



# Analysis of music tagging and listening patterns: Do tags really function as retrieval aids?

Jared Lorince<sup>1</sup>, Kenny Joseph<sup>2</sup>, & Peter M. Todd<sup>1</sup>

<sup>1</sup>Adaptive Behavior & Cognition Lab, Indiana University

<sup>2</sup>Computation, Organization, and Society Program, Carnegie Mellon University

# In a nutshell:

**The problem:** It is assumed tags are used as cues for one's own future retrieval, but there is no behavioral evidence for this.

**The research question:** Is there evidence that tagging content increases the probability of future interaction with that content?

**The findings:** Tagging seems to lead to *small* increases in future content interaction, and only for some tags: Future retrieval does *not* appear to be the primary motivation for tagging.



Dethklok

[Tracks](#) [Albums](#) [Pictures](#) [Videos](#) [Events](#) [More...](#)

## Tags

00s 100 times bigger than the beatles adult swim american awesome black metal **blacker than the blackest black times infinity** brendon small

brutal brutal death metal brutal heavy metal brutal metal cartoon **comedy** comedy metal death **death metal** deja vu psycho metal

deth metal dethklok epic everything will be metal fucking awesome funny heavy metal in control look behind you **melodic death metal**

melodic metal **metal** metalocalypse metalocaylpse music that makes me want to beat children music that makes your head explode of amazement

nathan explosion parody pet cemetery planet x power metal psycho metal psycho metal 4 life psycho metal tag radio

psycho metal this is psycho psychometal psychometal4life real coffee from the hills of columbia rock sacrifice psycho metal soundtrack stfu

swisgar swiself taller than a tree technical death metal the most metal thing i have ever heard in my whole life this is psycho thrash metal

toki wartooth not a bumblebee true metal usa virtual band william murderface murderface murderface



Dethklok

Tags

00s

100 times bigger than the beatles

adult swim

brutal

brutal death metal

brutal heavy metal

brutal metal

can

deth metal

dethklok

epic

everything will be metal

fucking awesome

funny

metal

melodic metal

metalocalypse

metalocaylpse

music that makes me want to beat children

nathan explosion

parody

pet cemetery

planet x

power metal

psycho metal

psycho metal 4 life

psycho metal 4 life

psycho metal this is psycho

psychometal

psychometal4life

real coffee from the hills of columbia

rock

sacrafice psycho metal

soundtrack

swisgar swiself taller than a tree

technical death metal

the most metal thing i have ever heard in my whole life

this is psycho

thrash metal

toki wartooth not a bumblebee

true metal

usa

virtual band

william murderface murderface murderface

00s 90s acoustic alternative alternative rock amazing ambient art rock atmospheric  
avant-garde awesome beautiful chill chillout downtempo dream pop dreamy easy listening electronic  
electronica emo emotional epic ethereal experimental experimental rock favorites favourites folk iceland  
icelandic indie indie pop indie rock instrumental love male vocalists melancholic melancholy  
mellow minimalist nordic pop post rock post-rock progressive progressive rock  
psychedelic relax relaxing rock scandinavian shoegaze sigur ros singer-songwriter sleep slow-motion rock space  
rock trip-hop want to see live



Dethklok

Tags

00s

100 times bigger than the beatles

adult swim

brutal

brutal death metal

brutal heavy metal

brutal metal

car

deth metal

dethklok

epic

everything will be metal

funny

melodic

metal

metalcore

metalcore

metalcore

metalcore

nathan explosion

parody

pet cemetery

planet x

power metal

psycho metal

psycho metal 4 life

psycho metal 4 life

psycho metal this is psycho

psychometal

psychometal4life

real coffee from the hills of columbia

rock

sacrafice psycho metal

soundtrack

swisgar swiself taller than a tree

technical death metal

the most metal thing i have ever heard in my whole life

this is psycho

thrash metal

toki wartooth not a bumblebee

true metal

usa

virtual band

william murderface murderface murderface

00s 90s acoustic alternative alternative rock amazing ambient art rock atmospheric

avant-garde awesome beautiful chill chillout downtempo dream pop dreamy easy listening electronic

electronica emo emotional epic ethereal experimental experimental rock favorites favourites folk iceland

icelandic indie indie pop indie rock instrumental love male vocalists melancholic melancholy

mellow minimalist nordic pop post rock post-rock progressive rock

psychedelic relax relaxing rock scandinavian shoegaze slow rock space

sleep slow-motion rock space

# Collaborative Tagging



Dethklok

Tags

00s

100 times bigger than the beatles

adult swim

brutal

brutal death metal

brutal heavy metal

brutal metal

ca

deth metal

dethklok

epic

ever

ing will be met

lucking awesome

metal

metalocalyp

metalocalypse

metalocalypse

metalocalypse

metalocalypse

nathan explosion

parody

pet cemetery

planet x

power metal

psycho metal

psycho metal 4 life

psycho metal 4

psycho metal this is psycho

psychometal

psychometal4life

real coffee from the hills of columbia

rock

sacrafice psycho metal

soundtrack

swisgar swiself taller than a tree

technical death metal

the most metal thing i have ever heard in my whole life

this is psycho

thrash metal

toki wartooth not a bumblebee

true metal

usa

virtual band

william murderface murderface murderface

# Folksonomy

00s 90s acoustic alternative alternative rock amazing ambient art rock atmospheric  
avant-garde awesome beautiful chill chillout downtempo dream pop dreamy easy listening electronic  
electronica emo emotional epic ethereal experimental experimental rock favorites favourites folk iceland  
icelandic indie indie pop indie rock instrumental love male vocalists melancholic melancholy  
mellow minimalist nordic pop post rock post-rock progressive progressive rock  
psychedelic relax relaxing rock scandinavian shoegaze singer-songwriter sleep slow-motion rock space

00s 90s acoustic alternative alternative rock amazing ambient art rock atmospheric  
avant-garde awesome beautiful chill chillout downtempo dream pop dreamy easy listening electronic  
electronica emo emotional epic ethereal experimental experimental rock favorites favourites folk iceland  
icelandic indie indie pop indie rock instrumental love male vocalists melancholic melancholy  
mellow minimalist nordic pop post rock post-rock progressive progressive rock  
psychedelic relax relaxing rock scandinavian shoegaze singer-songwriter sleep slow-motion rock space

last.fm



flickr



LibraryThing  
What's on your bookshelf?



- **Tagging expertise**  
Yeung, Noll, Gibbins, Meinel, & Shadbolt (2009,2011)
- **Modeling of tagging behavior**
  - **Mathematical**  
Cattuto, Loreto, & Pietronero (2007).
  - **Multi-agent**  
Lorince & Todd (2013)
- **Consensus in collaborative tagging**  
Golder & Huberman (2006), Robu, Halpin, & Shepherd (2009)
- **Social networks and tagging**  
Schifanella, Barrat, Cattuto, Markines, & Menczer (2010)
- **“Supertagger” behavior**  
Lorince, Zorowitz, Murdock, & Todd (2014)



- Tagging expertise  
Yeung, Noll, Gibbins, Meinel, & Shadbolt (2009,2011)
- Modeling of tagging behavior
  - Mathematical  
Cattuto, Loreto, & Pietronero (2007).
  - Multi-agent  
Lorince & Todd (2013)
- Consensus in collaborative tagging  
Golder & Huberman (2006), Robu, Halpin, & Shepherd (2009)
- Social networks and tagging  
Schifanella, Barrat, Cattuto, Markines, & Menczer (2010)
- “Supertagger” behavior  
Lorince, Zorowitz, Murdock, & Todd (2014)

We know plenty about **how**  
people tag, but **why** they tag is  
a much trickier question

“Folksonomy is the result of personal free tagging of information and objects (anything with a URL) for one's own retrieval.”

Thomas Vander Wal

<http://www.vanderwal.net/random/entrysel.php?blog=1750>

“Folksonomy is the result of personal free tagging of information and objects (anything with a URL) ***for one's own retrieval.***”

Thomas Vander Wal

<http://www.vanderwal.net/random/entrysel.php?blog=1750>

# The retrieval aid hypothesis

“Folksonomy is the result of personal free tagging of information and objects (anything with a URL) *for one's own retrieval.*”

Thomas Vander Wal

<http://www.vanderwal.net/random/entrysel.php?blog=1750>

# There are alternatives...

Resource sharing vs. personal information management (Heckner, et al. 2009)

Opinion expression, performance, and activism (Zollers, 2007)

Self presentation, play/competition (Marlow, et al., 2006)

Sociality and function (Ames & Naaman, 2007)

Categorizers and describers (Körner & colleagues, 2010, 2011)

# There are alternatives...

Resource sharing vs. personal information management (Heckner, et al. 2009)

Opinion expression, performance, and activism (Zollers, 2007)

Self presentation, play/competition (Marlow, et al., 2006)

Sociality and function (Ames & Naaman, 2007)

Categorizers and describers (Körner & colleagues, 2010, 2011)

...but they all face a serious obstacle

We must infer *why* people tag from *how*  
they tag

We must infer *why* people tag from *how*  
they tag

There is simply no data available on how  
people *use* tags once they've applied them.



# Dataset

last.fm

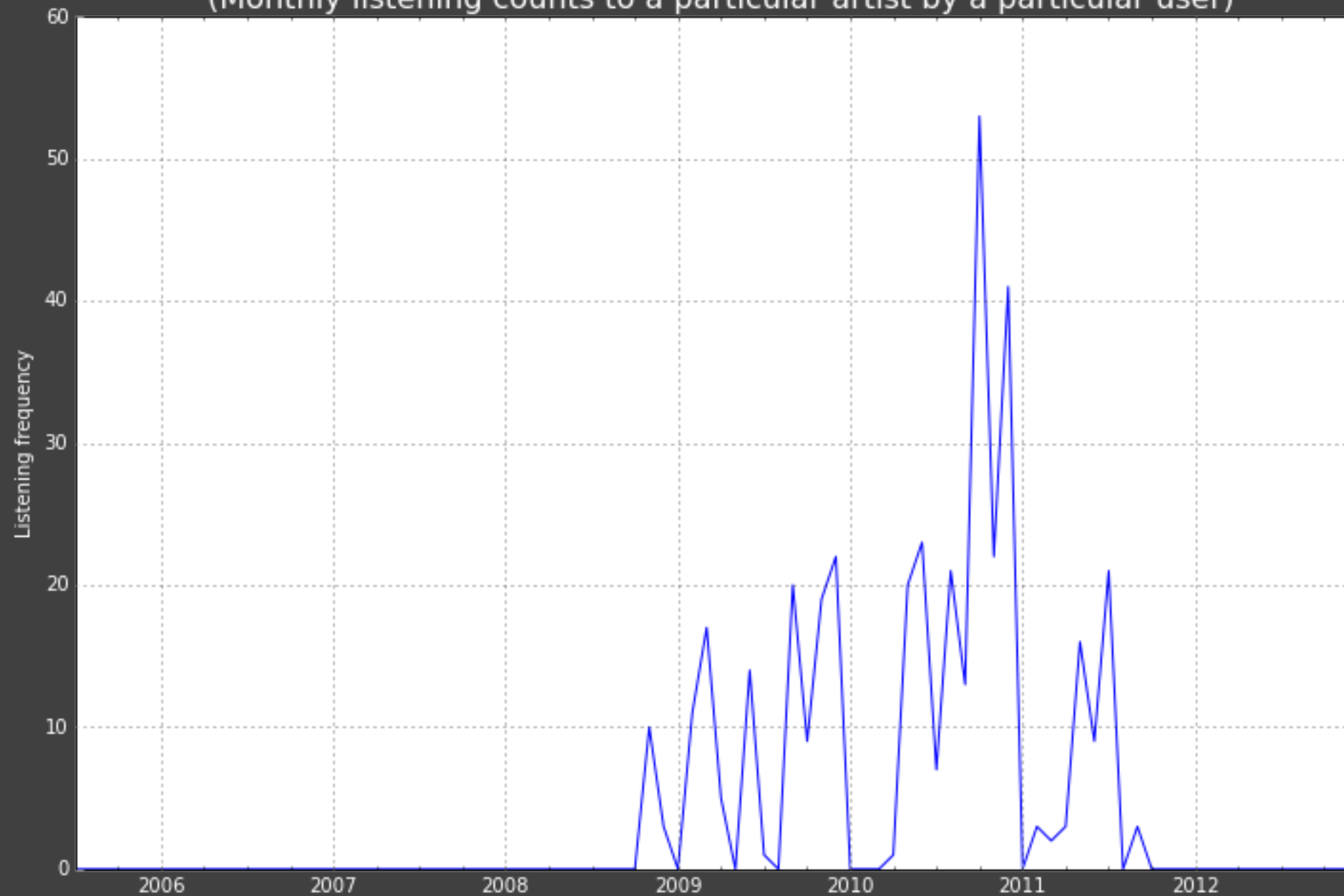
Total users	1,884,597
Total friendship relations	24,320,919
Total group memberships	5,458,935
Total annotations	50,372,893
Users with $\geq 1$ annotation	521,780
Total unique tags	1,029,091
Unique items tagged	4,477,591
Unique artists tagged	1,049,263
Users with listening data recorded	105,425
Total listens	2,089,473,214
Unique songs listened	45,732,350
Unique artists listened	4,444,119
Total loved tracks	162,788,213
Users with $\geq 1$ loved track	1,335,859
Total banned tracks	23,321,347
Users with $\geq 1$ banned track	502,758

# Dataset

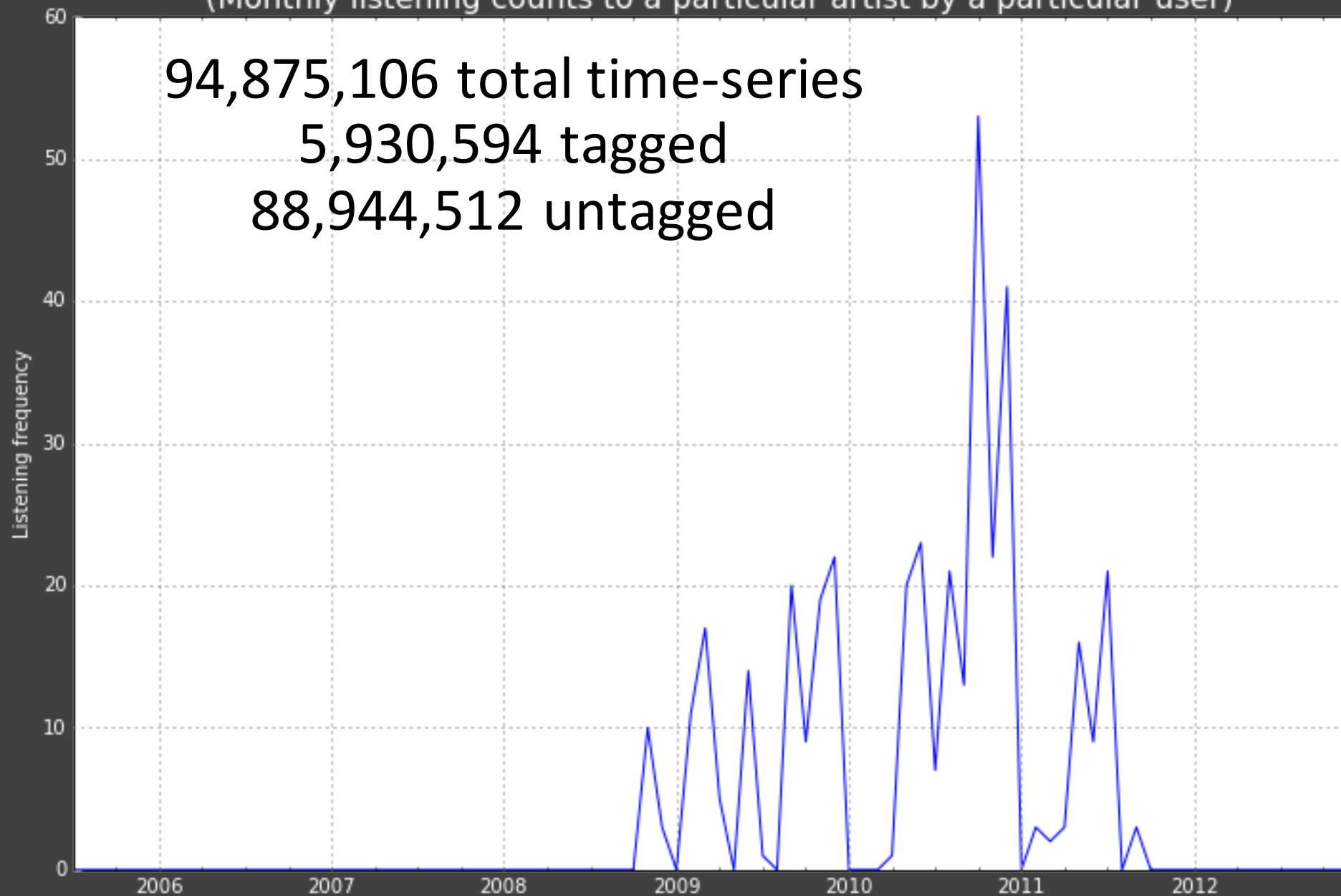
last.fm

Total users	1,884,597
Total friendship relations	24,320,919
Total group memberships	5,458,935
Total annotations	50,372,893
Users with $\geq 1$ annotation	521,780
Total unique tags	1,029,091
Unique items tagged	4,477,591
Unique artists tagged	1,049,263
Users with listening data recorded	105,425
Total listens	2,089,473,214
Unique songs listened	45,732,350
Unique artists listened	4,444,119
Total loved tracks	162,788,213
Users with $\geq 1$ loved track	1,335,859
Total banned tracks	23,321,347
Users with $\geq 1$ banned track	502,758

Example user-artist listening time series  
(Monthly listening counts to a particular artist by a particular user)



Example user-artist listening time series  
(Monthly listening counts to a particular artist by a particular user)



Our approach: Compare tagged and untagged *user-artist listening time-series*.

The hypothesis: If tags really do serve as retrieval aids, we should observe an increase in listening rates caused by application of tags.

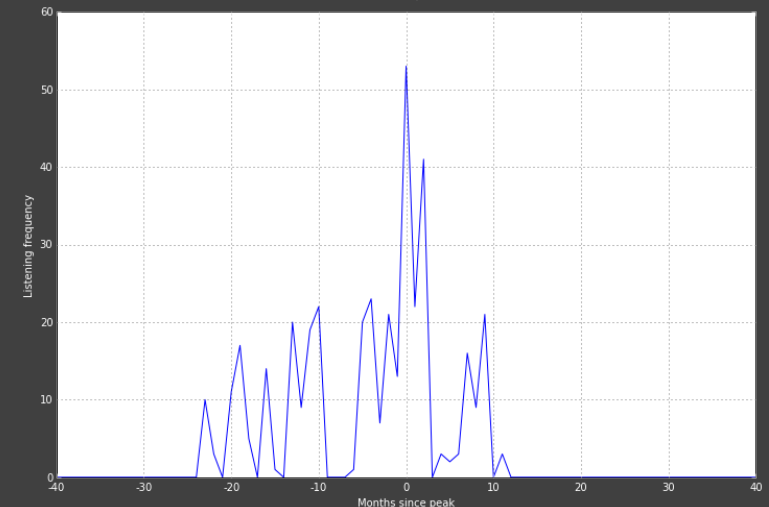
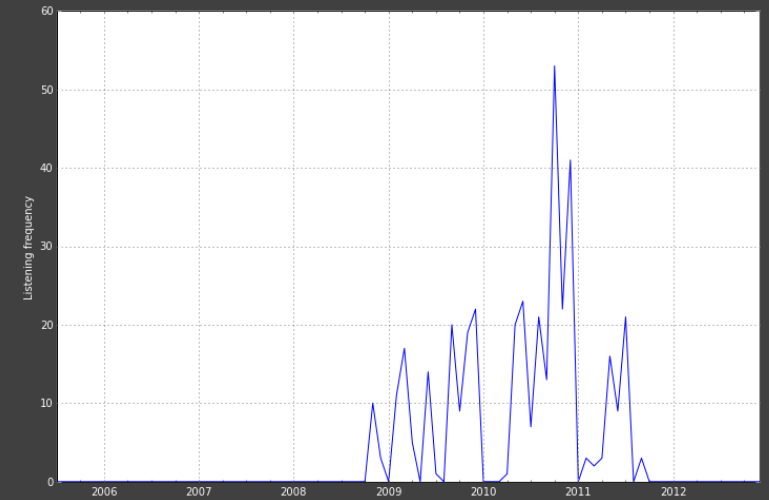
Listening rates for tagged time series are much higher when tagged, but this is not a fair comparison.

We need a *counterfactual*

# Generating a (pseudo-)counterfactual

1. Align all time series to the month of peak listening.
2. Limit analysis to a 13 month period extending from 6 months prior to the peak month to 6 months after the peak.
3. For tagged time-series, only consider those where tag was applied in the peak month

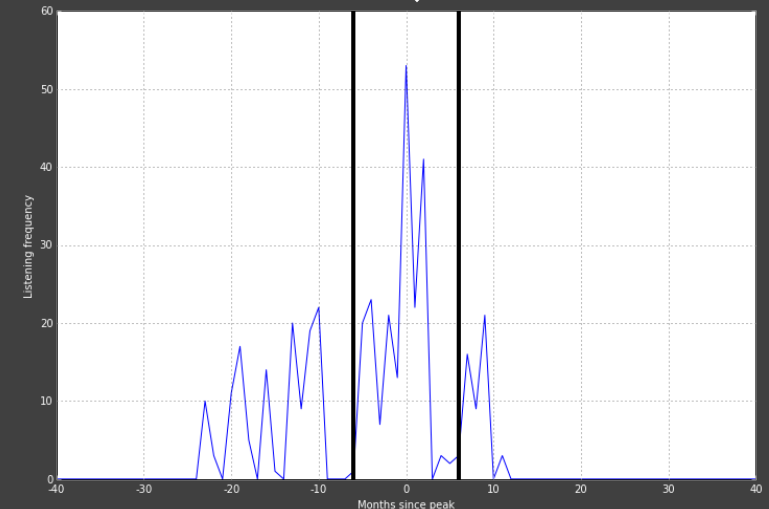
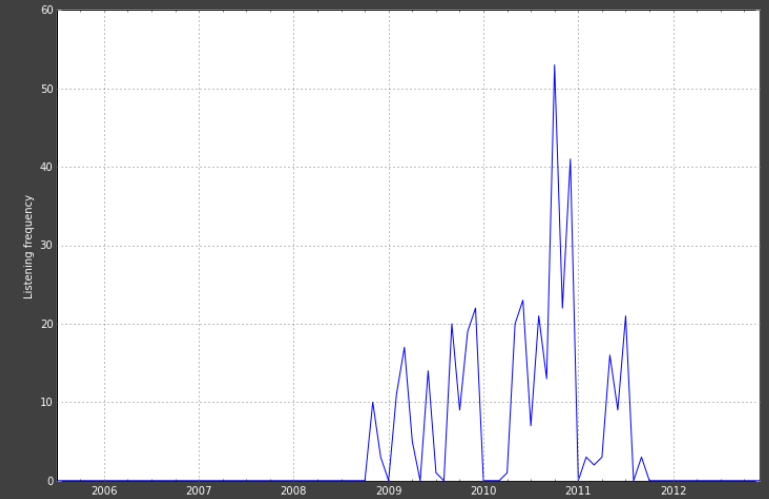
*This give us a basis to compare temporally aligned tagged and untagged time-series data.*



# Generating a (pseudo-)counterfactual

1. Align all time series to the month of peak listening.
2. Limit analysis to a 13 month period extending from 6 months prior to the peak month to 6 months after the peak.
3. For tagged time-series, only consider those where tag was applied in the peak month

*This give us a basis to compare temporally aligned tagged and untagged time-series data.*

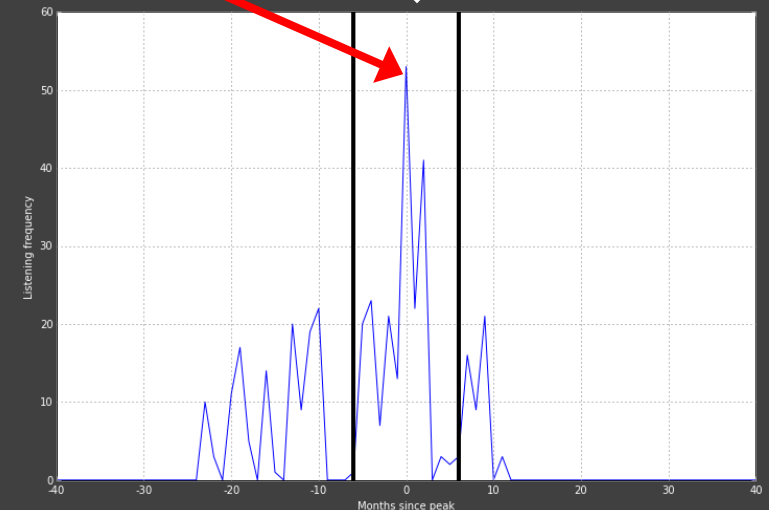
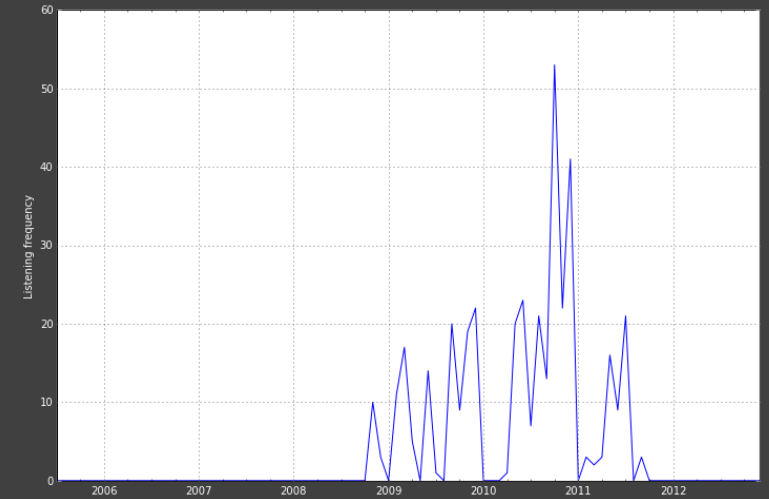




# Generating a (pseudo-)counterfactual

1. Align all time series to the month of peak listening.
2. Limit analysis to a 13 month period extending from 6 months prior to the peak month to 6 months after the peak.
3. For tagged time-series, only consider those where tag was applied in the peak month

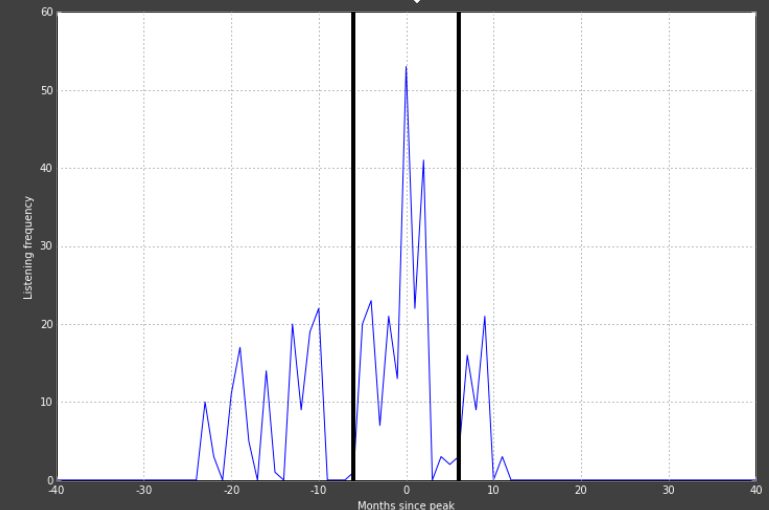
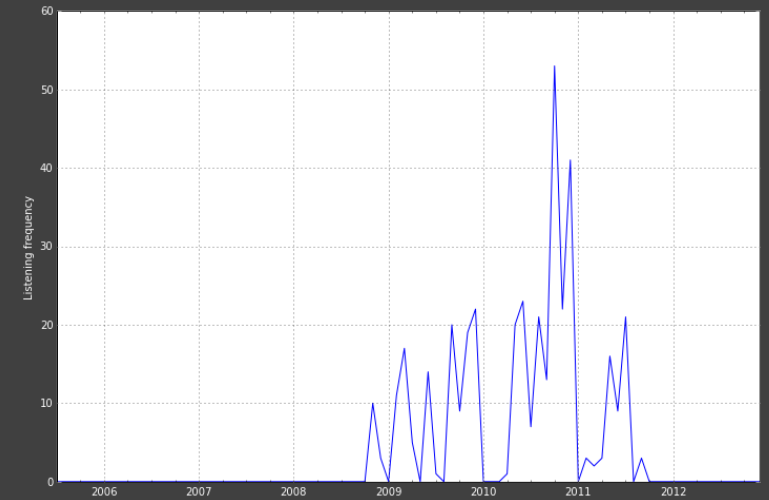
*This give us a basis to compare temporally aligned tagged and untagged time-series data.*

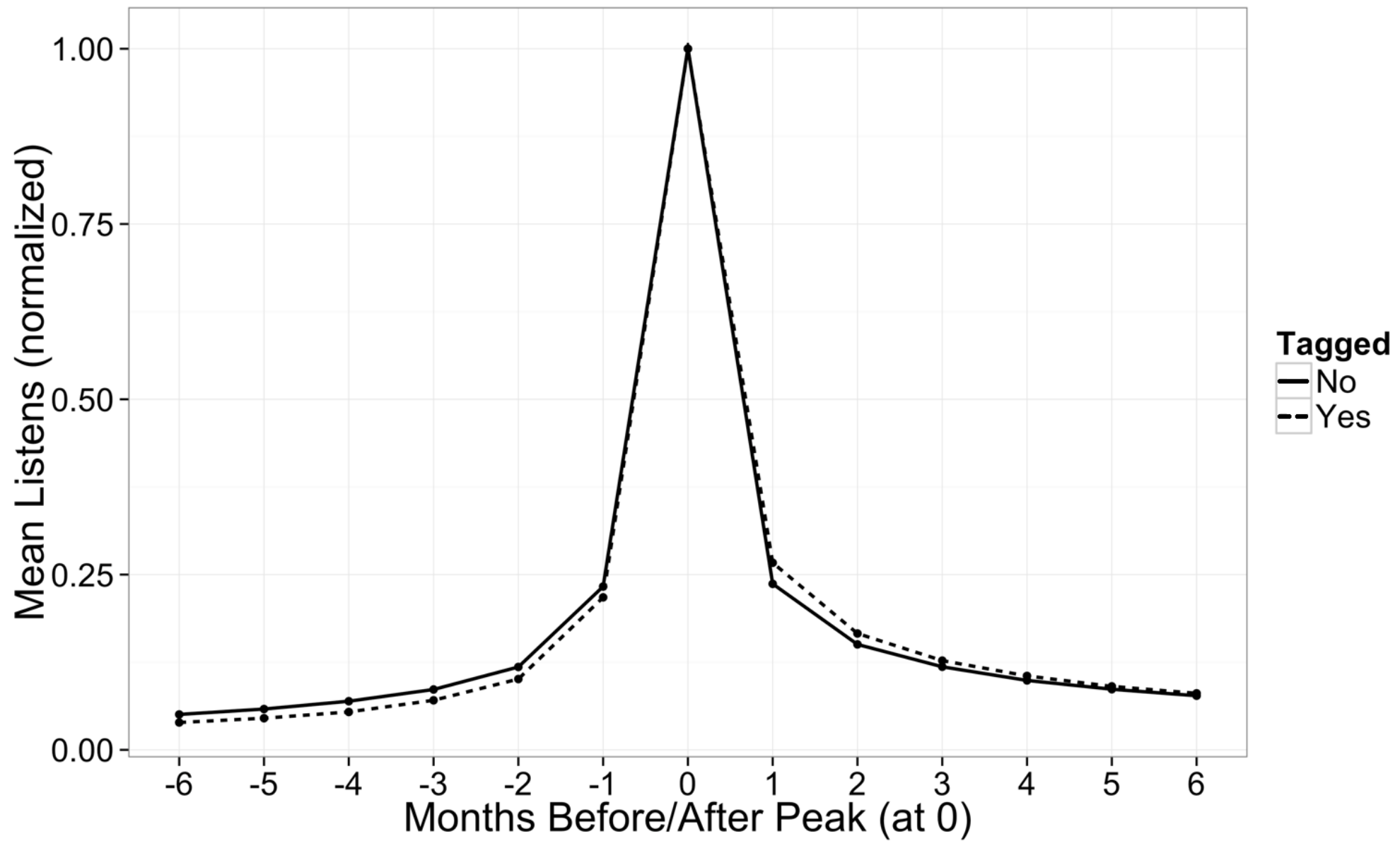


# Generating a (pseudo-)counterfactual

1. Align all time series to the month of peak listening.
2. Limit analysis to a 13 month period extending from 6 months prior to the peak month to 6 months after the peak.
3. For tagged time-series, only consider those where tag was applied in the peak month

*This give us a basis to compare temporally aligned tagged and untagged time-series data.*





# A regression model (GAM)

Pre-peak listening behavior + tagged? → post-peak listening

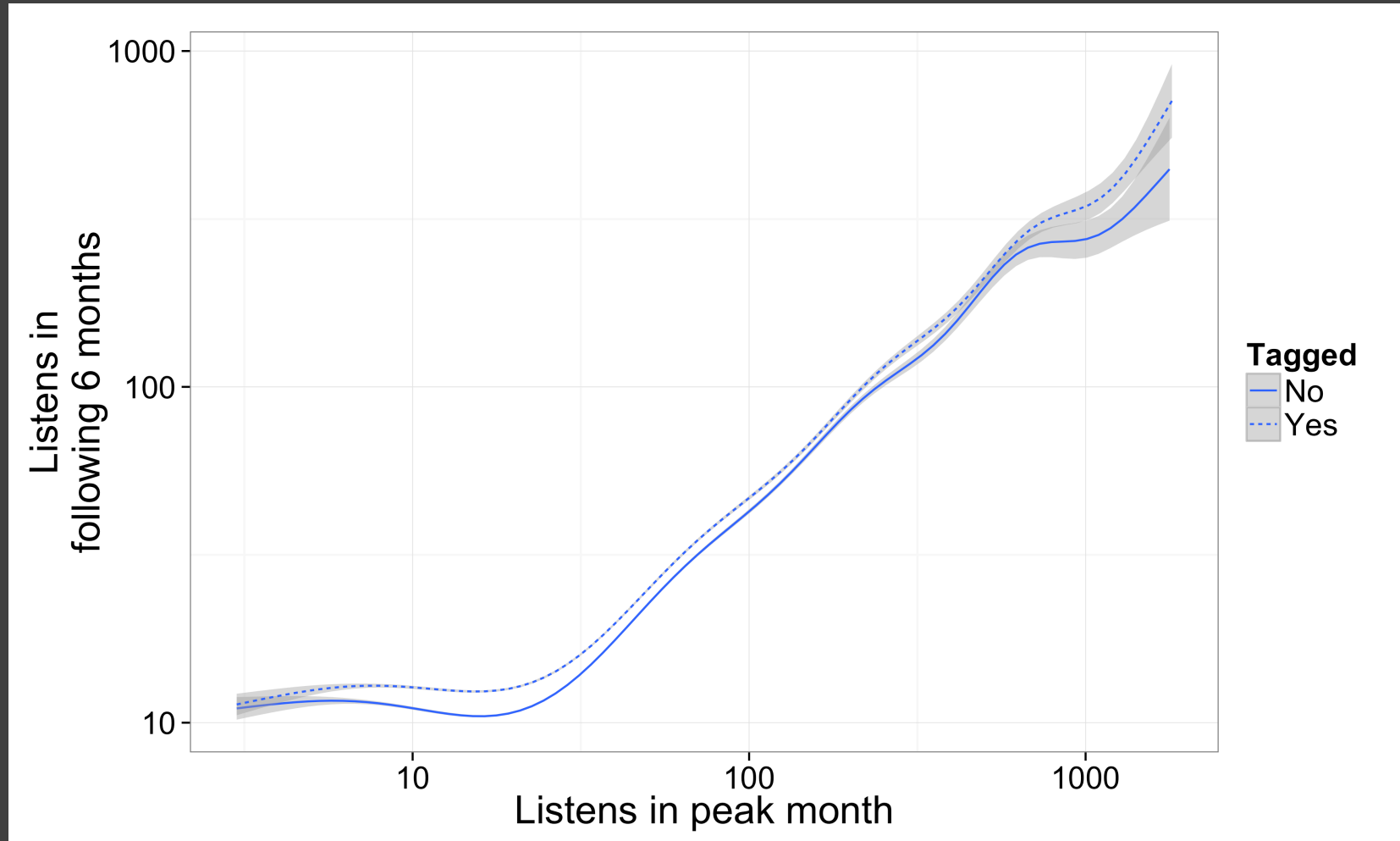
Independent variables (8 total):

Binary tagged/untagged indicator

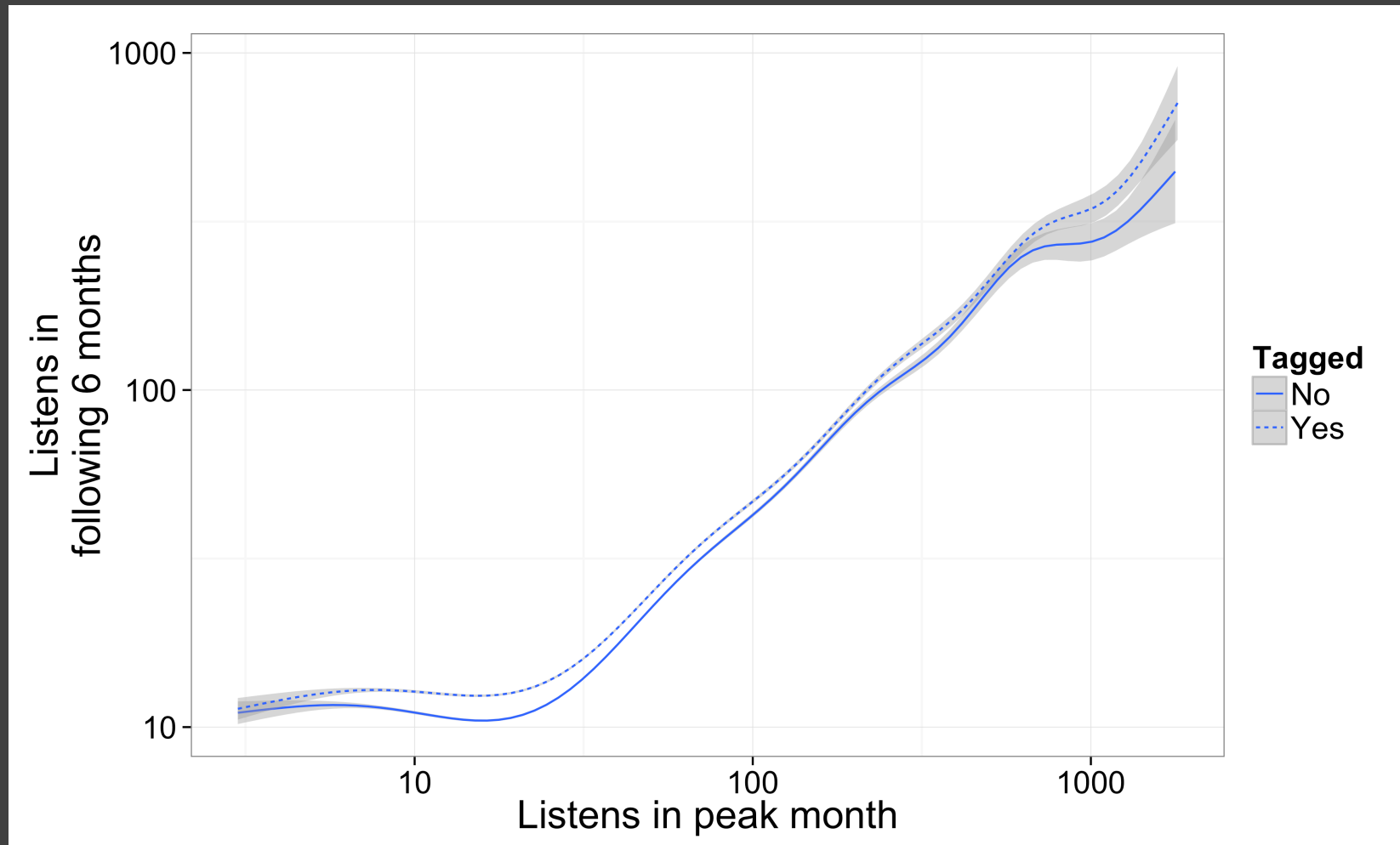
$$\{T_{peak-6}, T_{peak-5}, \dots, T_{peak}\}$$

Dependent variable:  $\sum_{i=1}^6 T_{peak+i}$

# A regression model (GAM)



# A regression model (GAM)



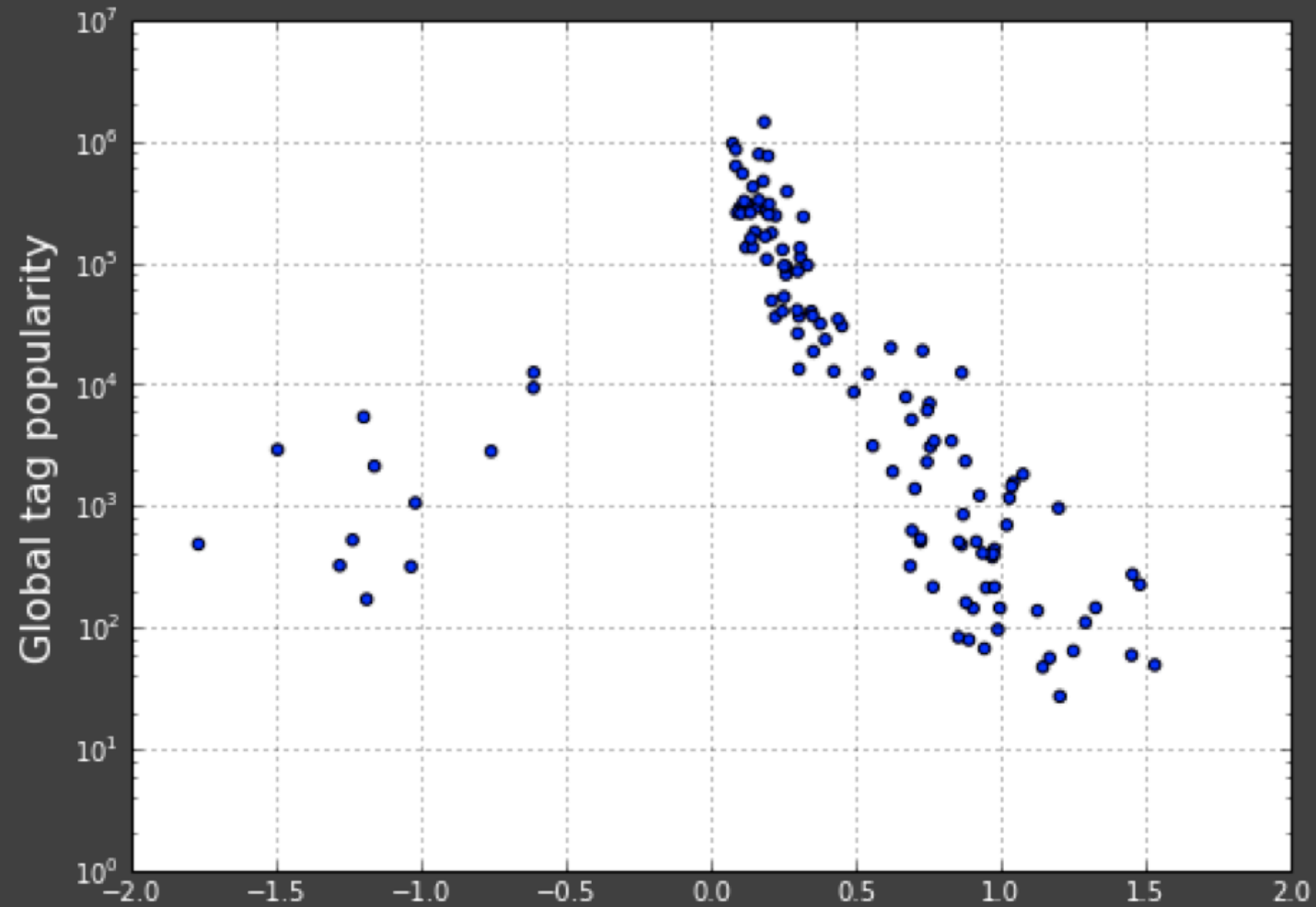
(but the difference is small: 1-2 listens over 6 months)

# What about particular tags?

Another regression model, with 2 changes:

1. Binary regressors for all tags with  $\geq 5$  occurrences (2,290).
2. Control for listening in peak month, but not preceding months

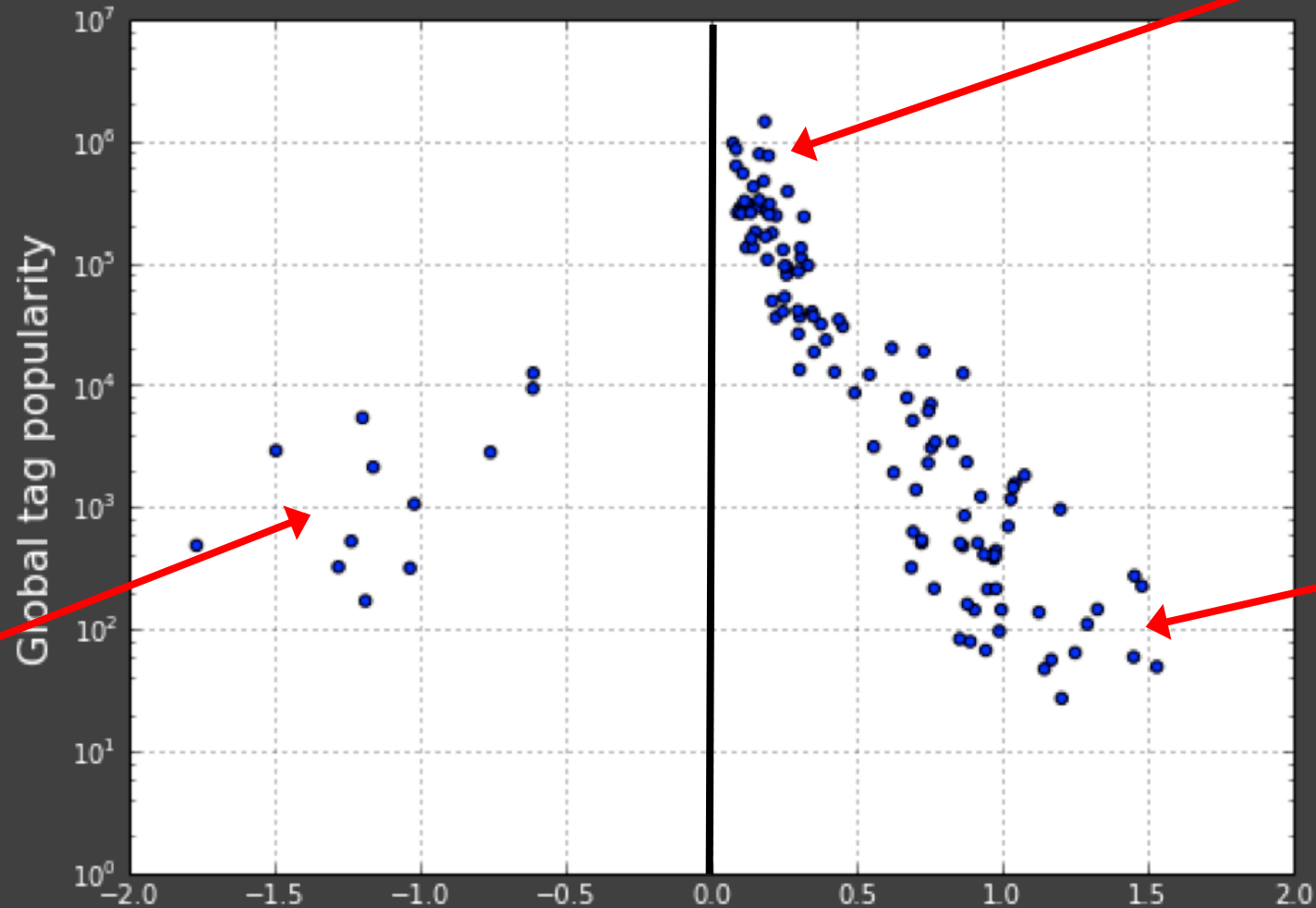
# What about particular tags?





# What about particular tags?

“seen live”, “electronic”, “death metal”,  
“hip hop”, “female vocalists”



“meaningful songs”,  
“mymusic”, “meine  
liebsten”, “i love”,  
“artmunz”

“saw them live”,  
“5432fun”, “loved  
artists”, “great song”,  
“Omix”

So, *do* tags really serve as retrieval aids?

Tags certainly do not ***always*** function as retrieval aids, and our results suggest that retrieval may actually be an uncommon tagging motivation.

Tags certainly do not *always* function as retrieval aids, and our results suggest that retrieval may actually be an uncommon tagging motivation.

Moving forward...

Refinement of modeling approaches

In-depth tag analyses

Other datasets?

jlorince@indiana.edu  
<https://pages.iu.edu/~jlorince>

