

# ♪ jazz songs, predictive modeling with LSTM

## dataset

webscraping of 456 songs and 22 columns with jazz information

source: "The Jazzomat Research Project"

No real patterns found

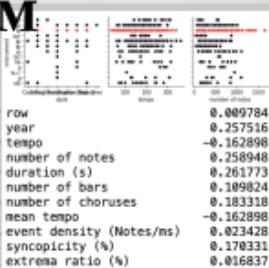
2. looking at which 'ts' songs are the biggest outliers. Because the more the songs are alike, the better the model works

tempo	notes	duration	bars	choruses	mean tempo	event density (Notes/ms)	syncoplicity (%)	extrema ratio (%)
new	456	non-mel11	are04	0.009784				
title	456	non-mel11	0.257516					
style	456	non-mel11	0.162898					
year	456	non-mel11	0.261773					
tempo	456	non-mel11	0.258948					
notes	456	non-mel11	0.189824					
key	456	non-mel11	0.183318					
duration	456	non-mel11	0.162898					
bars	456	non-mel11	0.023428					
choruses	456	non-mel11	0.178331					
mean tempo	456	non-mel11	0.016837					
event density (Notes/ms)	456	non-mel11	0.009784					
syncoplicity (%)	456	non-mel11	0.257516					
extrema ratio (%)	456	non-mel11	0.162898					

## data exploring

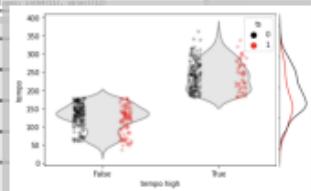
Trying to figure out which songs to take with me next step

1. correlation matrix with dependent variable (inst)



## preprocessing

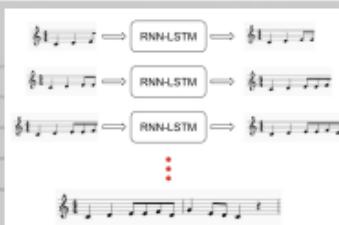
Looking at chosen 113 midi files



[10. 0. 0. ... 0. 0. 0.]  
[1. 0. 0. ... 0. 0. 0.]  
[1. 0. 0. ... 0. 0. 0.]  
  
...  
  
[1. 0. 0. ... 0. 0. 0.]  
[0. 0. 0. ... 0. 0. 0.]  
[0. 0. 1. ... 0. 0. 0.]]  
[37 6 0 ... 46 46 46]  
r r r 68 62 55 57 r 52  
r r r 74 74 74 r r r  
62 66 59 62 64 59 52 57 57  
66 69 \_ r \_ 65 r 71 71 r

## predictive model

After generative sequences the model will be build. And it will predict the next note in the sequence. And repeat



While building the model, more exploratory data analysis will be used and the overall results will be put into a midi

