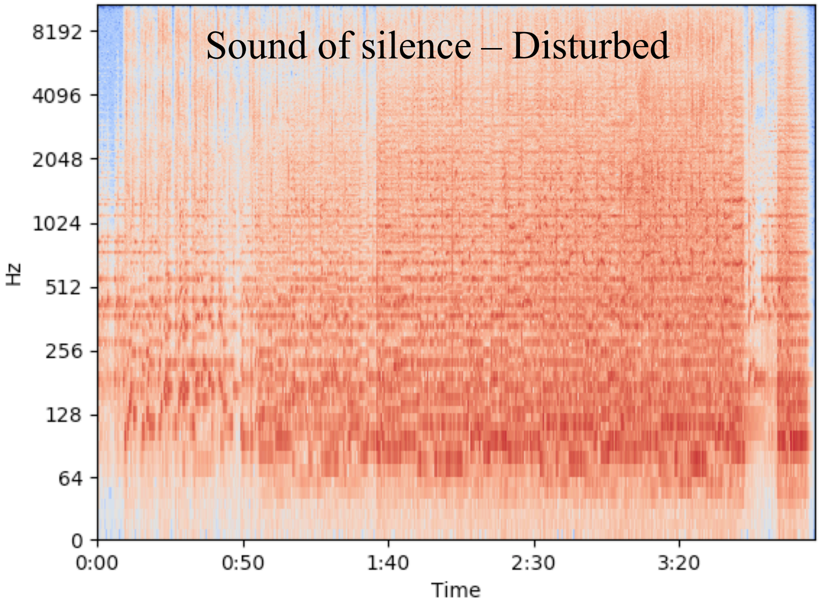
Do you/would you like this song? Why?

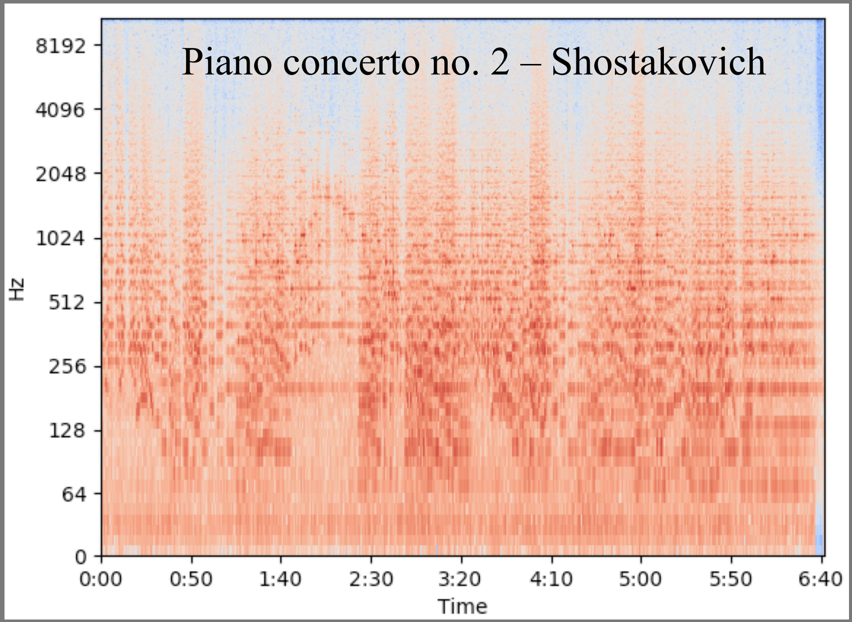
You may love/hate the original, you may love/hate the cover, for different reasons?

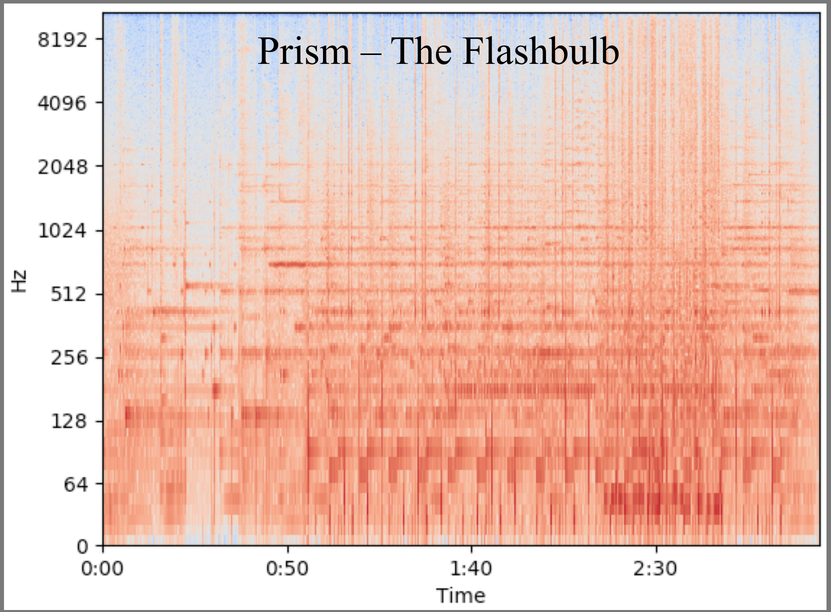
You may not enjoy heavy metal in general, but enjoy their cover of this song

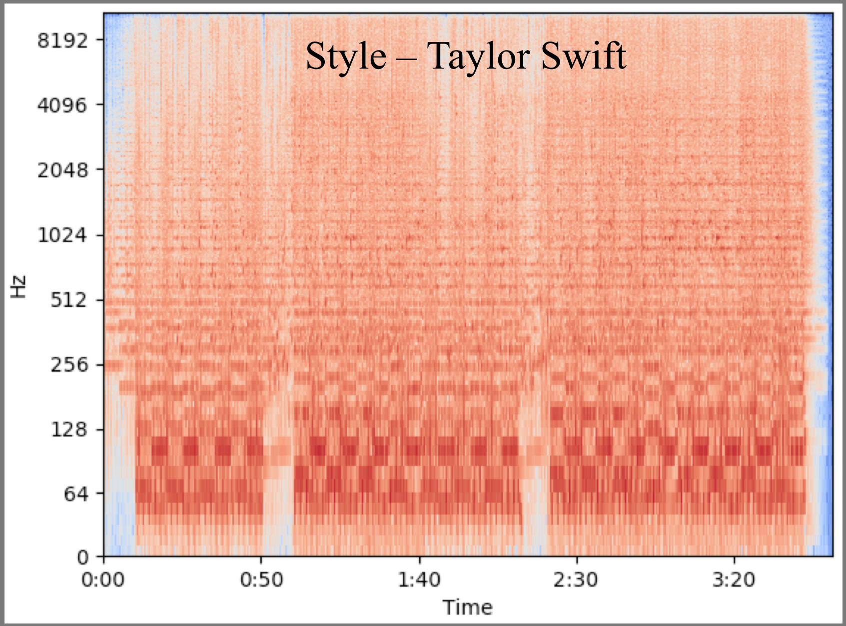
How can we get a computer to quantify all this?

# Musical Features – turning songs into data

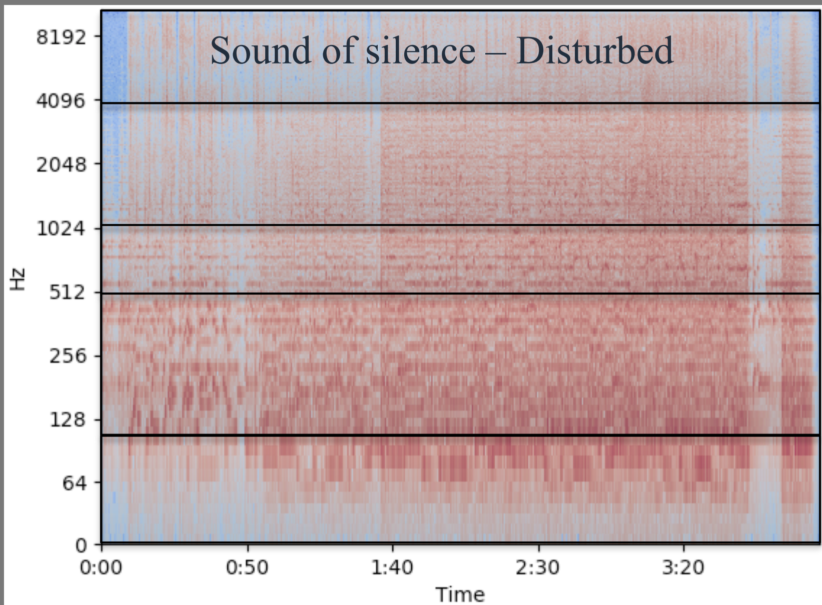


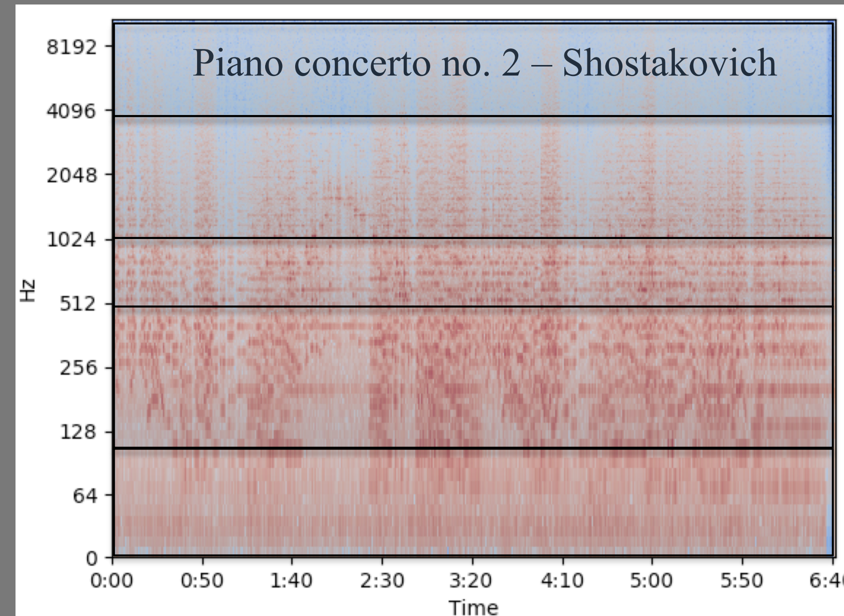



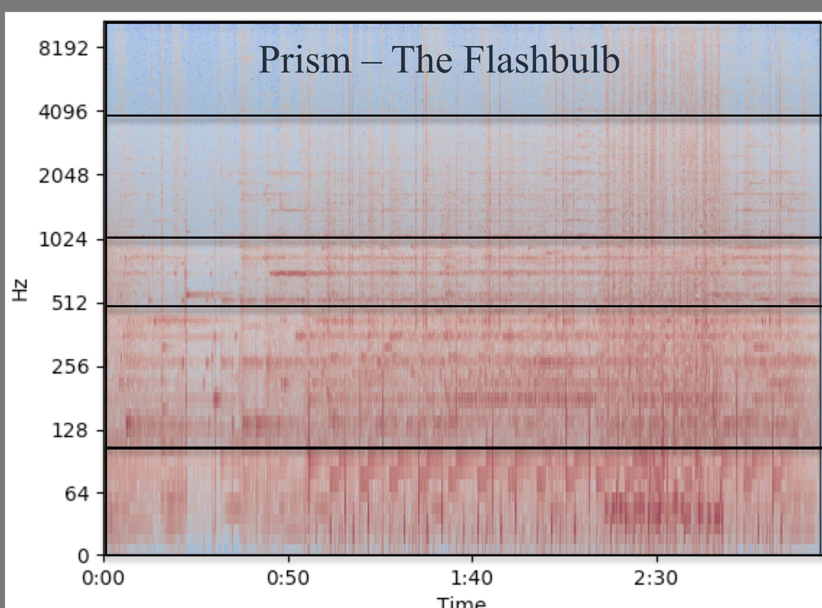



# Musical Features – turning songs into data



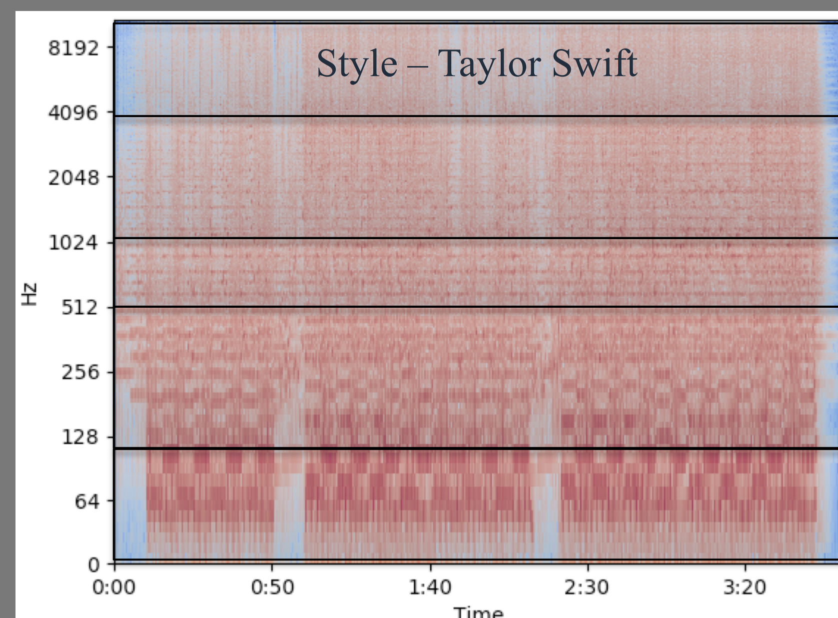

band 5

band 4

band 3

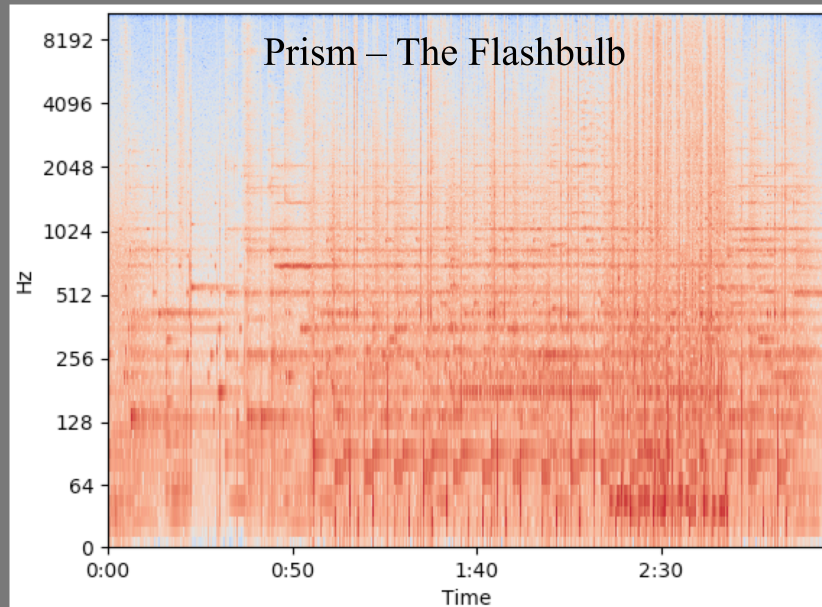
band 2

band 1



# Harmonic/Percussive separation

Median Filtering



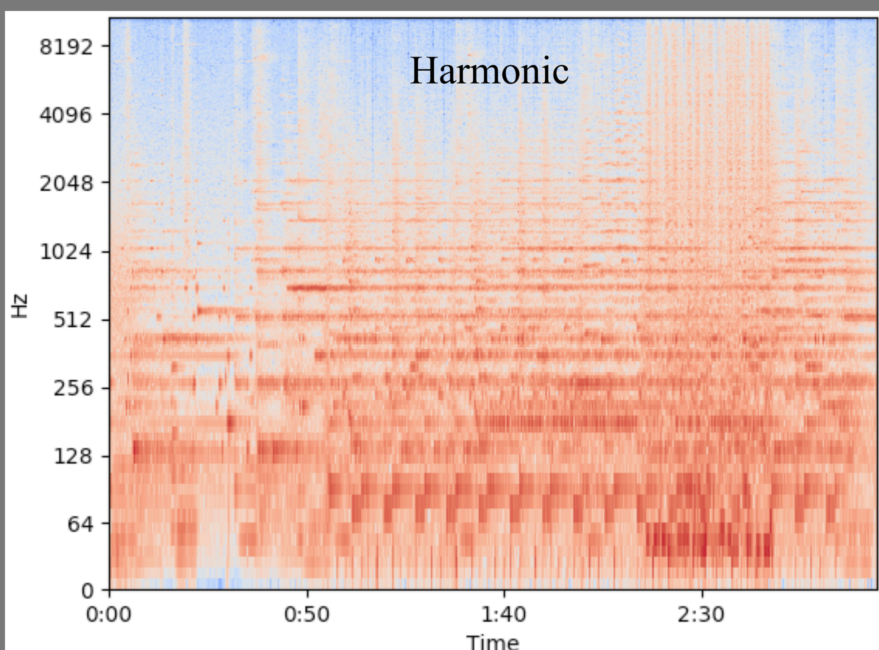

$$J(\mathbf{H},\mathbf{P}) = \frac{1}{2\sigma_H^2}\sum_{h,i}(H_{h,i-1} - H_{h,i})^2 + \frac{1}{2\sigma_P^2}\sum_{h,i}(P_{h-1,i} - P_{h,i})^2$$

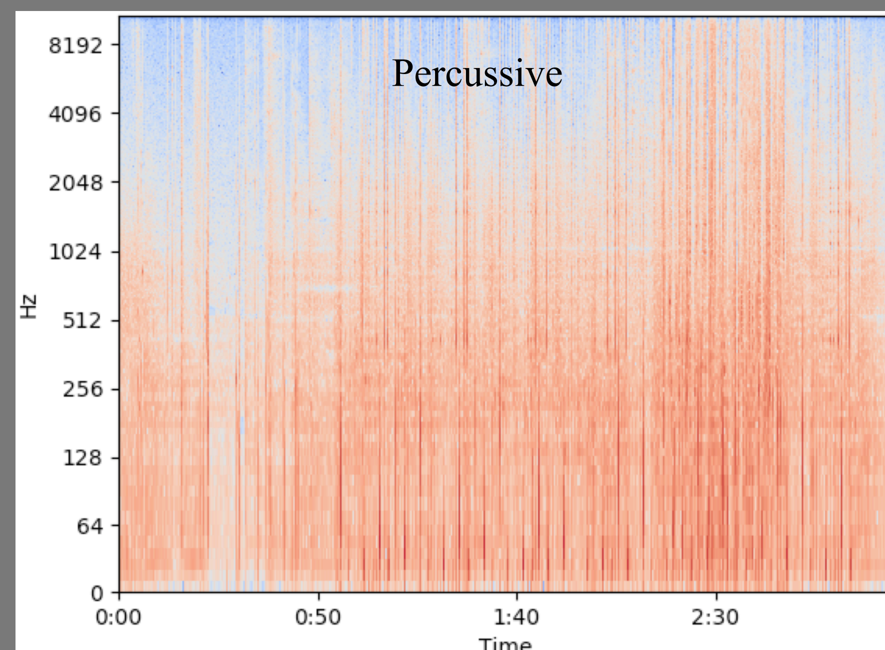
$$H_{h,i} + P_{h,i} = W_{h,i}$$

$$H_{h,i} \geq 0, \quad P_{h,i} \geq 0$$

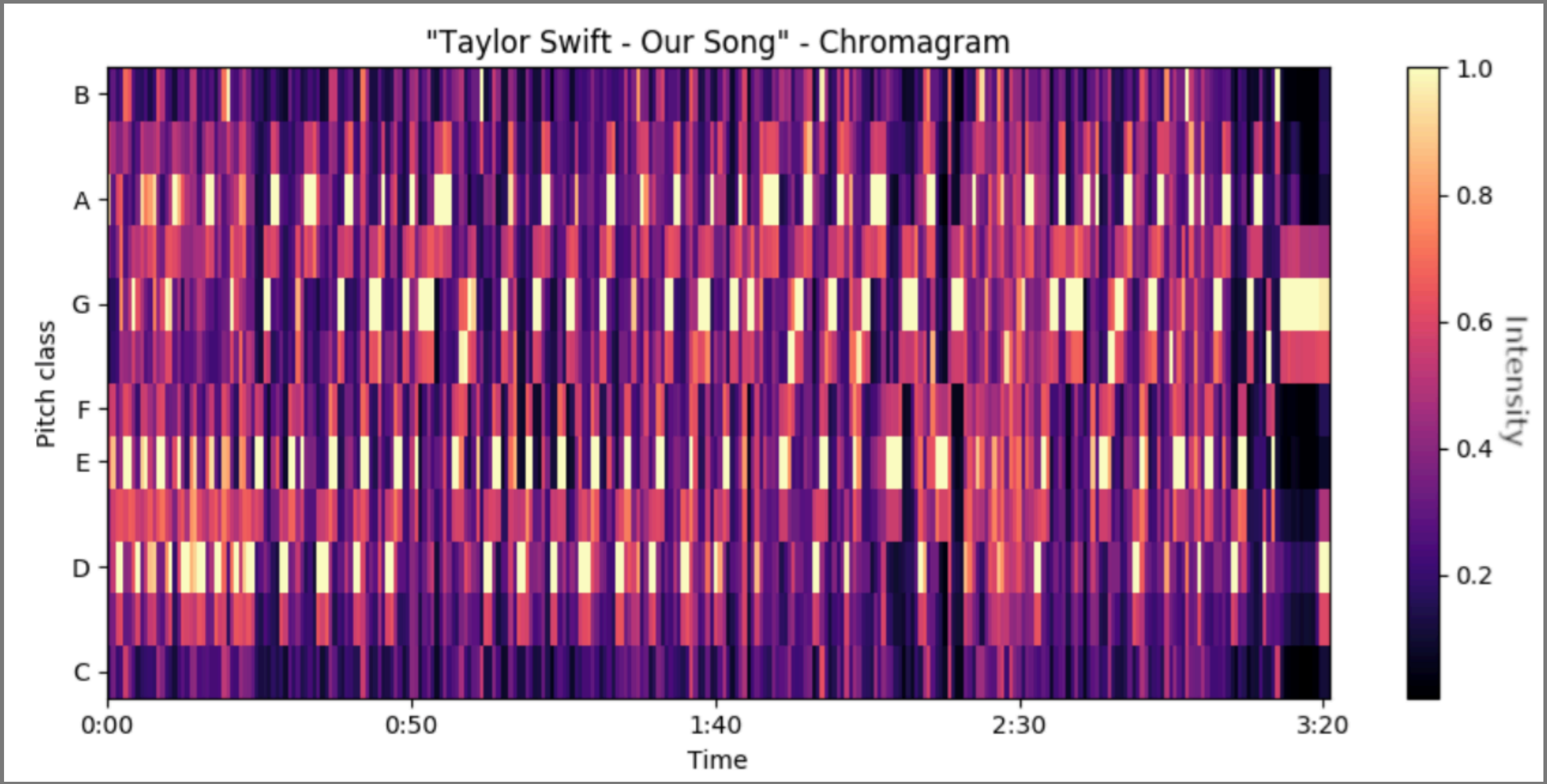
$$\tilde{\mathbf{W}} = \mathbf{W}^\gamma, \text{ where } 0 < \gamma \leq 1$$

Fitzgerald, Derry. "Harmonic/percussive separation using median filtering." 13th International Conference on Digital Audio Effects (DAFX10), Graz, Austria, 2010.

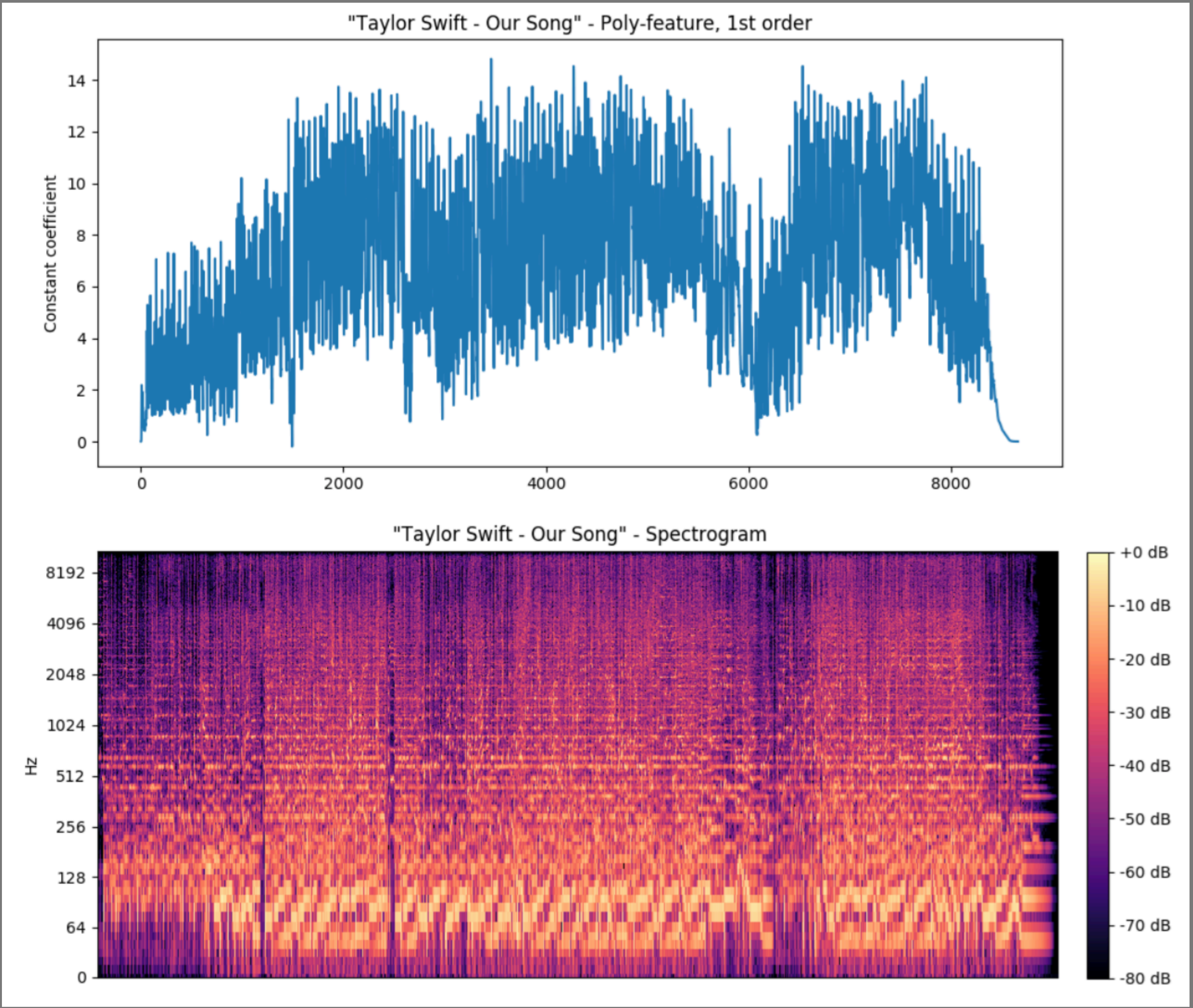


# Note classes



# Fitting polynomials to the spectrogram



# 419 Spectral and temporal features

## A.1 Spectral features

- RMS energy average - percussive
- RMS energy std - percussive
- RMS energy skew - percussive
- RMS energy kurtosis - percussive
- RMS energy average - harmonic
- RMS energy std - harmonic
- RMS energy skew - harmonic
- RMS energy kurtosis - harmonic
- spectral centroid average
- spectral centroid std
- bandwidth average
- bandwidth std
- spectral contrast average
- spectral contrast std
- poly-features gradient coefficient average
- poly-features gradient coefficient std
- poly-features constant coefficient average
- poly-features constant coefficient std
- zero-crossing rate average
- zero-crossing rate std
- onset average
- onset std
- tempo (beats per minute)

## A.3 Frequency-band features

These features were calculated for each band of 4, 11, and 24 logarithmically-spaced frequency bands.

- RMS energy average - harmonic
- RMS energy std - harmonic
- RMS energy skew - harmonic
- RMS energy kurtosis - harmonic
- RMS energy average - percussive
- RMS energy std - percussive
- RMS energy skew - percussive
- RMS energy kurtosis - percussive

## A.4 Time-domain windowing

These features were calculated for windows of width 0.5, 1, 3, and 5 seconds.

- poly-features gradient coefficient std
- poly-features gradient coefficient skew
- poly-features gradient coefficient kurtosis
- poly-features constant coefficient std
- poly-features constant coefficient skew
- poly-features constant coefficient kurtosis
- harmonic std
- harmonic skew
- harmonic kurtosis
- percussive std
- percussive skew
- percussive kurtosis

# 419 Spectral and temporal features

For each song, now have a list of 419 numbers

Now we can use Machine Learning to explore this data

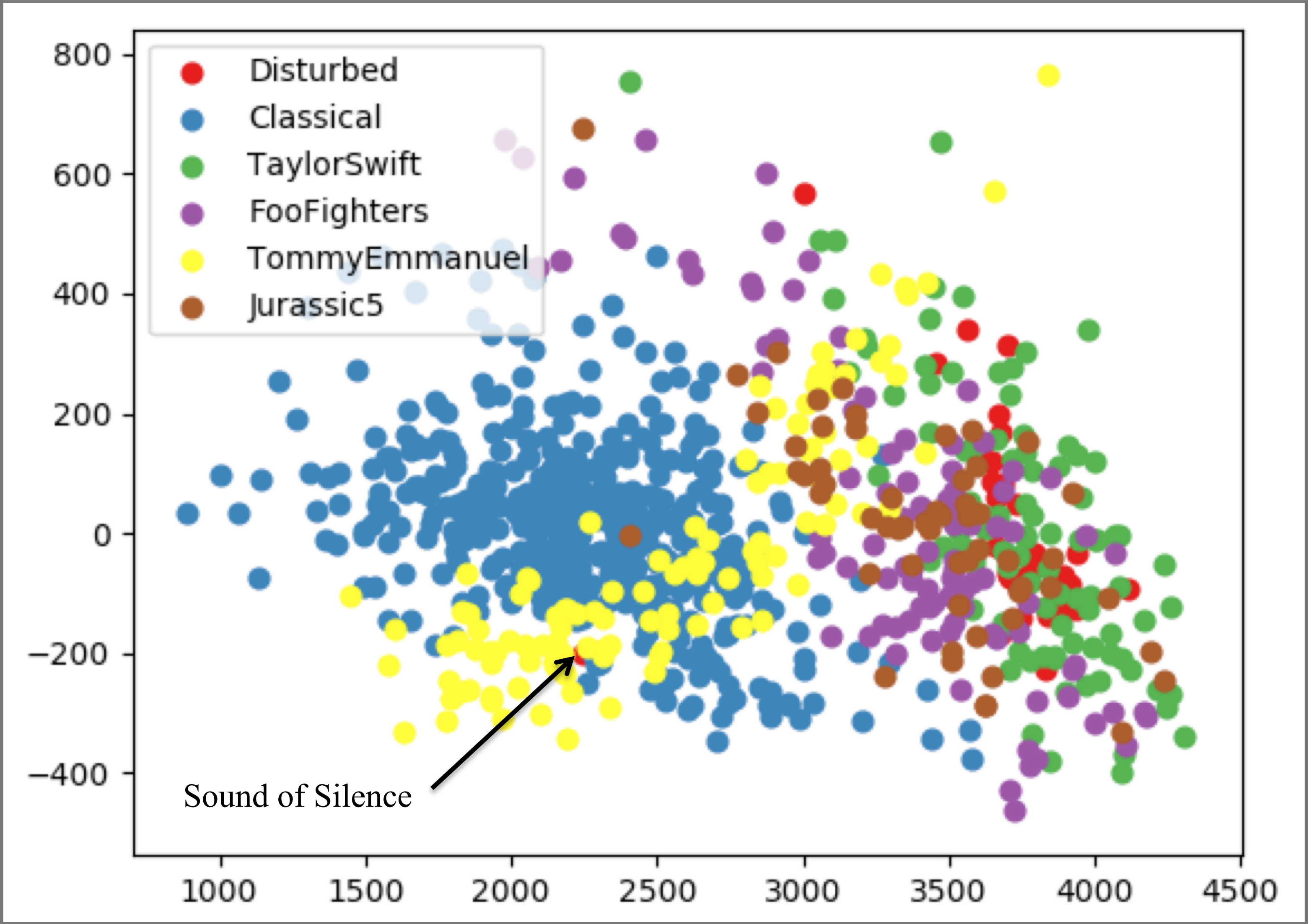
What kind of class labels might we want? Artists? Genres?

Start with unsupervised methods - clustering

# Unsupervised Clustering

Singular Value Decomposition – SVD

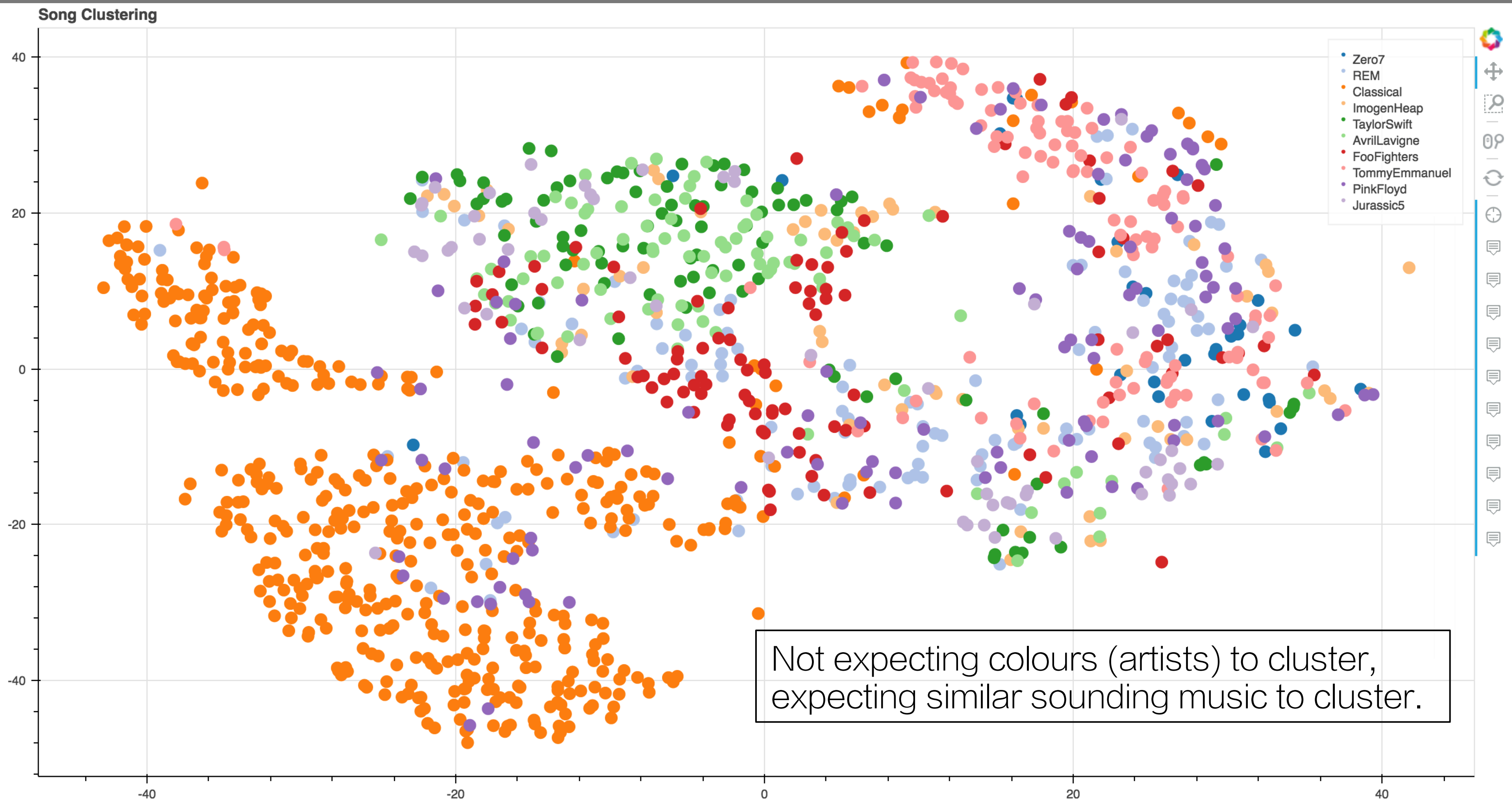
419 dimensions linearly reduced to 2 dimensions



# Unsupervised Clustering – going non-linear...

## t-distributed stochastic neighbor embedding (t-SNE)

Interactive version: github.com/informationcake/music-machine-learning



# Unsupervised Clustering – going non-linear...

t-distributed stochastic neighbor embedding (t-SNE)

## Class imbalance

The Flashbulb – 43 albums, 560 songs
Pink Floyd – 15 albums, 168 songs
REM – 9 albums, 120 songs
Taylor Swift – 5 albums, 95 songs
Avril Lavigne – 5 albums, 91 songs
Foo Fighters – 7 albums, 87 songs
Michael Jackson – 5 albums, 60 songs

.

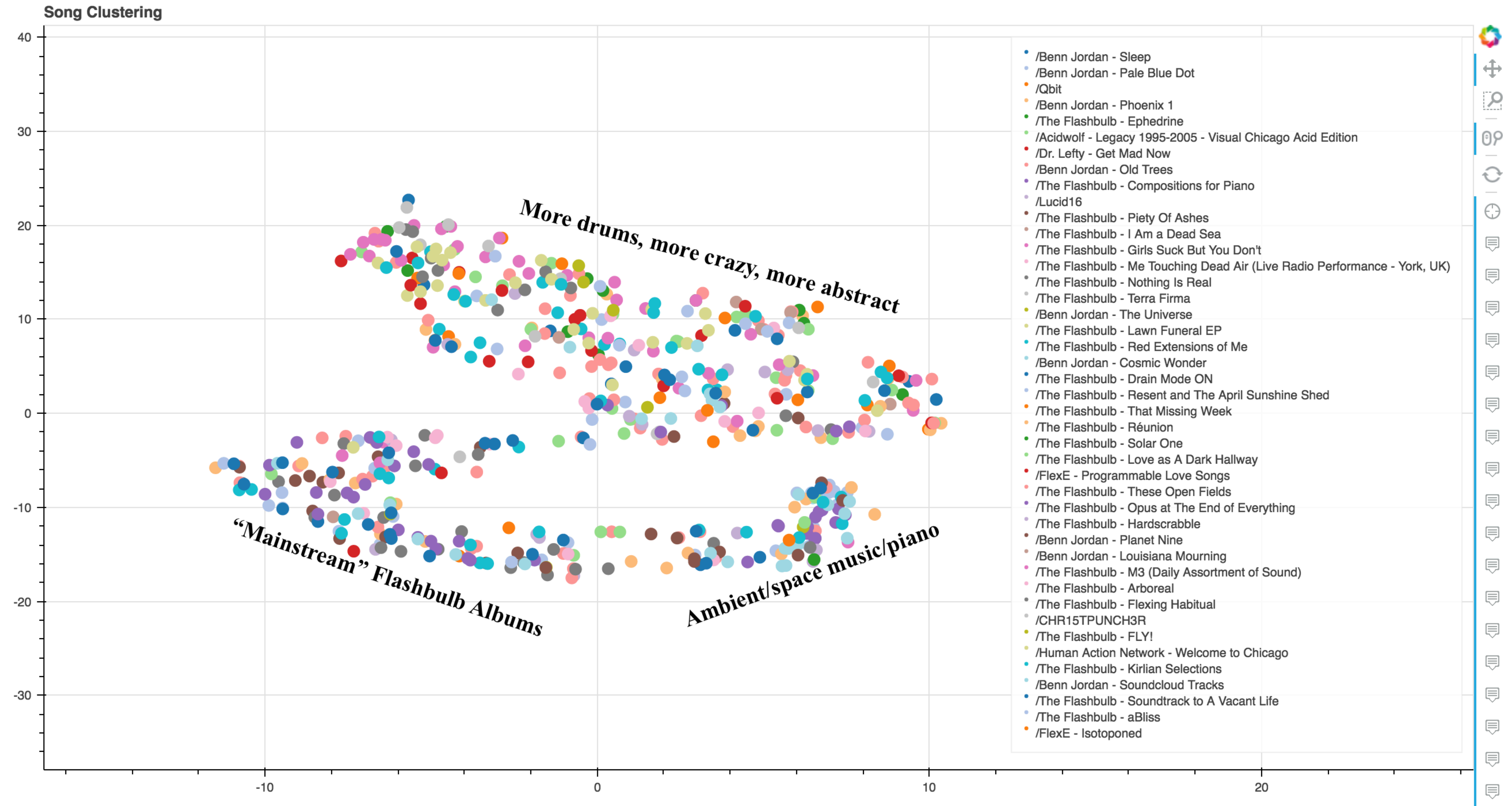
.

.

# Unsupervised Clustering – going non-linear...

## t-distributed stochastic neighbor embedding (t-SNE)

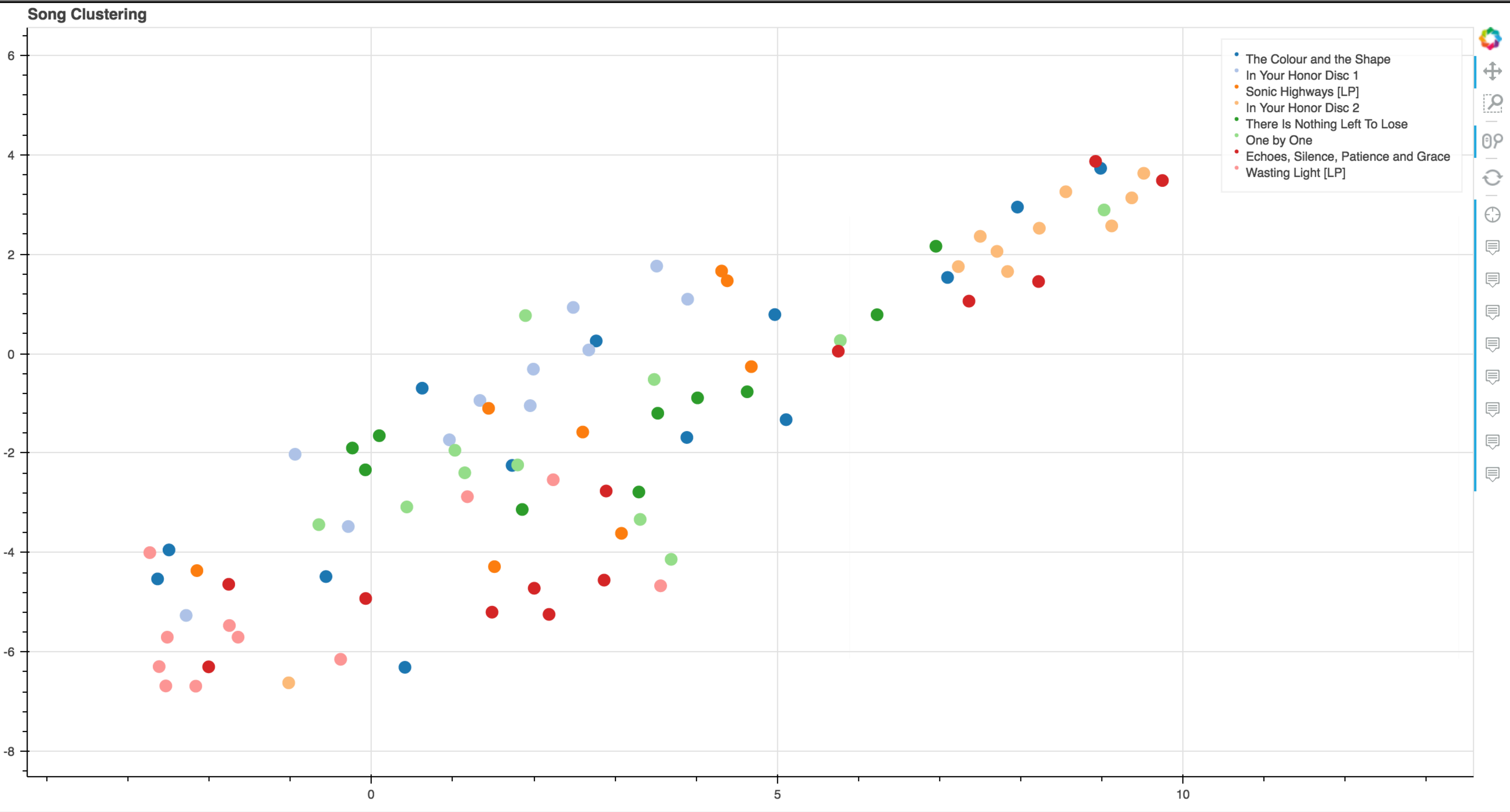
Individual artists: The Flashbulb



# Unsupervised Clustering – going non-linear...

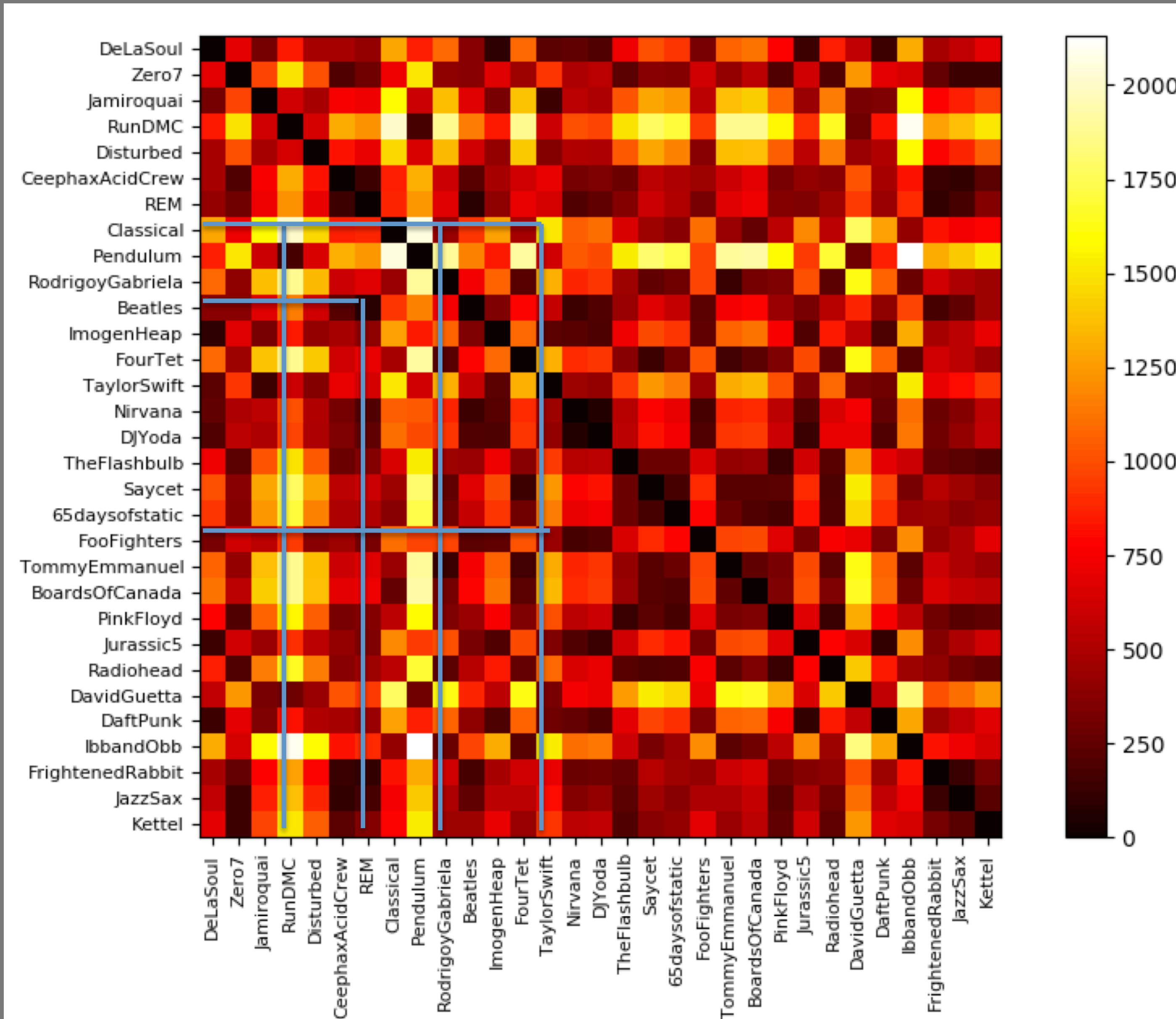
## t-distributed stochastic neighbor embedding (t-SNE)

Individual artists: Foo Fighters



# Classifiers

## Similarity matrix – Euclidian distance in 419 dimensions



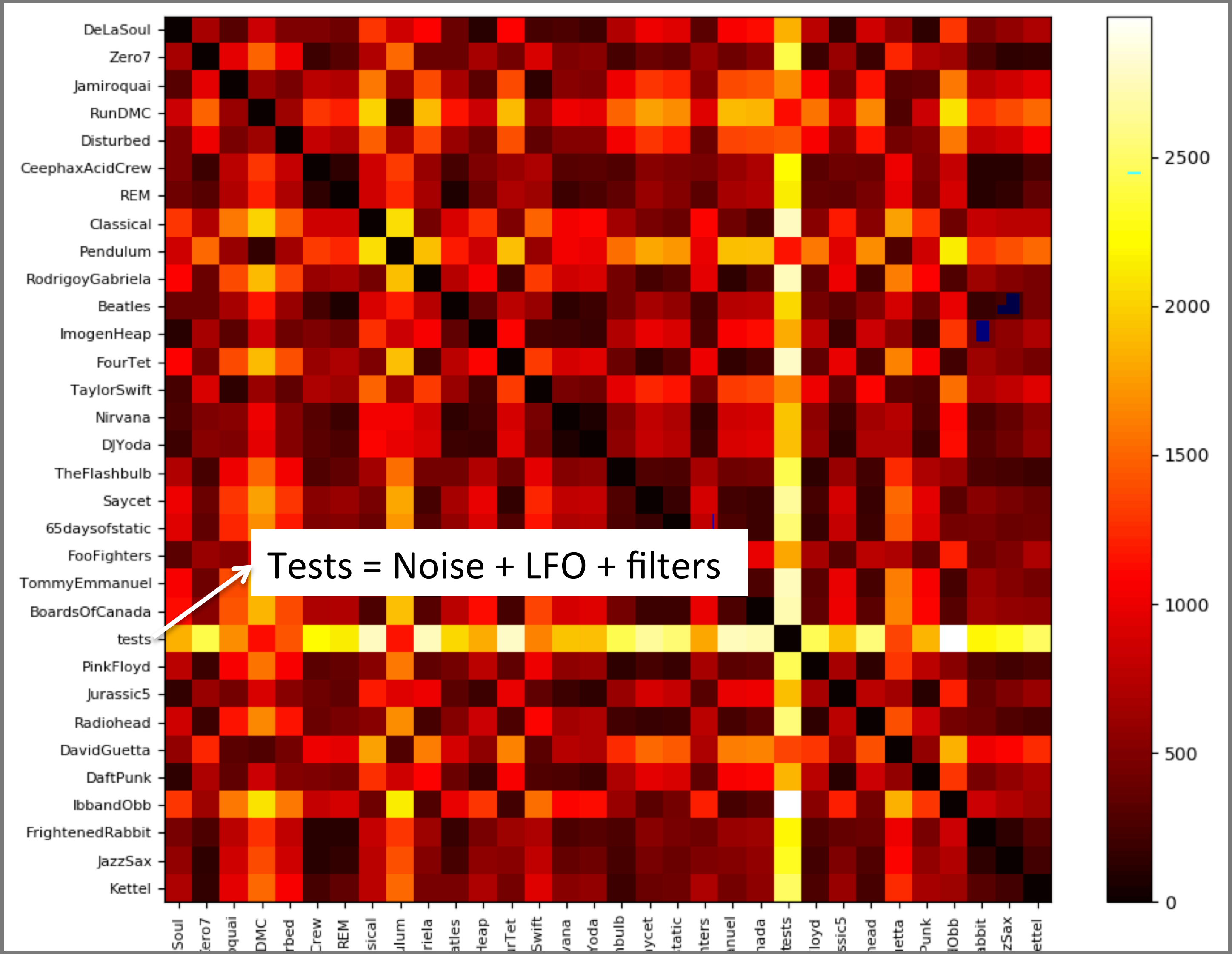

Different

Similar

Same

# Classifiers

Similarity matrix – Euclidian distance in 419 dimensions




Music...?

Different

Similar

Same

# Classical Music or Taylor Swift?

Take two artists

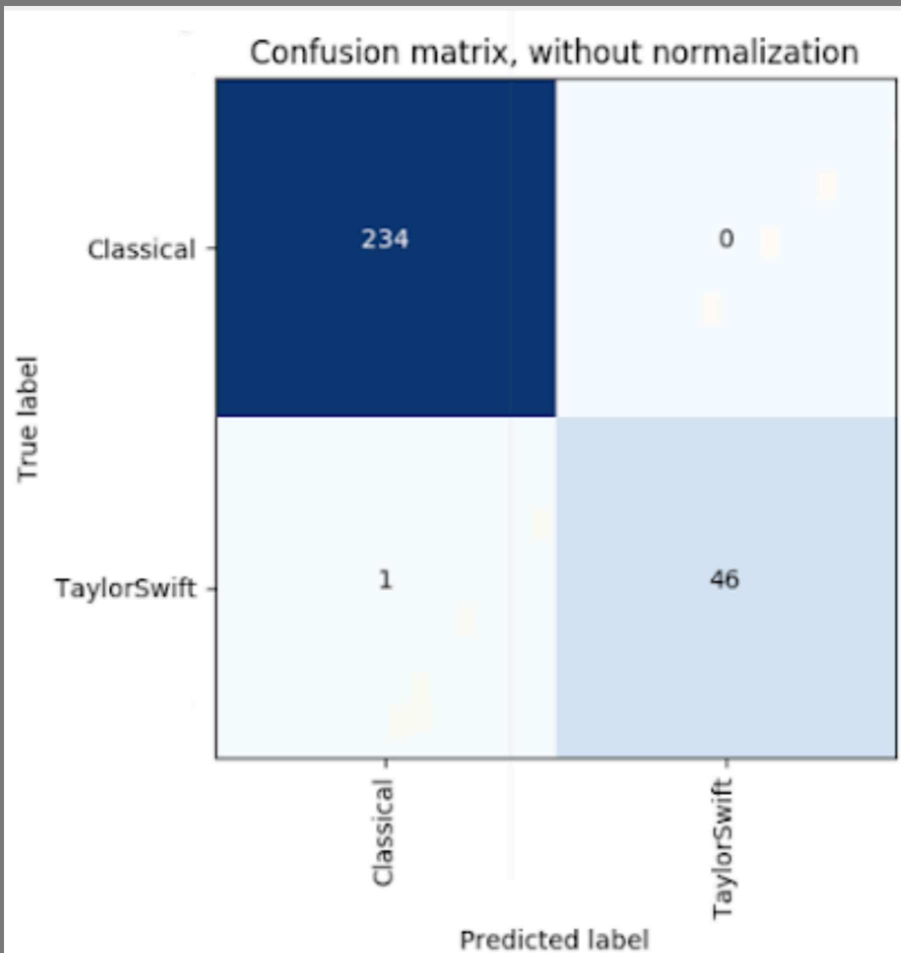
Give half of each artists songs to a Random Forest classifier to train on

Give the classifier the other half of each artists songs and let it predict who wrote them

Was it correct?

# Supervised Learning

## Classical Music or Taylor Swift?

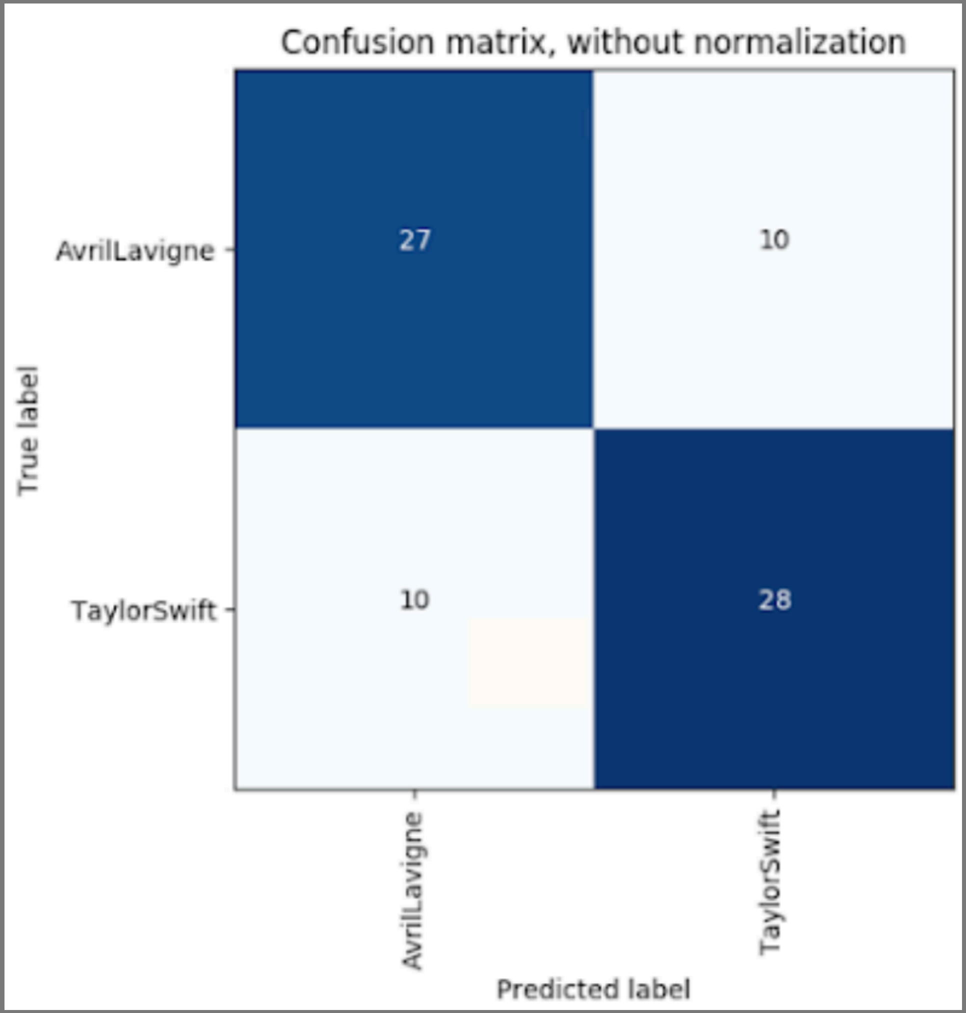



| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Classical | 1.00 | 1.00 | 1.00 | 234 |
| TaylorSwift | 1.00 | 0.98 | 0.99 | 47 |
| avg / total | 1.00 | 1.00 | 1.00 | 281 |

EASY!

# Supervised Learning

## Avril Lavigne or Taylor Swift?




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| AvrilLavigne | 0.73 | 0.73 | 0.73 | 37 |
| TaylorSwift | 0.74 | 0.74 | 0.74 | 38 |
| avg / total | 0.73 | 0.73 | 0.73 | 75 |

Harder... but still very good!

# Supervised Learning

## 12 musicians all together

| True label \ Predicted label | AvrilLavigne | BoardsOfCanada | Classical | FooFighters | ImogenHeap | MichaelJackson | PinkFloyd | REM | TaylorSwift | TheFlashbulb | TommyEmmanuel | Zero7 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AvrilLavigne | 17 | 0 | 0 | 7 | 0 | 1 | 0 | 2 | 8 | 1 | 0 | 0 |
| BoardsOfCanada | 0 | 36 | 14 | 0 | 2 | 0 | 3 | 0 | 0 | 31 | 0 | 0 |
| Classical | 0 | 0 | 183 | 0 | 0 | 0 | 3 | 0 | 0 | 1 | 0 | 0 |
| FooFighters | 0 | 0 | 0 | 29 | 0 | 0 | 1 | 4 | 1 | 4 | 0 | 1 |
| ImogenHeap | 0 | 1 | 1 | 0 | 11 | 0 | 0 | 1 | 0 | 10 | 0 | 0 |
| MichaelJackson | 0 | 0 | 0 | 1 | 0 | 16 | 0 | 0 | 1 | 6 | 0 | 0 |
| PinkFloyd | 0 | 5 | 1 | 0 | 1 | 0 | 33 | 3 | 0 | 0 | 0 | 0 |
| REM | 0 | 0 | 0 | 4 | 0 | 0 | 4 | 32 | 1 | 1 | 2 | 0 |
| TaylorSwift | 2 | 0 | 0 | 0 | 2 | 2 | 0 | 3 | 26 | 3 | 0 | 0 |
| TheFlashbulb | 1 | 4 | 2 | 0 | 0 | 0 | 2 | 1 | 0 | 214 | 0 | 0 |
| TommyEmmanuel | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 41 | 0 |
| Zero7 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 2 | 0 | 7 | 0 | 1 |

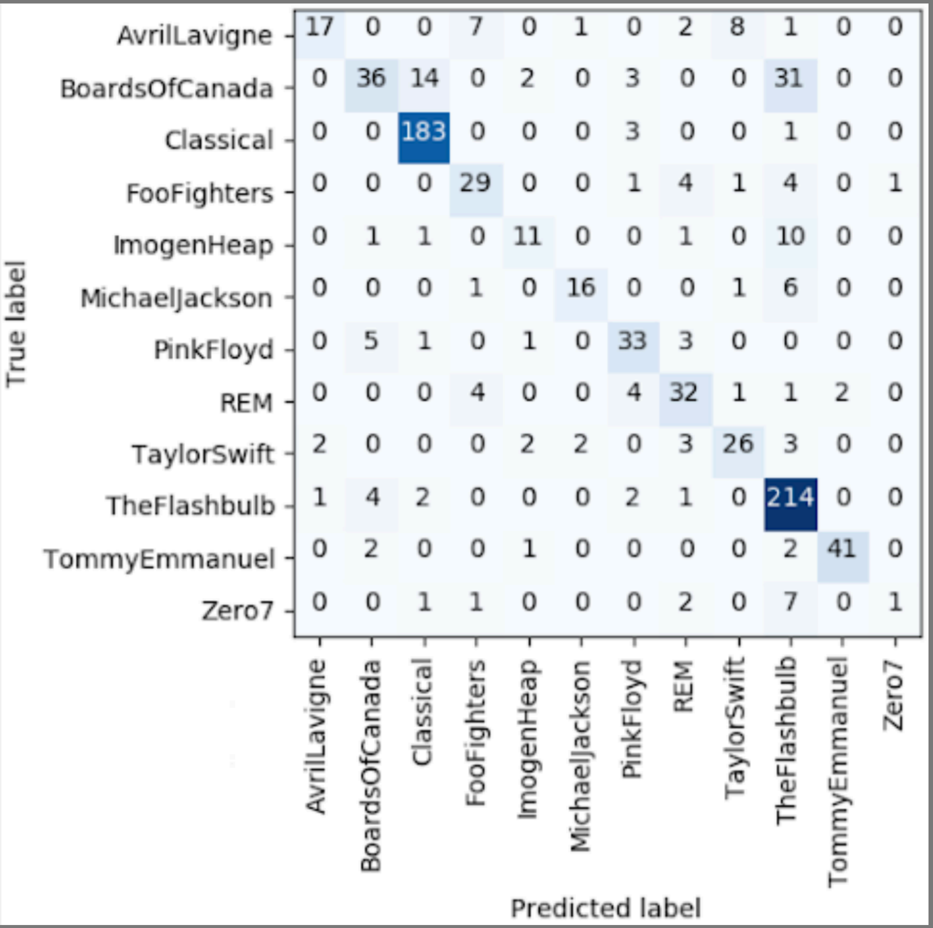
Predicted label

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| AvrilLavigne | 0.85 | 0.47 | 0.61 | 36 |
| BoardsOfCanada | 0.75 | 0.42 | 0.54 | 86 |
| Classical | 0.91 | 0.98 | 0.94 | 187 |
| FooFighters | 0.69 | 0.72 | 0.71 | 40 |
| ImogenHeap | 0.65 | 0.46 | 0.54 | 24 |
| MichaelJackson | 0.84 | 0.67 | 0.74 | 24 |
| PinkFloyd | 0.72 | 0.77 | 0.74 | 43 |
| REM | 0.67 | 0.73 | 0.70 | 44 |
| TaylorSwift | 0.70 | 0.68 | 0.69 | 38 |
| TheFlashbulb | 0.76 | 0.96 | 0.85 | 224 |
| TommyEmmanuel | 0.95 | 0.89 | 0.92 | 46 |
| Zero7 | 0.50 | 0.08 | 0.14 | 12 |
| avg / total | 0.79 | 0.79 | 0.78 | 804 |

Harder still... but overall great results!
Although some artists have relatively poor scores,
remember that random chance is now 1/11 (0. 09)

# Supervised Learning

## 12 musicians all together



| True label \ Predicted label | AvrilLavigne | BoardsOfCanada | Classical | FooFighters | ImogenHeap | MichaelJackson | PinkFloyd | REM | TaylorSwift | TheFlashbulb | TommyEmmanuel | Zero7 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AvrilLavigne | 17 | 0 | 0 | 7 | 0 | 1 | 0 | 2 | 8 | 1 | 0 | 0 |
| BoardsOfCanada | 0 | 36 | 14 | 0 | 2 | 0 | 3 | 0 | 0 | 31 | 0 | 0 |
| Classical | 0 | 0 | 183 | 0 | 0 | 0 | 3 | 0 | 0 | 1 | 0 | 0 |
| FooFighters | 0 | 0 | 0 | 29 | 0 | 0 | 1 | 4 | 1 | 4 | 0 | 1 |
| ImogenHeap | 0 | 1 | 1 | 0 | 11 | 0 | 0 | 1 | 0 | 10 | 0 | 0 |
| MichaelJackson | 0 | 0 | 0 | 1 | 0 | 16 | 0 | 0 | 1 | 6 | 0 | 0 |
| PinkFloyd | 0 | 5 | 1 | 0 | 1 | 0 | 33 | 3 | 0 | 0 | 0 | 0 |
| REM | 0 | 0 | 0 | 4 | 0 | 0 | 4 | 32 | 1 | 1 | 2 | 0 |
| TaylorSwift | 2 | 0 | 0 | 0 | 2 | 2 | 0 | 3 | 26 | 3 | 0 | 0 |
| TheFlashbulb | 1 | 4 | 2 | 0 | 0 | 0 | 2 | 1 | 0 | 214 | 0 | 0 |
| TommyEmmanuel | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 41 | 0 |
| Zero7 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 2 | 0 | 7 | 0 | 1 |

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| AvrilLavigne | 0.85 | 0.47 | 0.61 | 36 |
| BoardsOfCanada | 0.75 | 0.42 | 0.54 | 86 |
| Classical | 0.91 | 0.98 | 0.94 | 187 |
| FooFighters | 0.69 | 0.72 | 0.71 | 40 |
| ImogenHeap | 0.65 | 0.46 | 0.54 | 24 |
| MichaelJackson | 0.84 | 0.67 | 0.74 | 24 |
| PinkFloyd | 0.72 | 0.77 | 0.74 | 43 |
| REM | 0.67 | 0.73 | 0.70 | 44 |
| TaylorSwift | 0.70 | 0.68 | 0.69 | 38 |
| TheFlashbulb | 0.76 | 0.96 | 0.85 | 224 |
| TommyEmmanuel | 0.95 | 0.89 | 0.92 | 46 |
| Zero7 | 0.50 | 0.08 | 0.14 | 12 |
| avg / total | 0.79 | 0.79 | 0.78 | 804 |

Removing some classes (artists) can significantly affect the scores for artists with small training sets (e.g. Zero7)

# Supervised Learning

## ‘Xmas music’

| True label \ Predicted label | AvrilLavigne | Classical | FooFighters | ImogenHeap | MichaelJackson | PinkFloyd | REM | TaylorSwift | TheFlashbulb | TommyEmmanuel | Xmas | Zero7 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AvrilLavigne | 17 | 0 | 8 | 0 | 1 | 0 | 0 | 6 | 0 | 0 | 4 | 0 |
| Classical | 0 | 177 | 0 | 0 | 0 | 2 | 0 | 0 | 3 | 0 | 5 | 0 |
| FooFighters | 1 | 0 | 28 | 0 | 0 | 0 | 2 | 1 | 3 | 0 | 5 | 0 |
| ImogenHeap | 0 | 0 | 0 | 5 | 0 | 1 | 1 | 0 | 9 | 0 | 8 | 0 |
| MichaelJackson | 0 | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 3 | 0 | 7 | 0 |
| PinkFloyd | 0 | 3 | 0 | 0 | 0 | 22 | 0 | 0 | 1 | 1 | 16 | 0 |
| REM | 0 | 0 | 3 | 0 | 0 | 1 | 23 | 1 | 3 | 1 | 12 | 1 |
| TaylorSwift | 4 | 0 | 0 | 0 | 3 | 0 | 0 | 22 | 2 | 0 | 7 | 0 |
| TheFlashbulb | 0 | 5 | 0 | 0 | 1 | 0 | 1 | 1 | 210 | 4 | 2 | 0 |
| TommyEmmanuel | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 | 36 | 6 | 0 |
| Xmas | 2 | 10 | 4 | 0 | 0 | 3 | 3 | 1 | 7 | 0 | 98 | 1 |
| Zero7 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 5 | 0 | 1 | 4 |

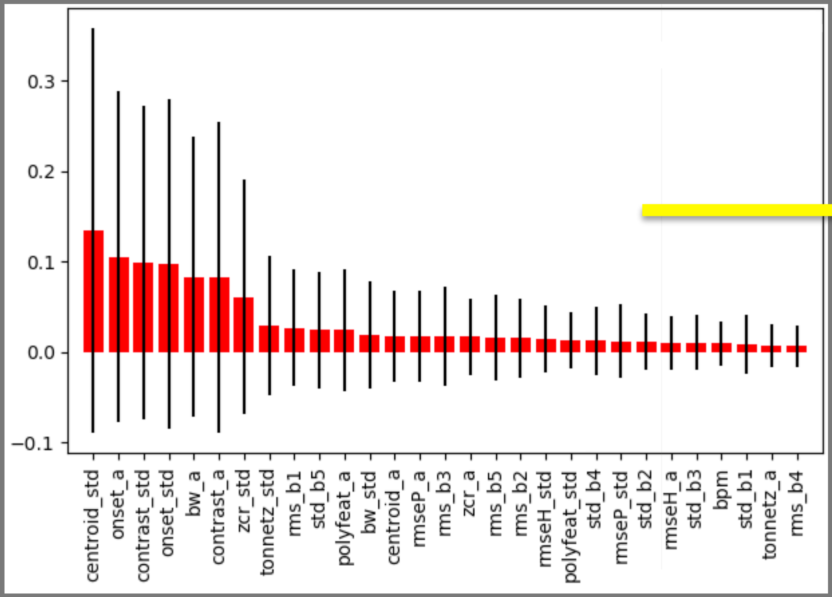
Note that Xmas music is often confused with Classical/Flashbulb/...

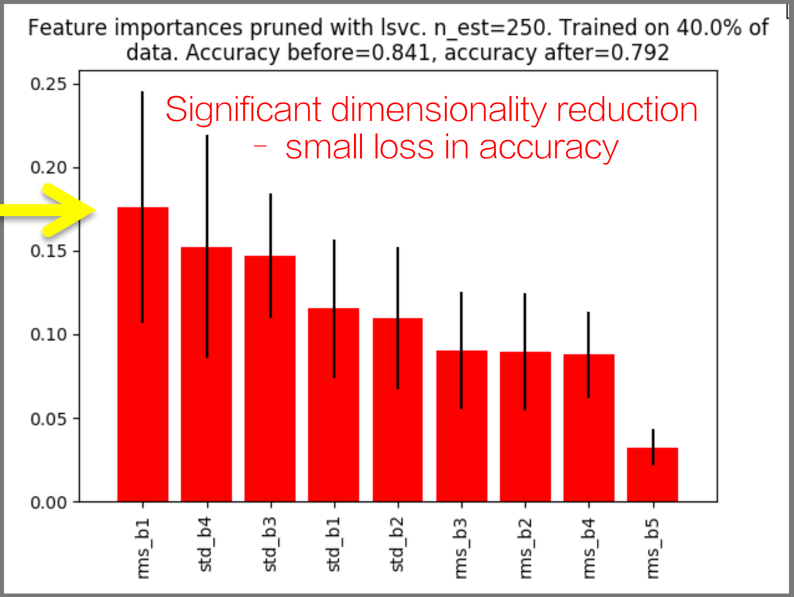
Pink Floyd and REM are often confused for being Xmas tunes...

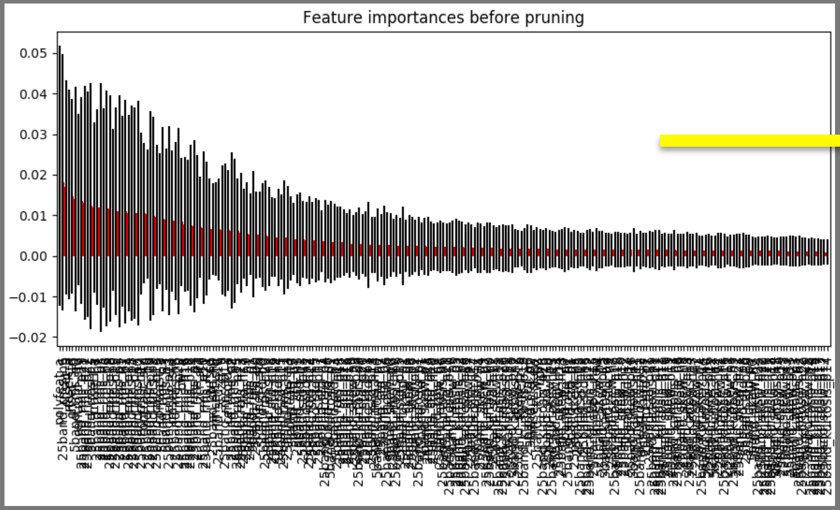
# Feature Importance – domain expertise matters

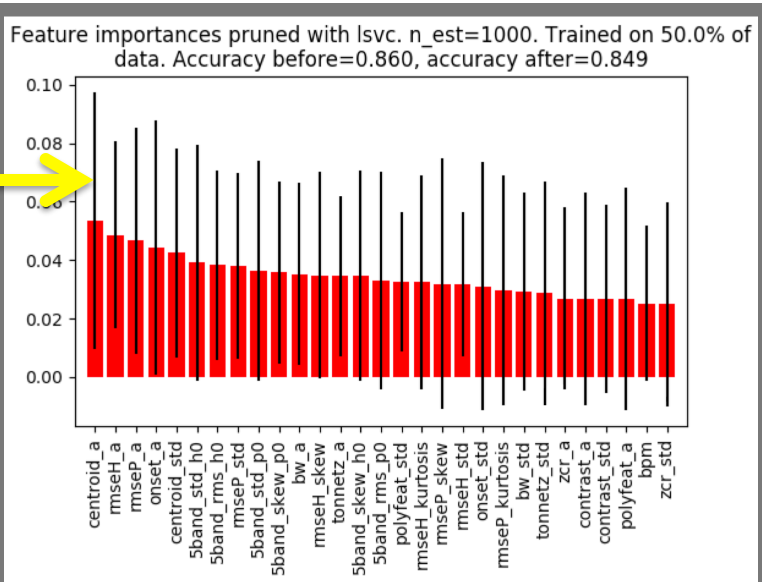
We want features to easily distinguish between different songs/artists

How many features to use? We can see which features were important in the model:



# Supervised Learning

## The music behind adverts?

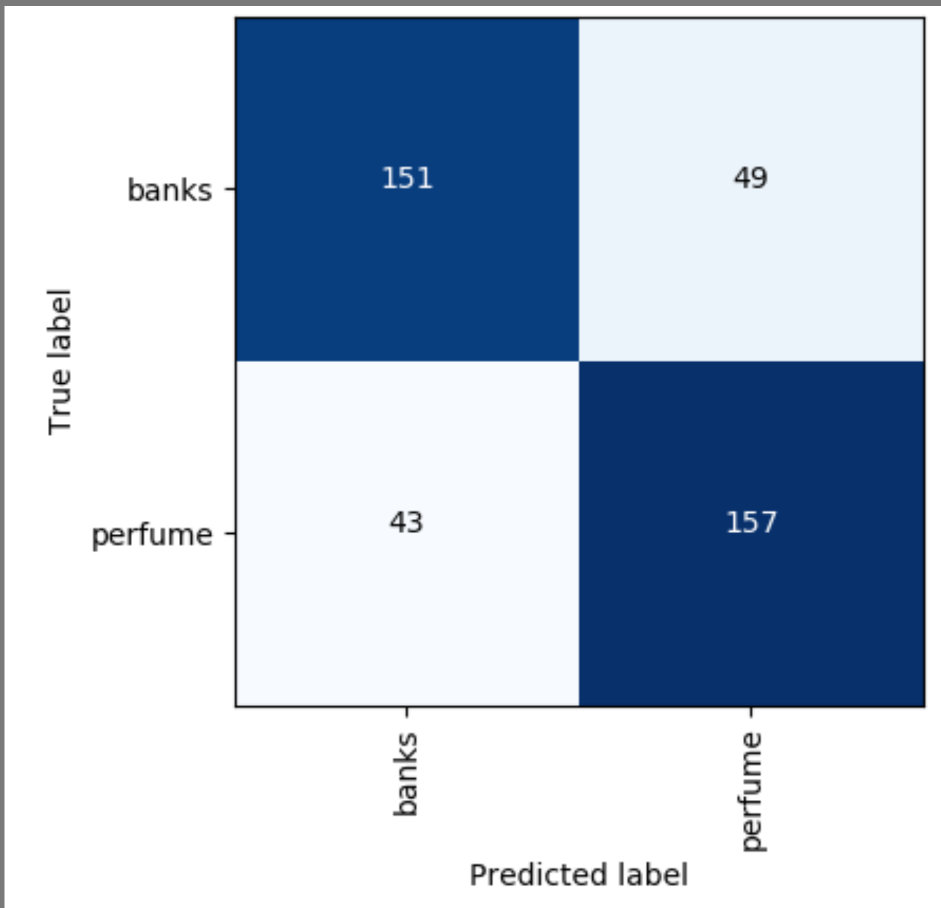
| Class name | Subclass names | | | | | # videos |
|---|---|---|---|---|---|---|
| Alcohol | Carling | Carlsberg | Fosters | Heineken | Strongbow | 500 |
| Banks | Barclays | Halifax | HSBC | Nationwide | Scottish Widows | 500 |
| Cars | Audi | Honda | Nissan | Renault | VW | 500 |
| Perfume | Bvlgari | Chanel | Dior | Gucci | Hugo Boss | 500 |
| Supermarkets | Asda | Morrisons | Sainsburys | Tesco | Waitrose | 500 |
| | | | | | Total support: | 2500 |

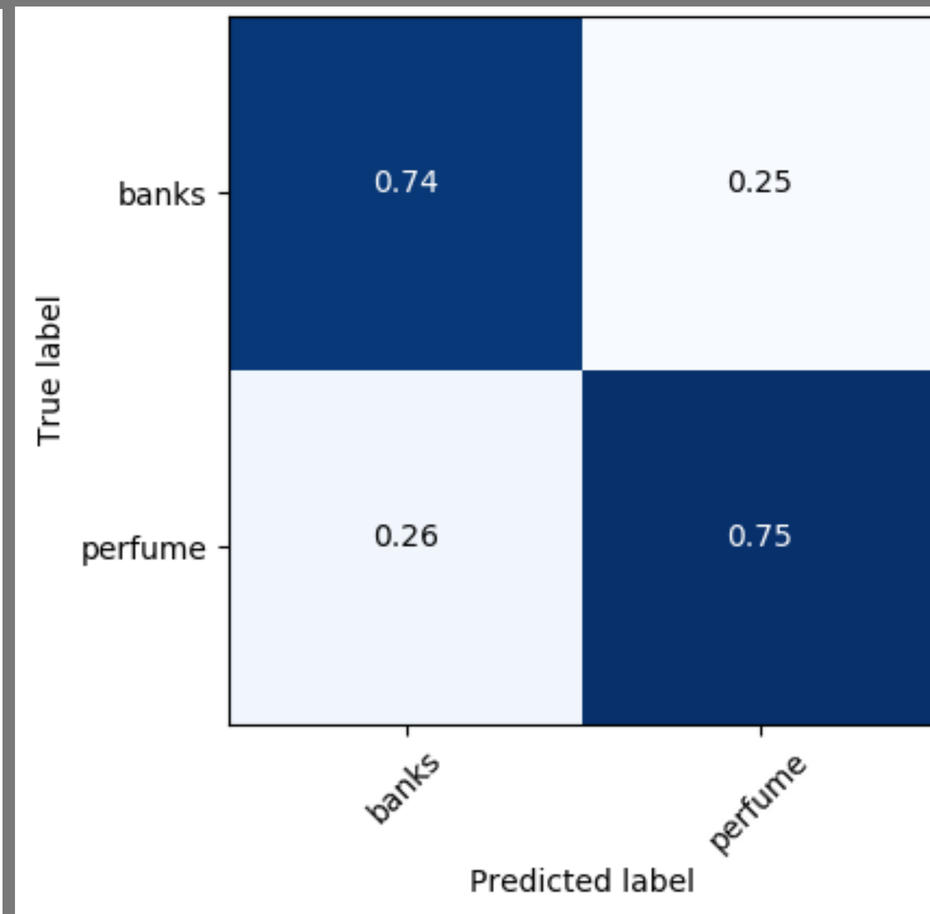
Train on 300 of them, test the remaining 200

*Work done by my summer student Naim Sen*

# Supervised Learning

Is it a band advert? Is it a perfume advert?



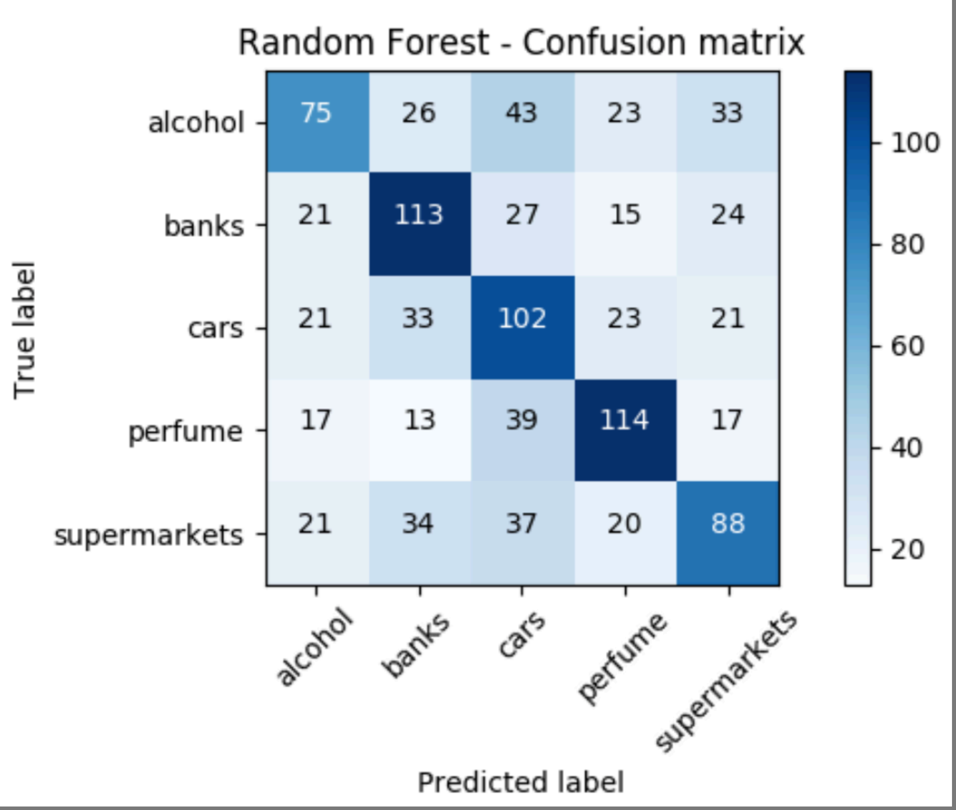




Precision = 0.75, Recall = 0.77

*Work done by my summer student Naim Sen*

# Supervised Learning

## The music behind adverts?




| Class | Precision | Recall | f1-score | Support |
|---|---|---|---|---|
| Alcohol | 0.48 | 0.38 | 0.42 | 200 |
| Banks | 0.52 | 0.56 | 0.54 | 200 |
| Cars | 0.41 | 0.51 | 0.46 | 200 |
| Perfume | 0.58 | 0.57 | 0.58 | 200 |
| Supermarkets | 0.48 | 0.44 | 0.46 | 200 |
| Mean | 0.50 | 0.49 | 0.49 | 1000 |

Random chance classification is 1/5, so scores are encouraging

What is the maximum expected performance?

Is specific music genuinely chosen for ads? Consciously/unconsciously?

Could a marketing company make music choices to better distinguish their adds?

*Work done by my summer student Naim Sen*

# Supervised Learning

## Voices: Women & Men

Can you work out gender based on voice alone? It is not as easy as it sounds...

You have probably never needed to train your brain to do this!

Is binary classification relevant? A better answer is a scale from 0 – 1.

"do you have a masculine or feminine voice?"

Scraping 1000 videos from Youtube with search terms like:
"pregnancy reviews"
"Drill reviews"
Gives Precision & Recall scores of around 0.8

Select TED talks from Youtube, but manually check if a male or female speaker.

A selection of 50 videos gives a Precision & Recall 0.98

Are these models transferable? Can you classify other recordings successfully?

Aim is to have an algorithm that can take audio clips and tell you the % a man or woman was speaking.

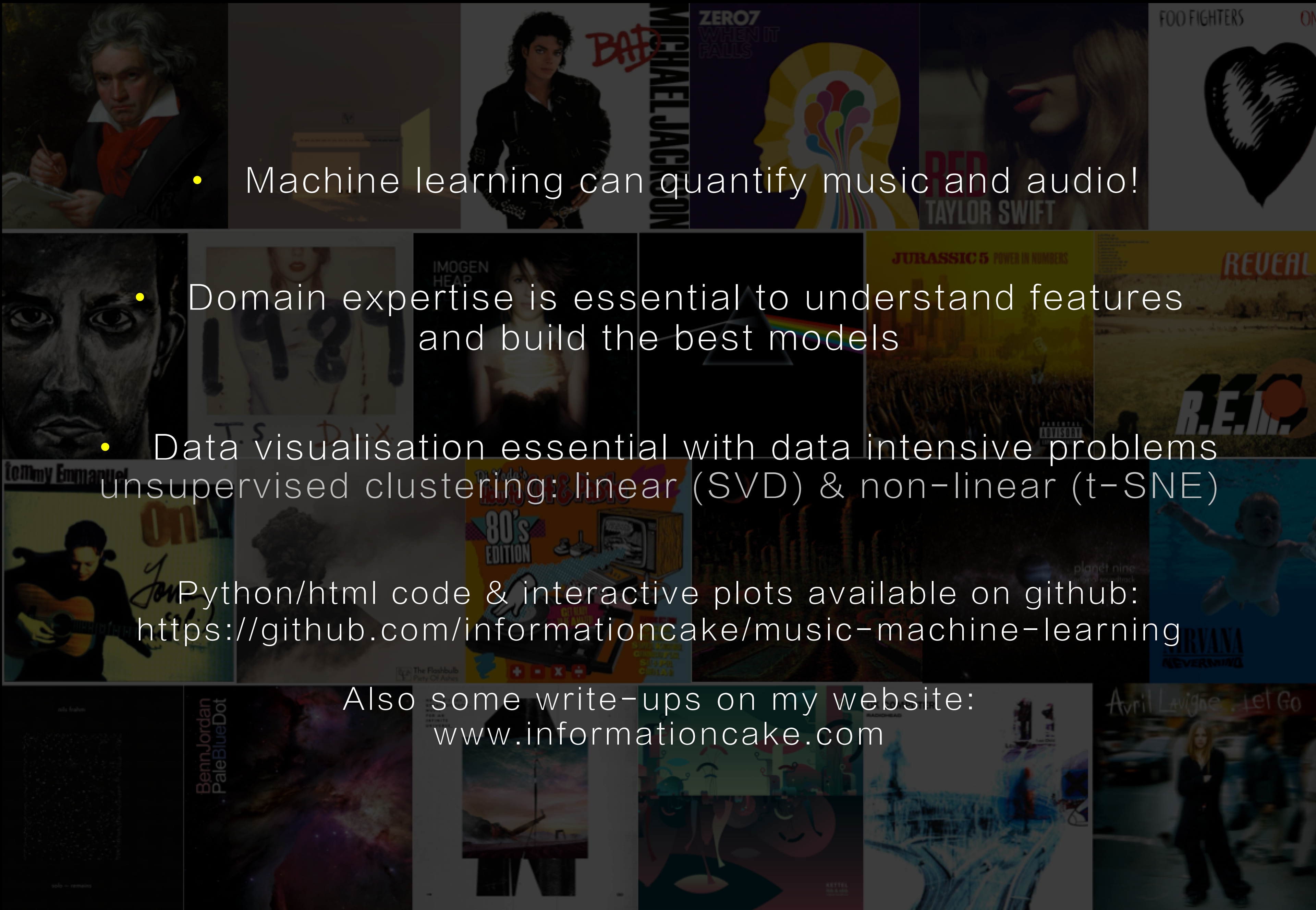
Preliminary tests on our department astronomy podcast (the Jodcast) are encouraging!

*Work done by my MPhys students Ana Zardoshti and Chloe Hutton*

# Benchmarking is difficult

The Piczak ESC-50 data set (a library of urban and natural sounds) has an accuracy of 81% for (*un-trained*) human listeners

# Machine learning with music




- Machine learning can quantify music and audio!

- Domain expertise is essential to understand features and build the best models

- Data visualisation essential with data intensive problems

unsupervised clustering: linear (SVD) & non-linear (t-SNE)

Python/html code & interactive plots available on github:
https://github.com/informationcake/music-machine-learning

Also some write-ups on my website:
www.informationcake.com