Learning a Playlist Representation For Playlist Generation

Spotify

Spotify

The world's leading music streaming provider

- Provides a music streaming service
- Free version with ads
- Premium version for a fee
- 75 million active users
- 20 million paying users

Spotify: The Soundtracks of Our Lives

Playlists, a popular way of consuming music

- **Millions of songs, Billions of playlists**
- **Itunes store controversy**
- **Tunigo (*Browse* Playlists)**
- “Playlists' uses include allowing a particular desired musical atmosphere to be created and maintained without constant user interaction, or to allow a variety of different styles of music be played, again without maintenance.” ***Wikipedia***

Thesis Outline

Project goal

- Make an initial step towards fully automated playlist generation
- In the form of candidate song selection

But

- Playlist generation is recommendation
- Recommender systems already exist

Recommender Systems

Collaborative Filtering

-

	Item.1	Item.2	Item.3	Item.4	Item.5	Item.6	Item.7
User 1	4	1	?	3	2	3	2
User 2	3	4	?	3	4	?	4
User 3	4	2	5	4	2	2	3
User 4	?	?	1	3	1	3	5
User 5	3	5	?	1	2	3	2
User 6	2	4	4	4	2	3	2
User 7	3	2	2	4	4	?	?

Recommender Systems

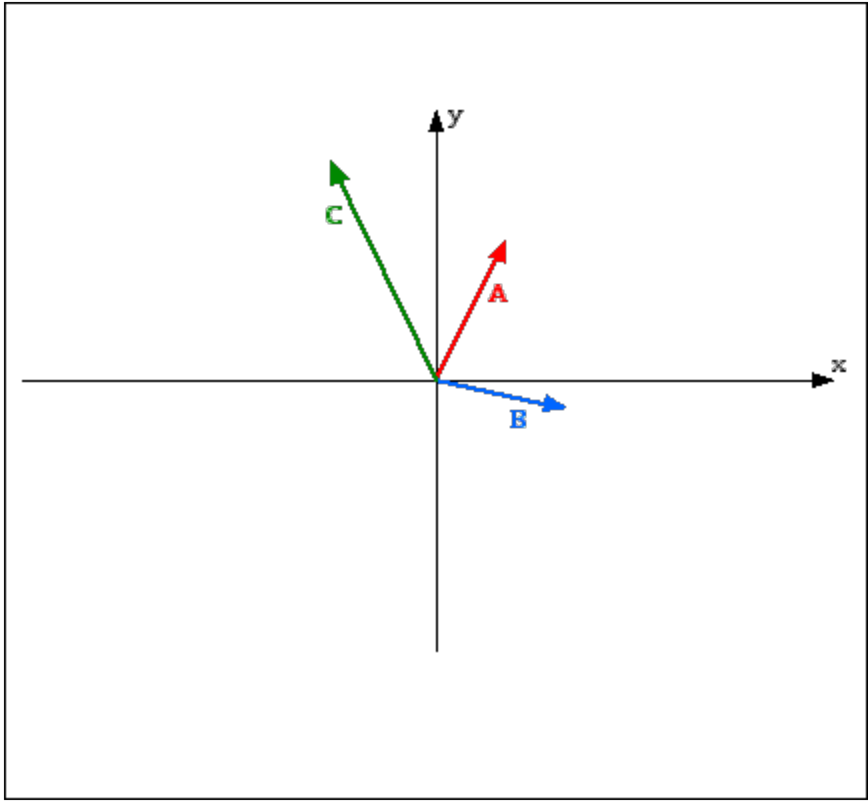
Collaborative Filtering

- Suffers from the cold start problem
- A previously unranked item cannot be recommended
- What if 20 000 songs are added each day..

Recommender Systems

Content Based Recommendation

- Vector space model
- Comparing items based on cosine similarity



Recommendation

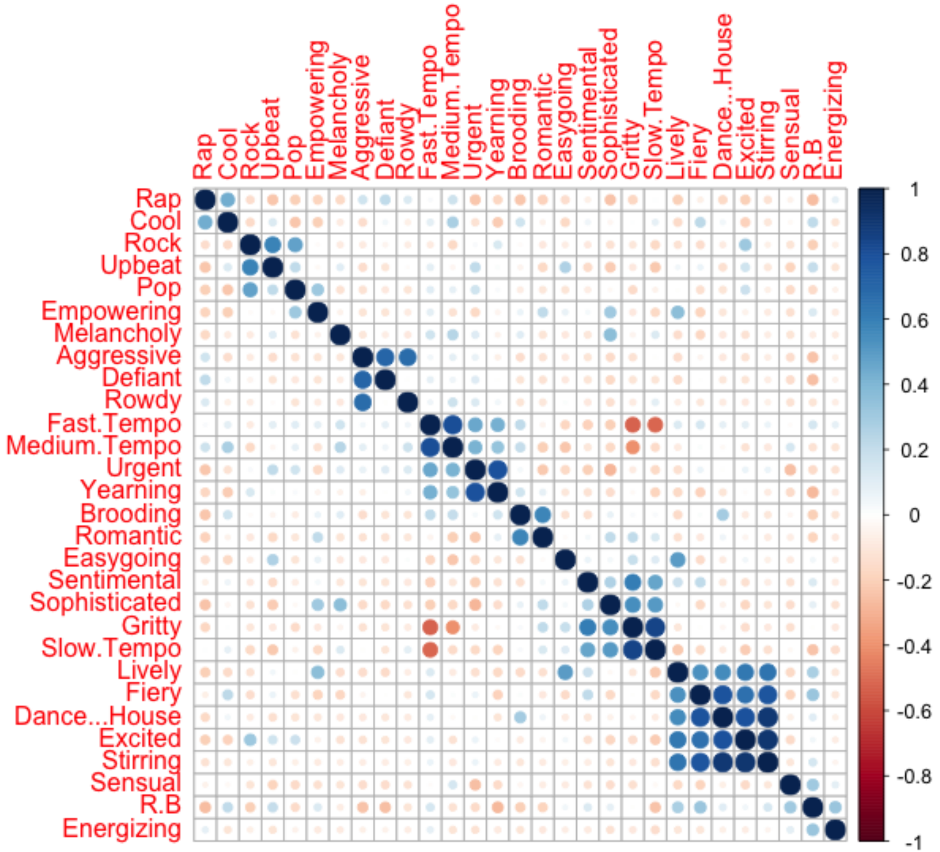
Downsides

- Traditional recommendation typically focuses on tracks, albums, items
- What if a playlist could be generated, with a seed playlist as training data?

Playlist Characteristics

What characterizes a playlist?

- **Variance**
- **How do features co-vary in a playlist?**



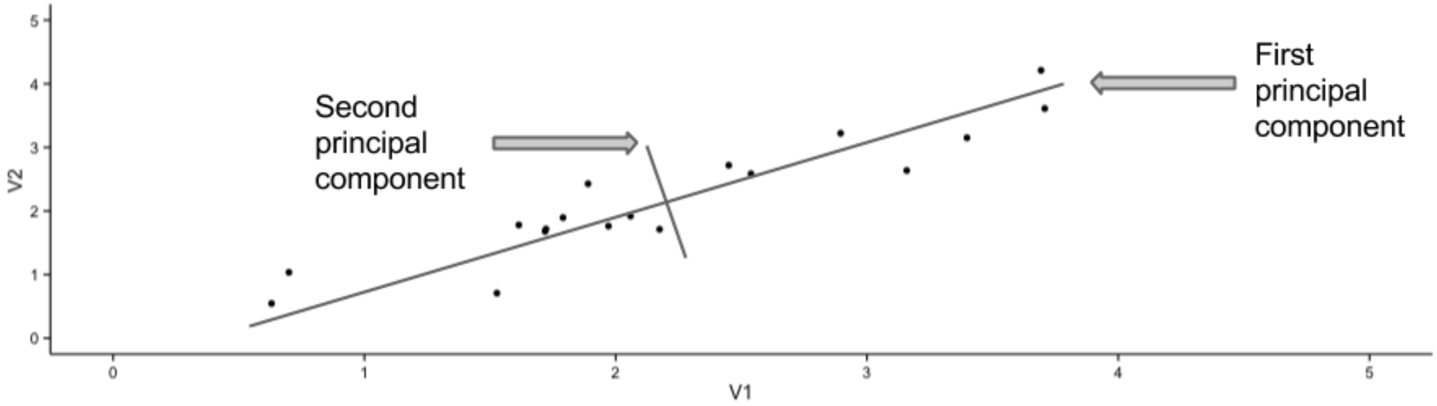
Representation Learning

How can playlist characteristics be represented?

- **Map features to latent factors**
- **Only take relevant variance into account**
- **Principal component analysis**

- **By approximating variance in playlists we learn a representation of playlist characteristics**

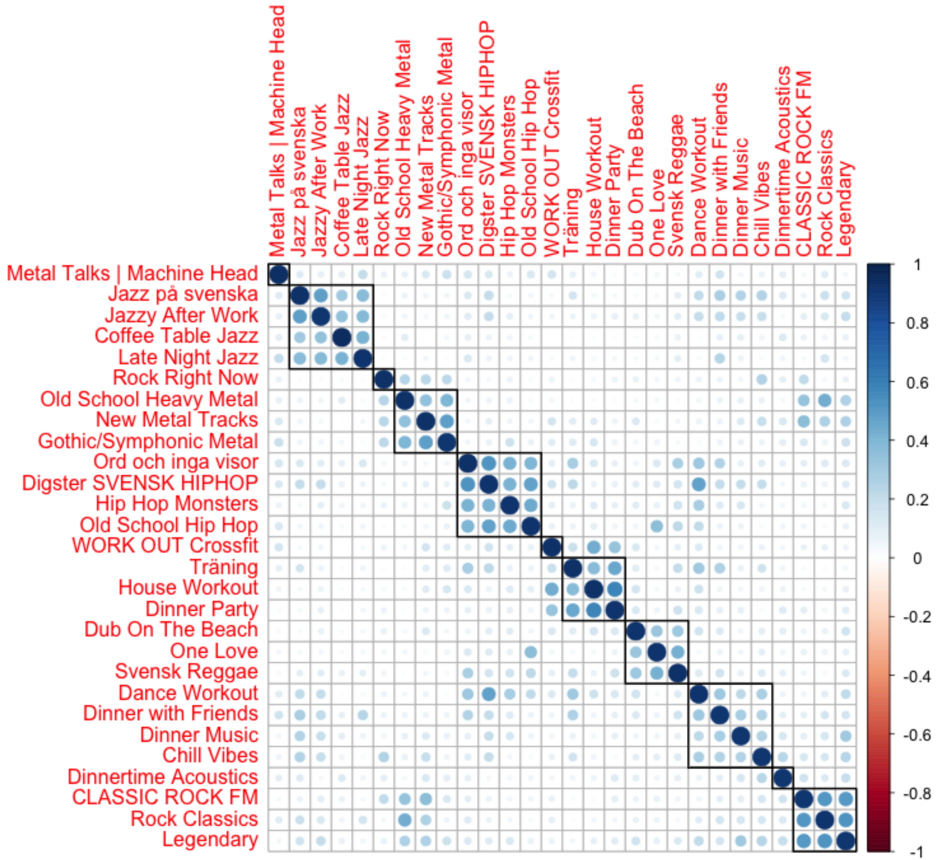
Principal Component Analysis



Playlist Characteristics

Does it work?

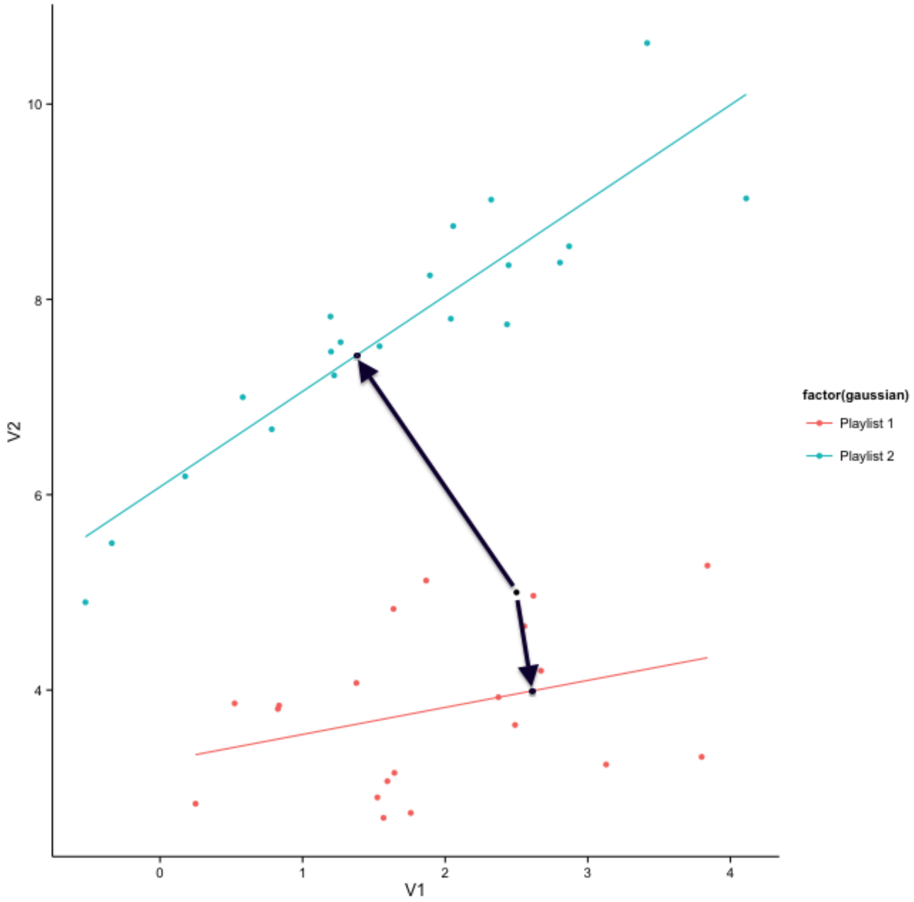
- Playlist comparison and clustering works fine
- Based on comparison of latent spaces



Candidate Song Selection

How to relate playlist variance to songs?

- Subspace method
- Relative change in magnitude under projection



Candidate Song Selection

How are songs selected?

- **Songs are projected into the principal component space of a playlist**
- **Ranking is according to the fit of songs to the principal component space**
- **Songs that have their “*feature mass*” in the directions of variance are ranked higher**

Theoretical shortcomings

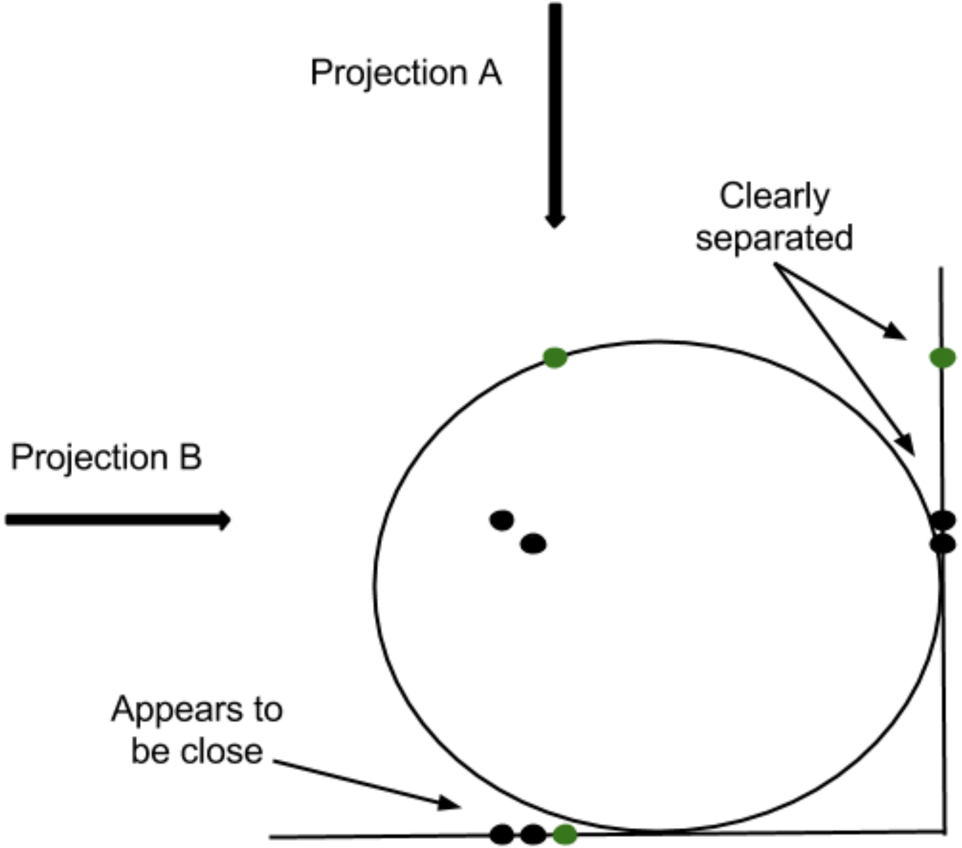
What are the weaknesses of the model?

- PCA is based on approximation of covariance in data
- Covariance is normalizing
- The model takes direction, but not location into account.

Approximate nearest neighbours

What if we could pre-filter based on location?

- **Approximate nearest neighbours can pre-filter based on location in sublinear time.**

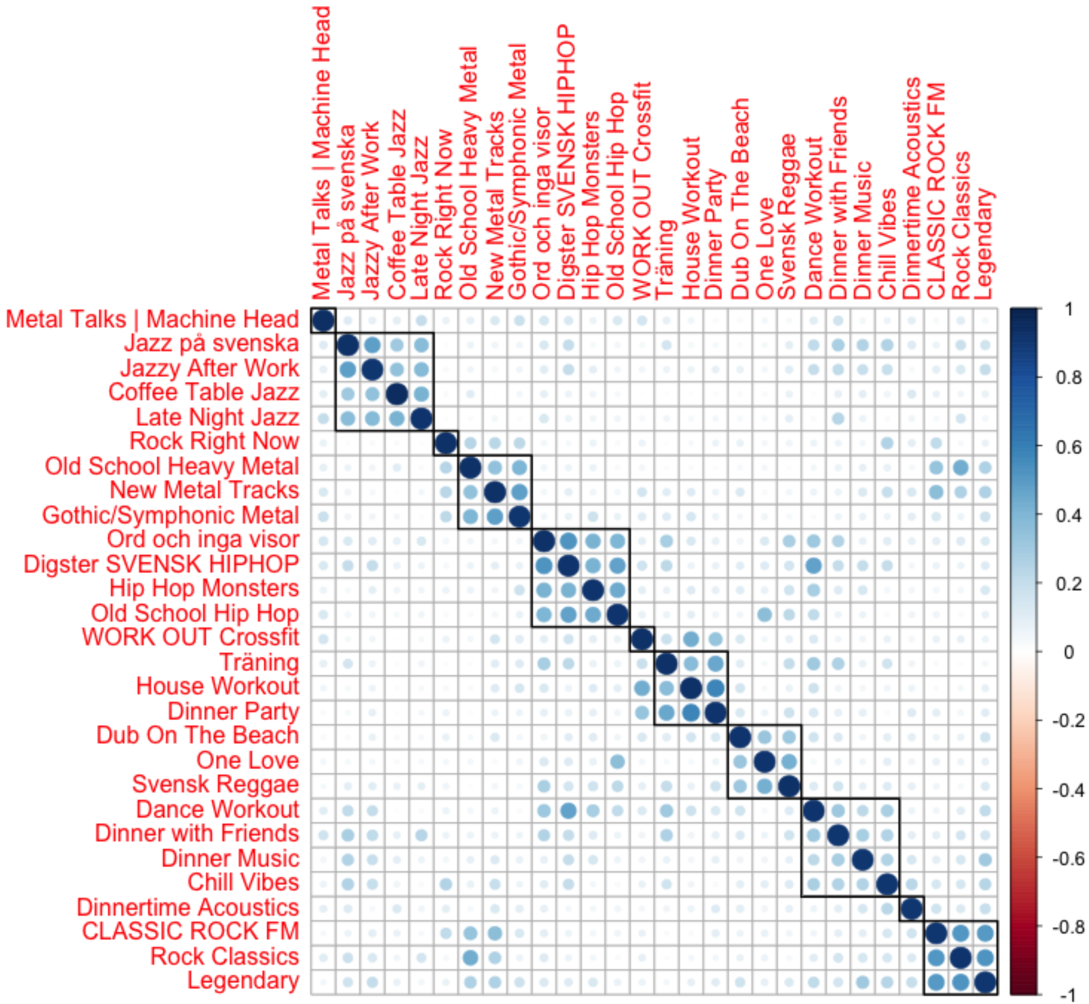


Evaluation

How playlist candidate songs be evaluated?

- **Why is evaluation difficult?**
  - Tautology
- **What is an appropriate metric?**
  - Precision
- **What is an appropriate baseline?**
  - Taking full variance of data into account

Evaluation


Metal Talks | Machine Head
Jazz på svenska
Jazzy After Work
Coffee Table Jazz
Late Night Jazz
Rock Right Now
Old School Heavy Metal
New Metal Tracks
Gothic/Symphonic Metal
Ord och inga visor
Digster SVENSK HIPHOP
Hip Hop Monsters
Old School Hip Hop
WORK OUT Crossfit
Träning
House Workout
Dinner Party
Dub On The Beach
One Love
Svensk Reggae
Dance Workout
Dinner with Friends
Dinner Music
Chill Vibes
Dinnertime Acoustics
CLASSIC ROCK FM
Rock Classics
Legendary
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

Evaluation

How playlist candidate songs be evaluated?

- **Why is evaluation difficult?**
  - Tautology
- **What is an appropriate metric?**
  - Precision
- **What is an appropriate baseline?**
  - Taking full variance of data into account

Quantitative results

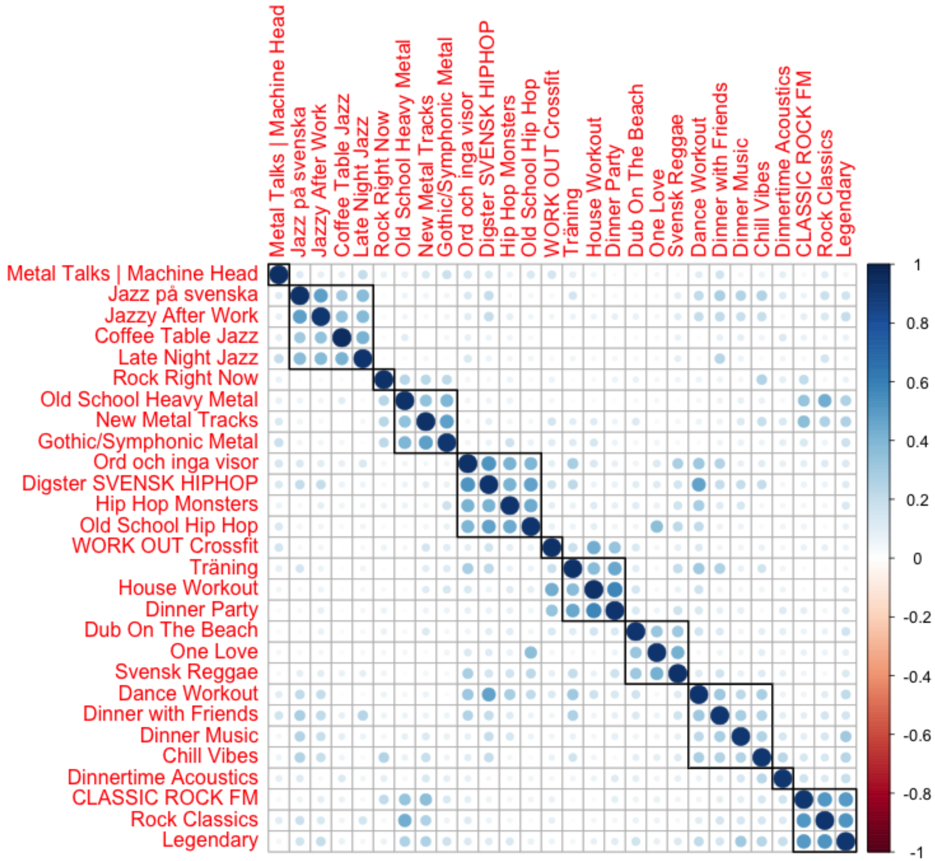
Precision scores

▶	**Model**	**Baseline**
p@10	6	3
p@20	10	6
p@30	16	10

False Positives

Let's revisit the clusters

- Where do FPs come from?
- It is reasonable that precision is overly conservative as a metric



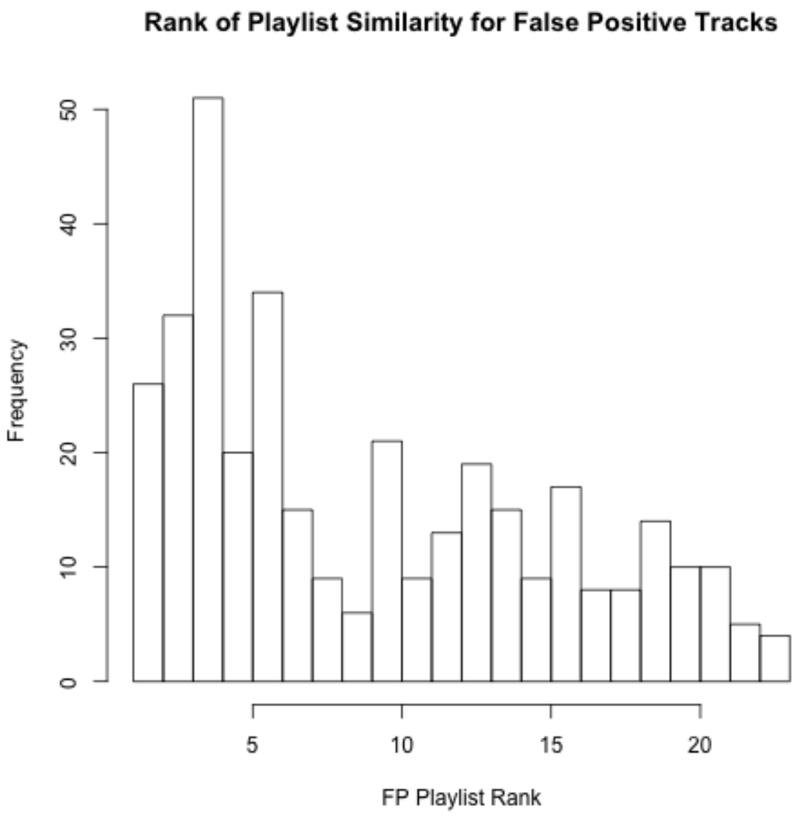
False Positives

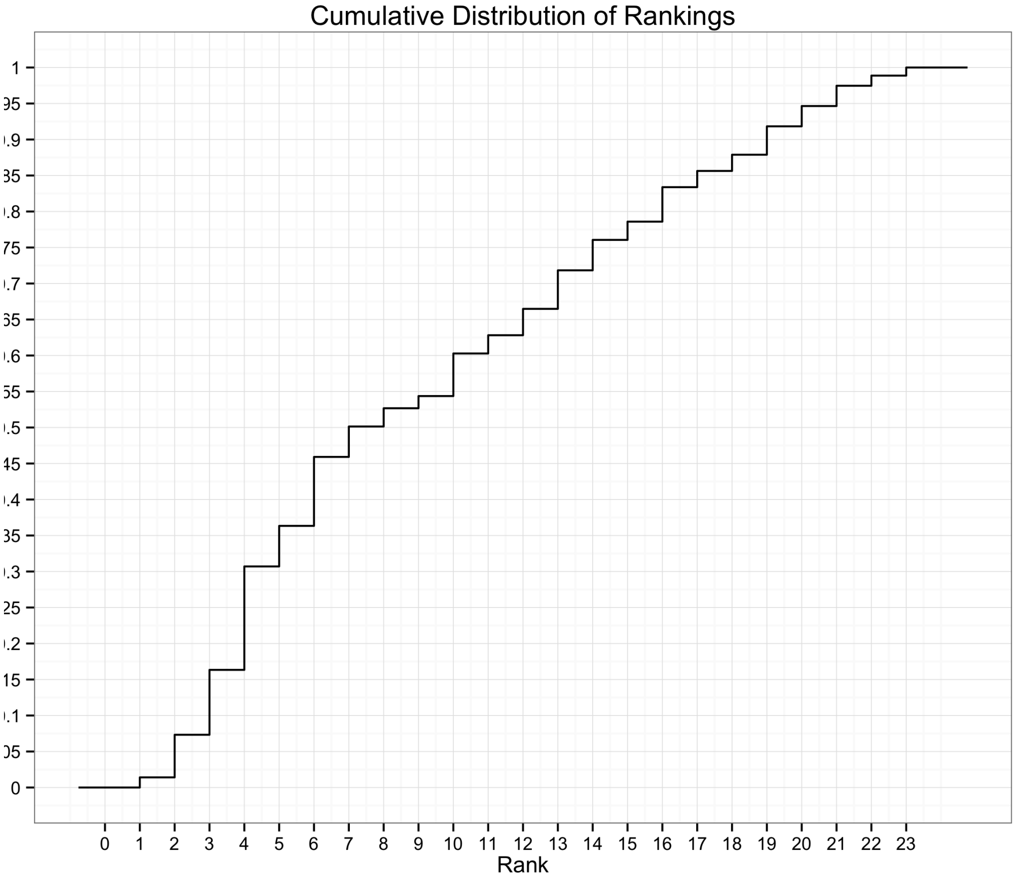
What can we see from rank distribution

- **Precision measure seems overly conservative**
- **"*True*" precision is probably higher**

- **Let us look at distribution of rankings of playlists which FPs belong to**

False Positives Rank Distributions





Qualitative Evaluations

Quantitative measures only reveal so much

- Are generated playlists listenable?
- How do they compare to curated playlists?
- Is actual precision really higher?

Qualitative Evaluations

What questions to ask listeners?

- How well songs matches playlist theme
- Number of outliers
- How well songs matches playlist theme with outliers removed.

Qualitative Evaluations

Difficulties with qualitative evaluations

- Users are biased
- How to handle user bias?

Qualitative Results

Moment of truth

- Two out of three playlist performed equally good or better than the reference playlists
- One performed significantly worse
- But that playlist was the 2nd worse in the entire experiment and had an uplift in precision over 100%

Scalability

Does it scale?

- Time complexity of creating a projection matrix is $O(D^3 + ND^2)$
- Should take less than 1 second on modern computer
- To do candidate song selection $O(nD^2)$
- Candidate song selection is parallelizable

Conclusions

What have we learned?

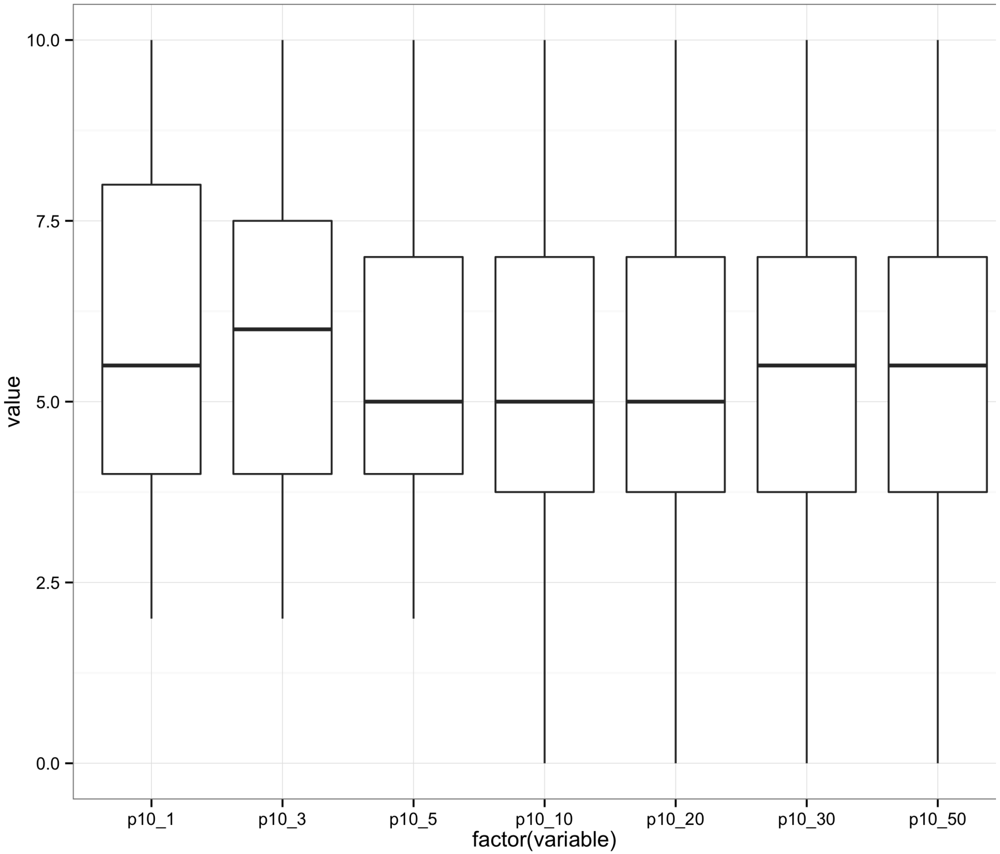
- Approximating relevant variance is shown more effective than using full variance as a baseline
- Promising qualitative results
- Performance is likely to increase with better features and more data

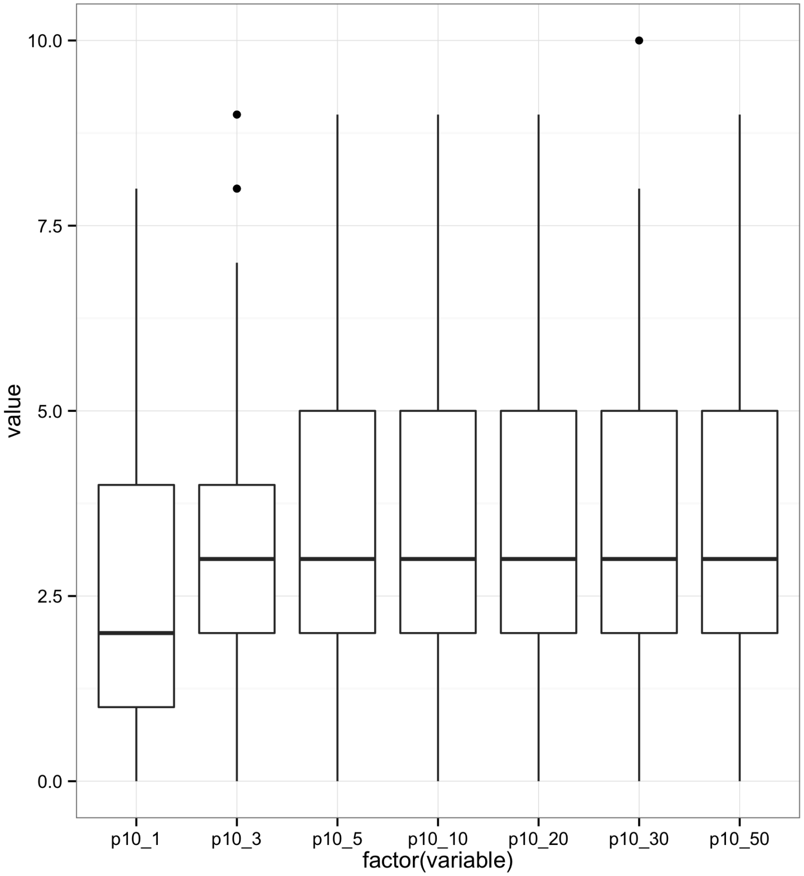
Questions?

Thank you for listening

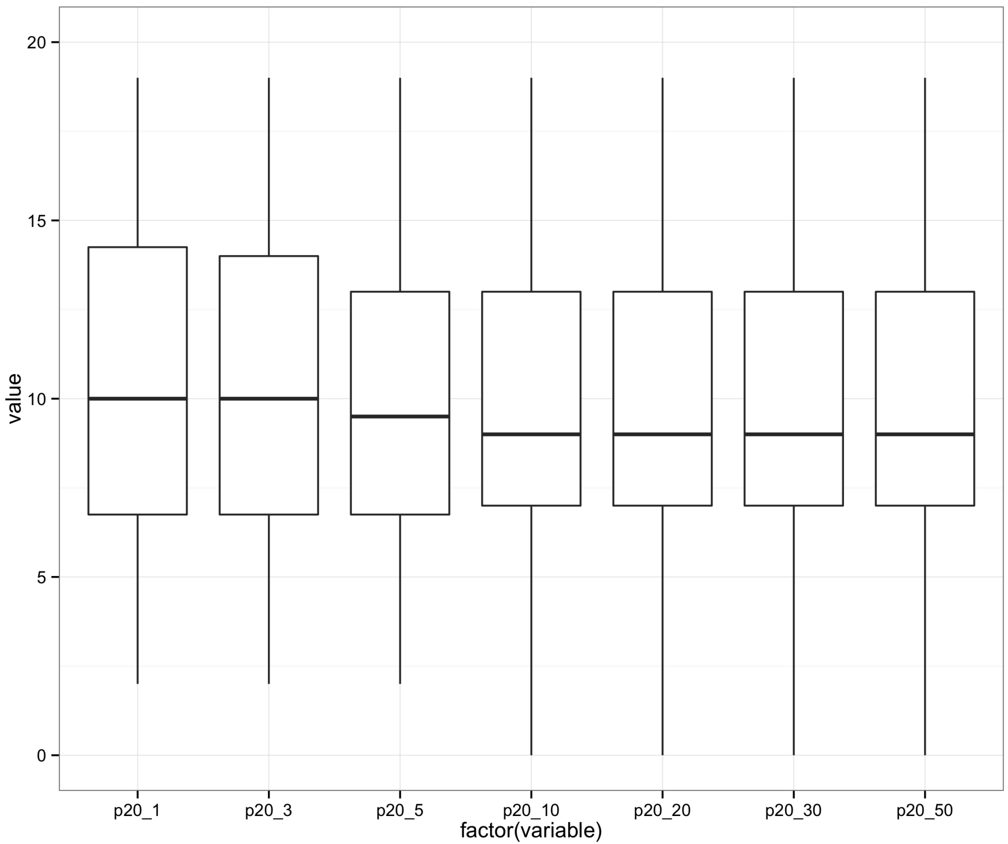
Special thanks to Spotify, Carl Henrik Ek, Boxun Zhang, Anders Petterson and Matteo Poletti

Results - p@10

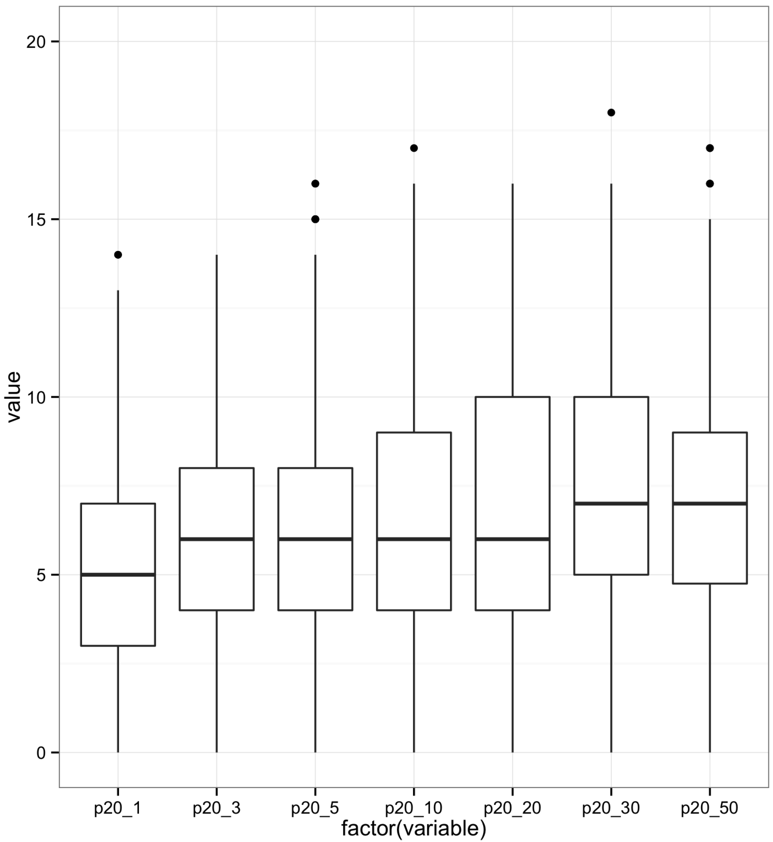

value
10.0
7.5
5.0
2.5
0.0
p10_1
p10_3
p10_5
p10_10
p10_20
p10_30
p10_50
factor(variable)


value
10.0
7.5
5.0
2.5
0.0
p10_1
p10_3
p10_5
p10_10
p10_20
p10_30
p10_50
factor(variable)

Results - p@20

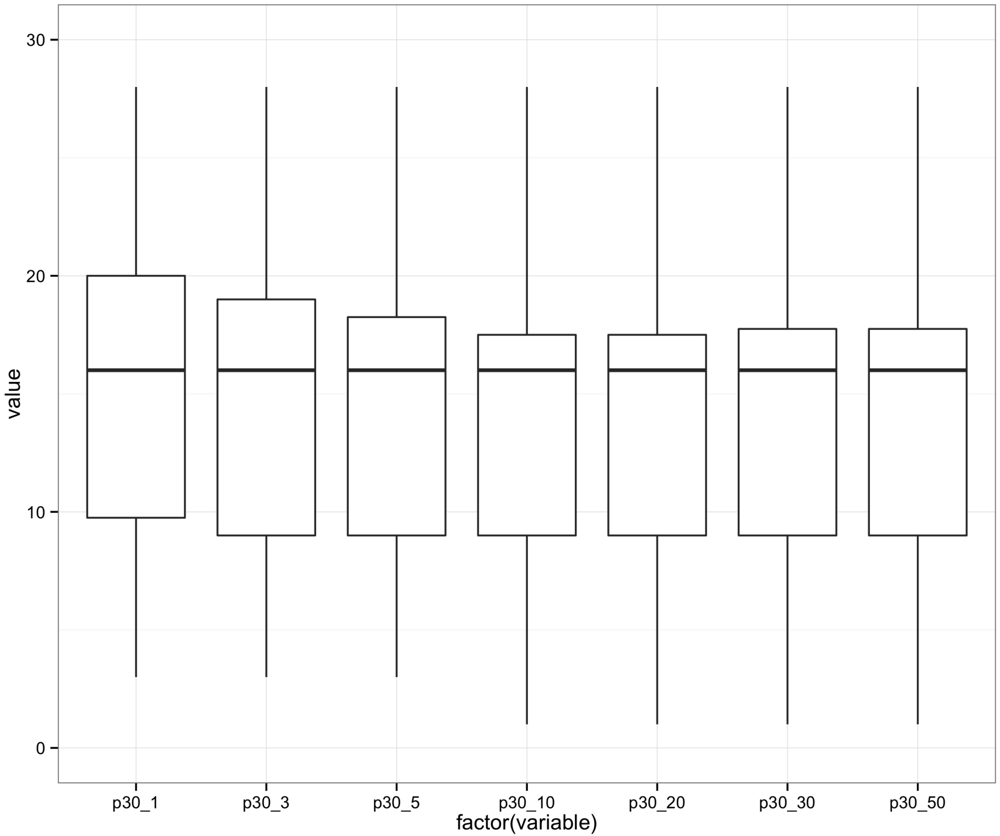


20
15
10
5
0
value
p20_1
p20_3
p20_5
p20_10
p20_20
p20_30
p20_50
factor(variable)

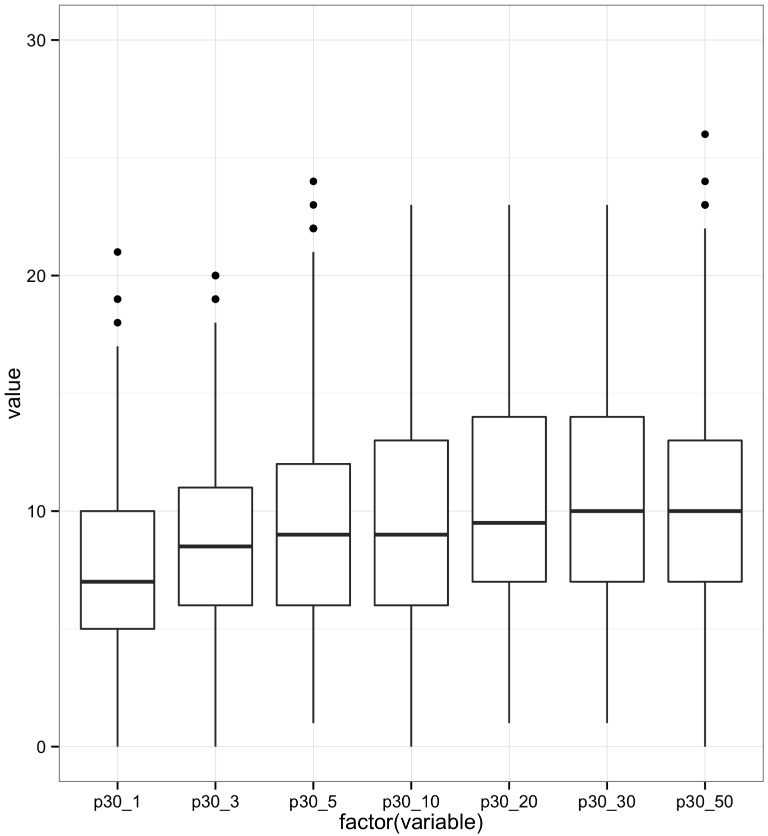


20
15
10
5
0
value
p20_1
p20_3
p20_5
p20_10
p20_20
p20_30
p20_50
factor(variable)

Results - p@30

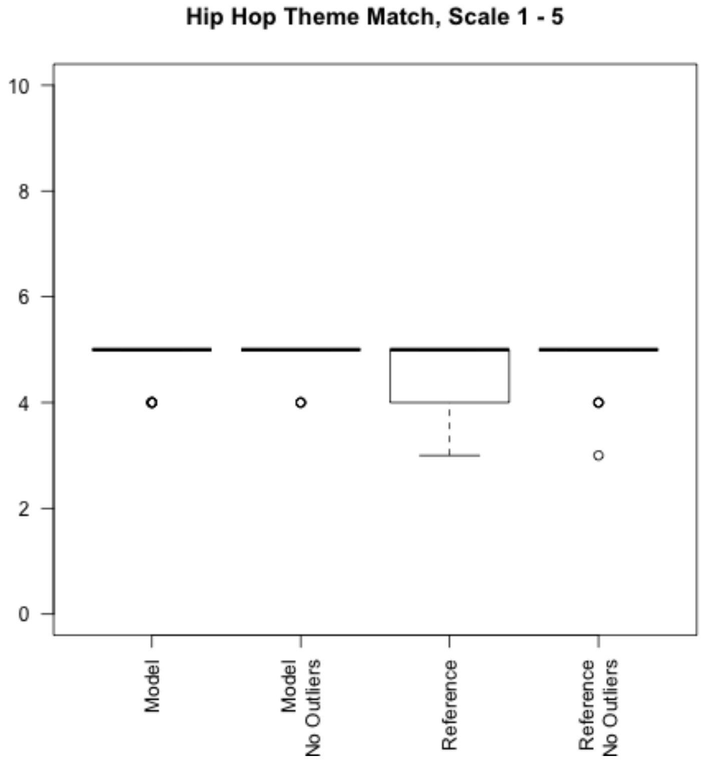


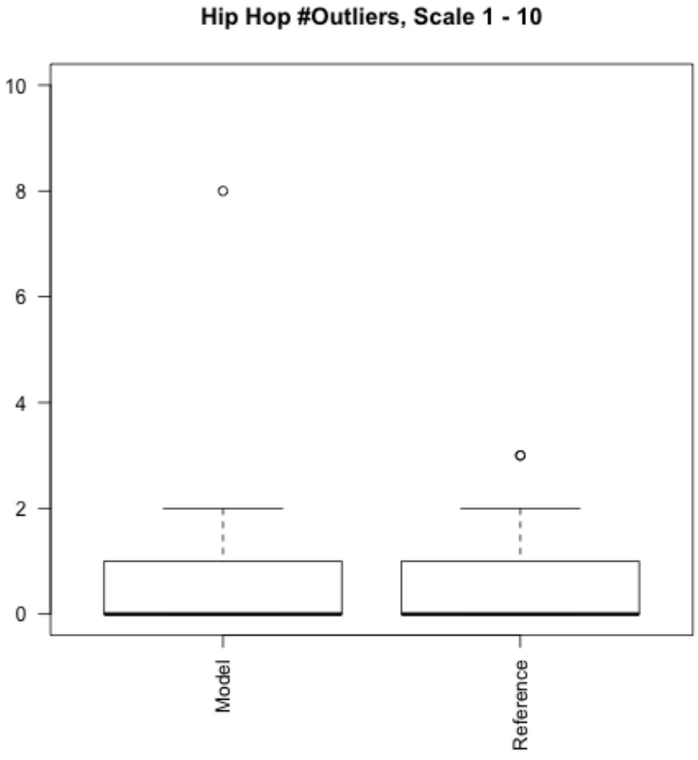
30
20
10
0
value
p30_1
p30_3
p30_5
p30_10
p30_20
p30_30
p30_50
factor(variable)



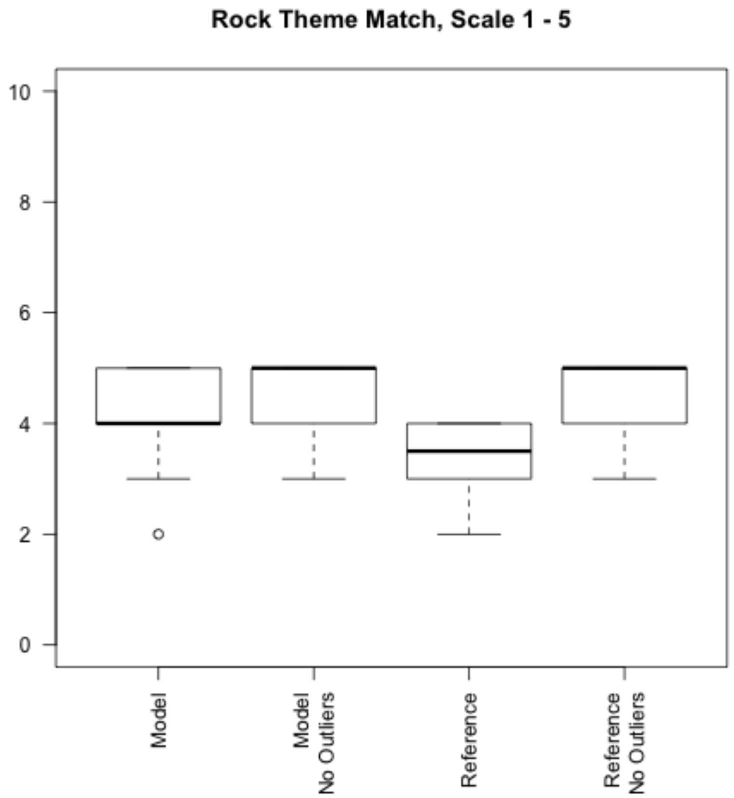
30
20
10
0
value
p30_1
p30_3
p30_5
p30_10
p30_20
p30_30
p30_50
factor(variable)

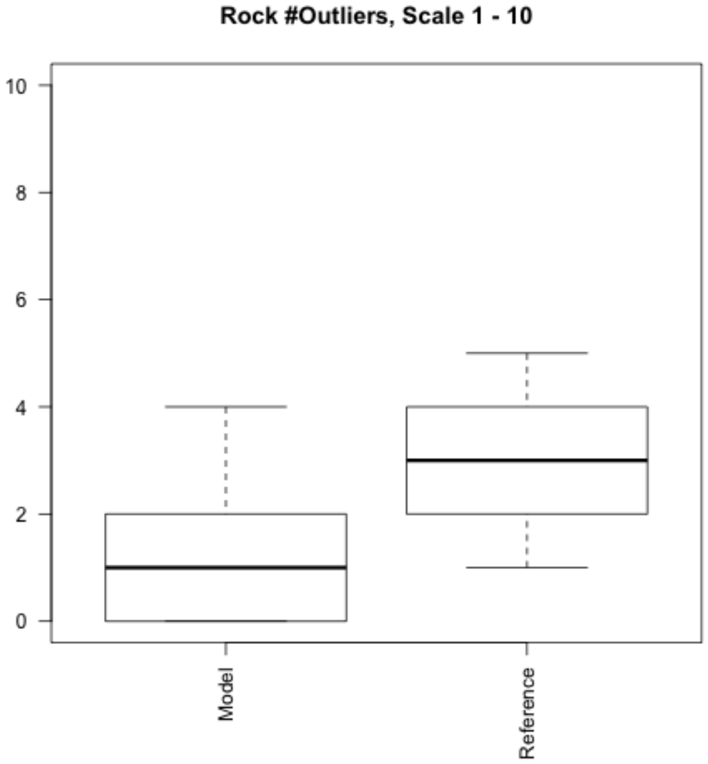
Qualitative Results Hip Hop





Qualitative Results Rock





Qualitative Results Metal