

9-26-2011

# Music Video Redundancy and Half-Life in YouTube

Matthias Prellwitz

Michael L. Nelson

*Old Dominion University*, [mnelson@odu.edu](mailto:mnelson@odu.edu)

Follow this and additional works at: [https://digitalcommons.odu.edu/computerscience\\_presentations](https://digitalcommons.odu.edu/computerscience_presentations)



Part of the [Archival Science Commons](#)

---

## Recommended Citation

Prellwitz, Matthias and Nelson, Michael L., "Music Video Redundancy and Half-Life in YouTube" (2011). *Computer Science Presentations*. 17.

[https://digitalcommons.odu.edu/computerscience\\_presentations/17](https://digitalcommons.odu.edu/computerscience_presentations/17)

This Book is brought to you for free and open access by the Computer Science at ODU Digital Commons. It has been accepted for inclusion in Computer Science Presentations by an authorized administrator of ODU Digital Commons. For more information, please contact [digitalcommons@odu.edu](mailto:digitalcommons@odu.edu).

# **Music Video Redundancy and Half-Life in YouTube**

**Matthias Prellwitz and Michael L. Nelson**

matthias.prellwitz@googlemail.com

mln@cs.odu.edu

# Linking to a particular copy “Rolling Stones - Satisfaction”

The screenshot shows a web browser window with the address bar containing the URL <http://www.youtube.com/watch?v=3a7CHPy04s8>. The page displays the YouTube interface for the video "Satisfaction-Rolling Stones" by user "xxampaxx". The video player shows the Rolling Stones' tongue logo with the text "Satisfaction-Rolling Stones" overlaid. The video has 1,230,104 views and is 4:03 minutes long. The page includes a search bar, navigation links, and a list of suggestions on the right side.

# Metadata lost when YouTube video disappears

The screenshot shows a YouTube page for the video "The Rolling Stones – Satisfaction". A white box highlights the video title and URL. A red banner at the top of the video area states: "This video is no longer available due to a copyright claim by ABKCO." Below this is a video player with a play button and a "Click For Sound" button. The video player shows a scene from the movie "The Graduate" with the text "Where does this silk come from?". Below the video player are sections for "Join the largest worldwide video-sharing community!" with "Create Account" and "Sign In" buttons, "Recommended for You" with four video thumbnails, and "Spotlight Videos" with two video thumbnails.

video title      The Rolling Stones – Satisfaction  
url                http://www.youtube.com/watch?v=214szPQBUYc

! This video is no longer available due to a copyright claim by ABKCO.

Being serious about entertainment means doing whatever it takes.  
Learn More

TAKE CHARGE: AMERICAN EXPRESS

Click For Sound  
Where does this silk come from?

0:14 / 2:01      240p

Join the largest worldwide video-sharing community!  
Create Account >      Already have an account? Sign In

Recommended for You      Learn More

Making Of "The Right Kind of Wrong"      3:49  
The Beatles - I Want To Hold Your Hand      2:47  
Beatsteaks - She Was Great      2:47  
Apple: Steve Jobs si mannia Sony      0:51

Spotlight Videos

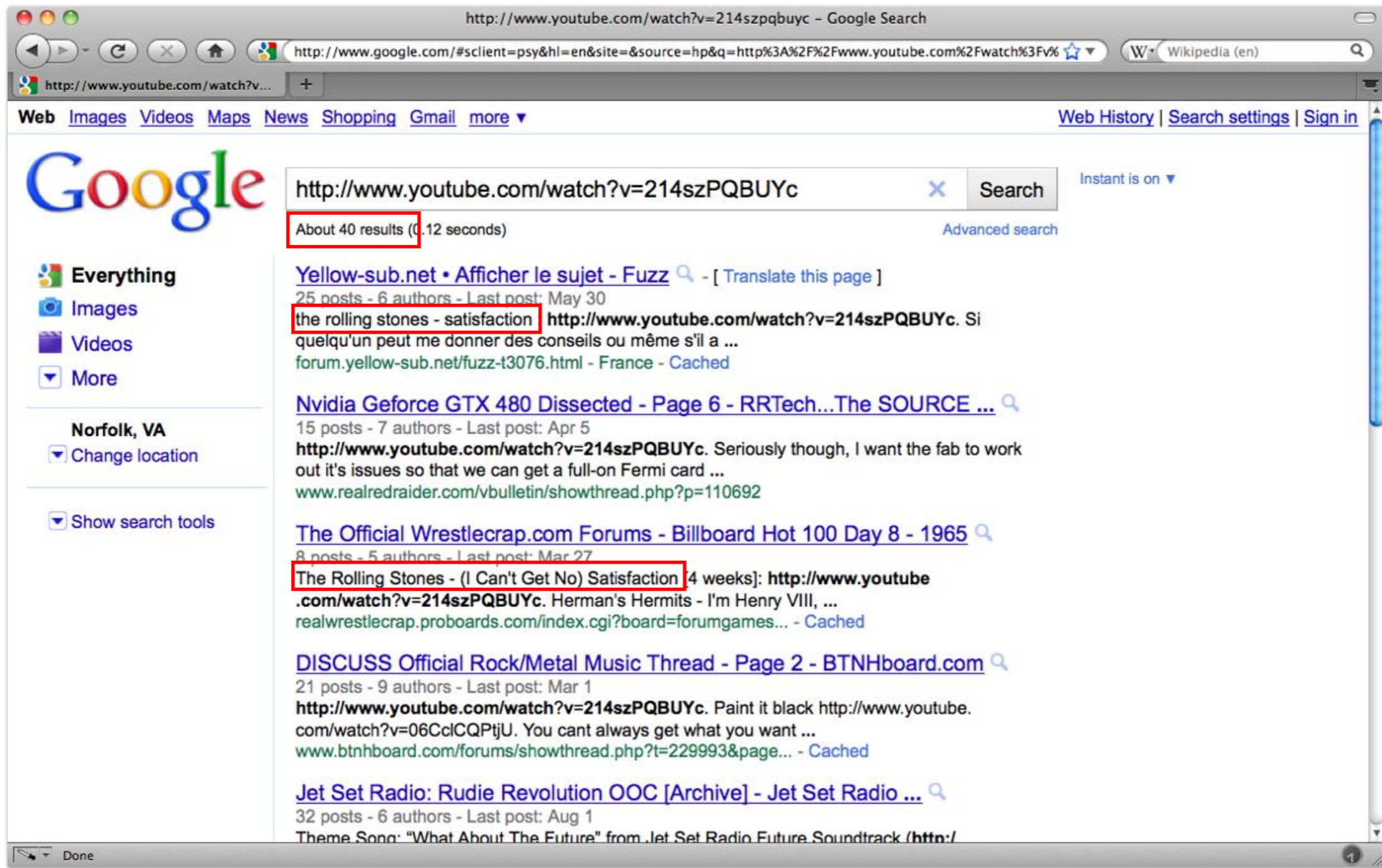
Tropfest  
The world's largest short film festival is now calling for entries on YouTube. Watch to learn more.  
Presented by: TROPFEST

Shock - Tropfest Winner 2010  
by TROPFEST  
32,139 views      4:18

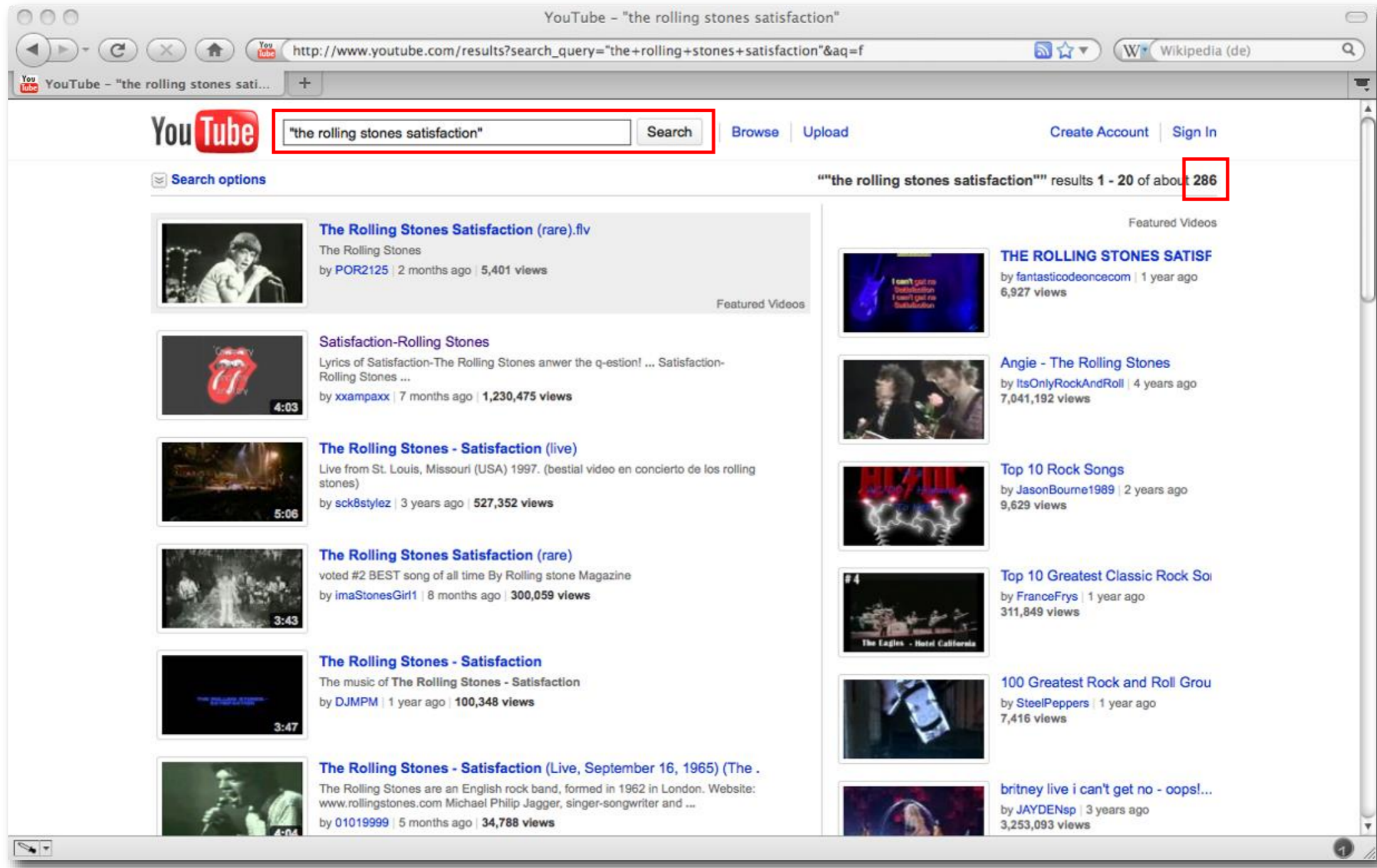
PRODUCTION MEETING  
by TROPFEST  
2,646 views      3:04

Transferring data from v7.lscache1.c.youtube.com...

# Metadata hard to recover from Search Engines



# But nearly 300 copies remain in YouTube



# Linking music-related URIs

- ▶ Transparent URI semantics

- ▶ [http://www.last.fm/music/John+Lennon/\\_/Imagine](http://www.last.fm/music/John+Lennon/_/Imagine)
- ▶ <http://www.ilike.com/artist/The+Cribs/track/I%27m+A+Realist>
- ▶ [http://www.last.fm/music/Johnny+Cash/\\_/Highwayman](http://www.last.fm/music/Johnny+Cash/_/Highwayman)

- ▶ Opaque URI semantics

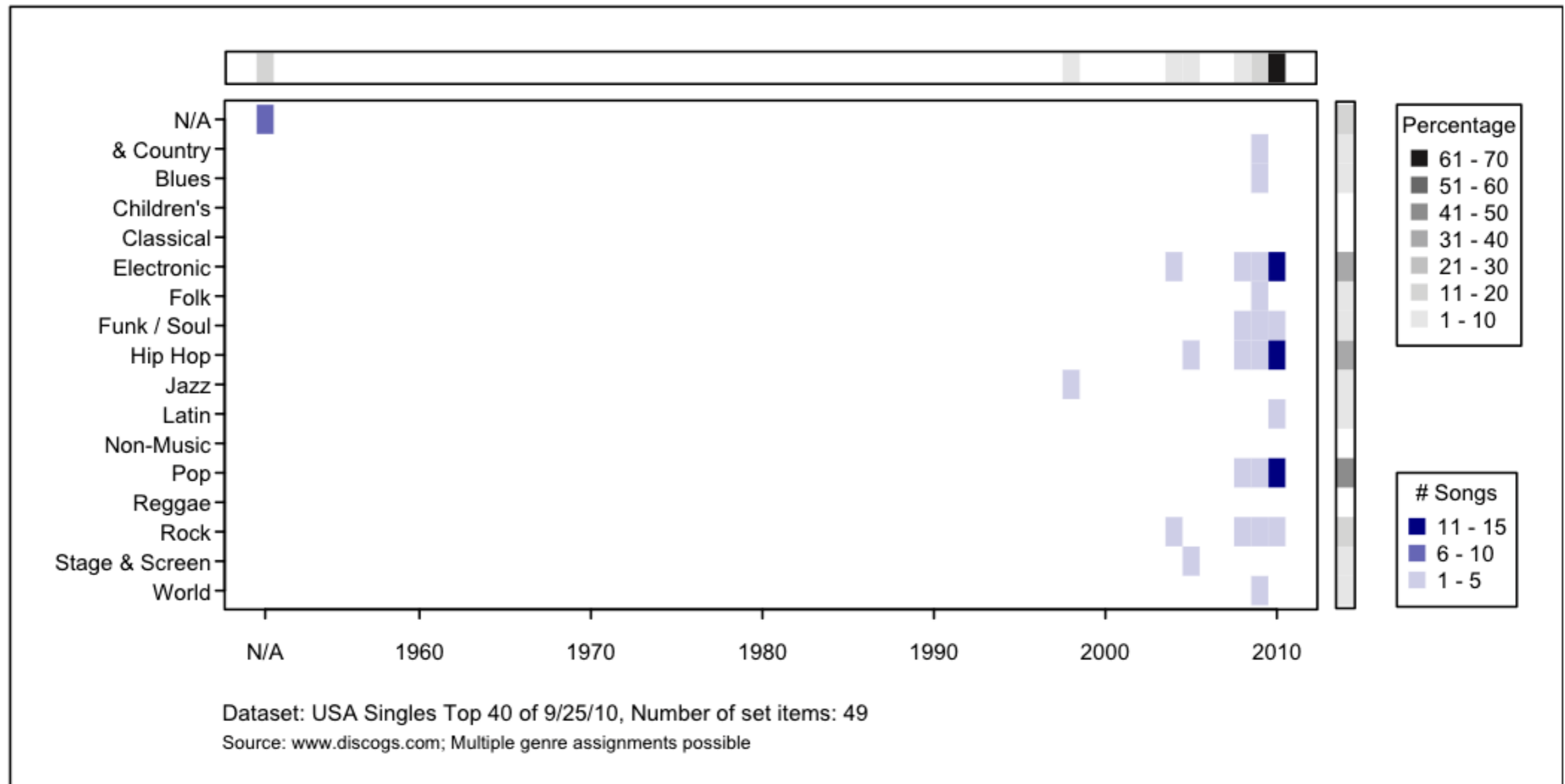
- ▶ <http://vids.myspace.com/index.cfm?fuseaction=vids.individual&videoid=5168491>
- ▶ <http://www.youtube.com/watch?v=VST2KKIYn50>

Website	Rank
youtube.com	3
myspace.com	61
dailymotion.com	109
vimeo.com	159
last.fm	508
myvideo.com	1,445

**Alexa-Ranking**  
Global, Dec 10  
<http://www.alexacom/topsites>

# Popular Music

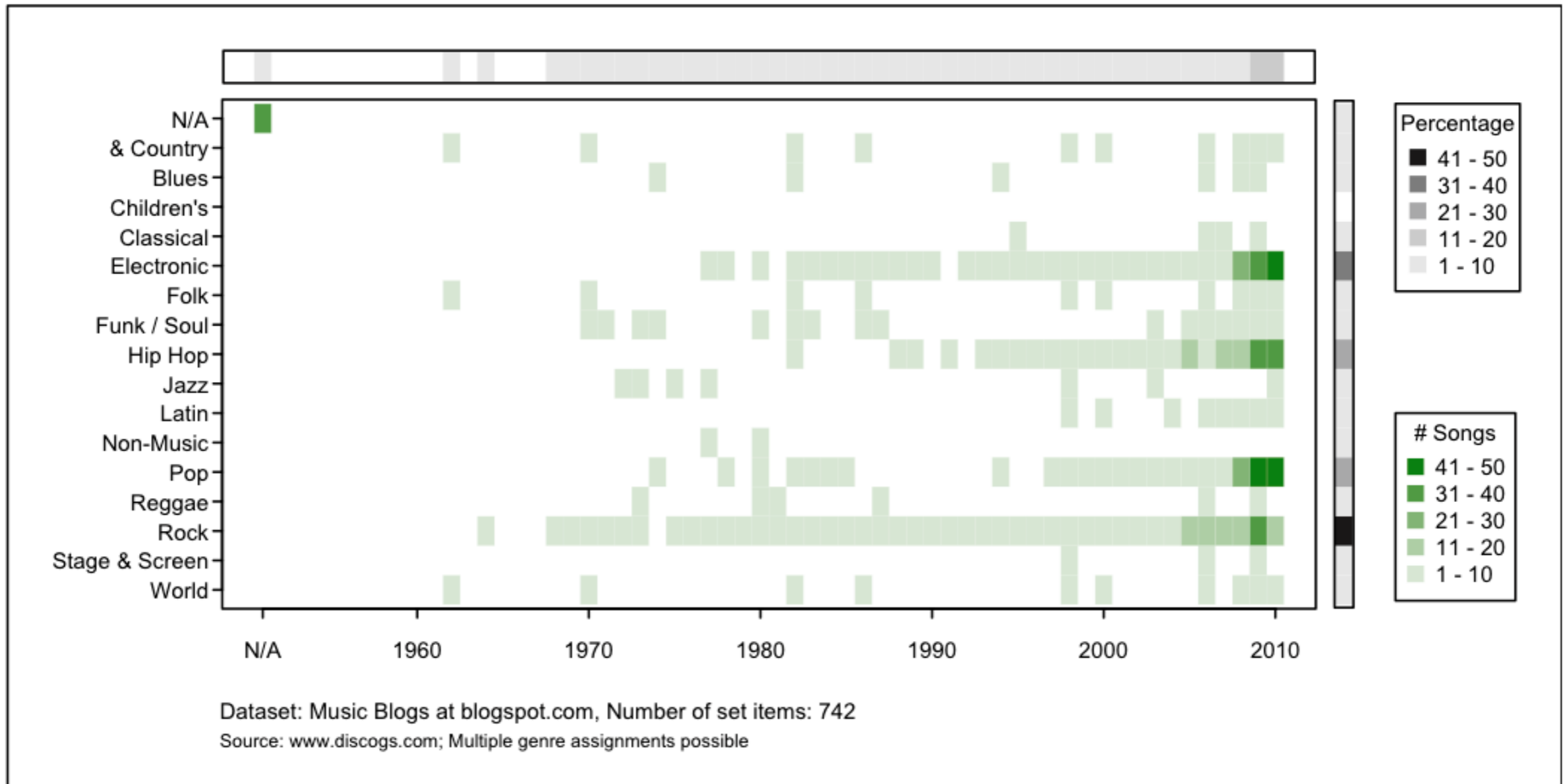
## US Top 40 Singles Charts of 9/25/10





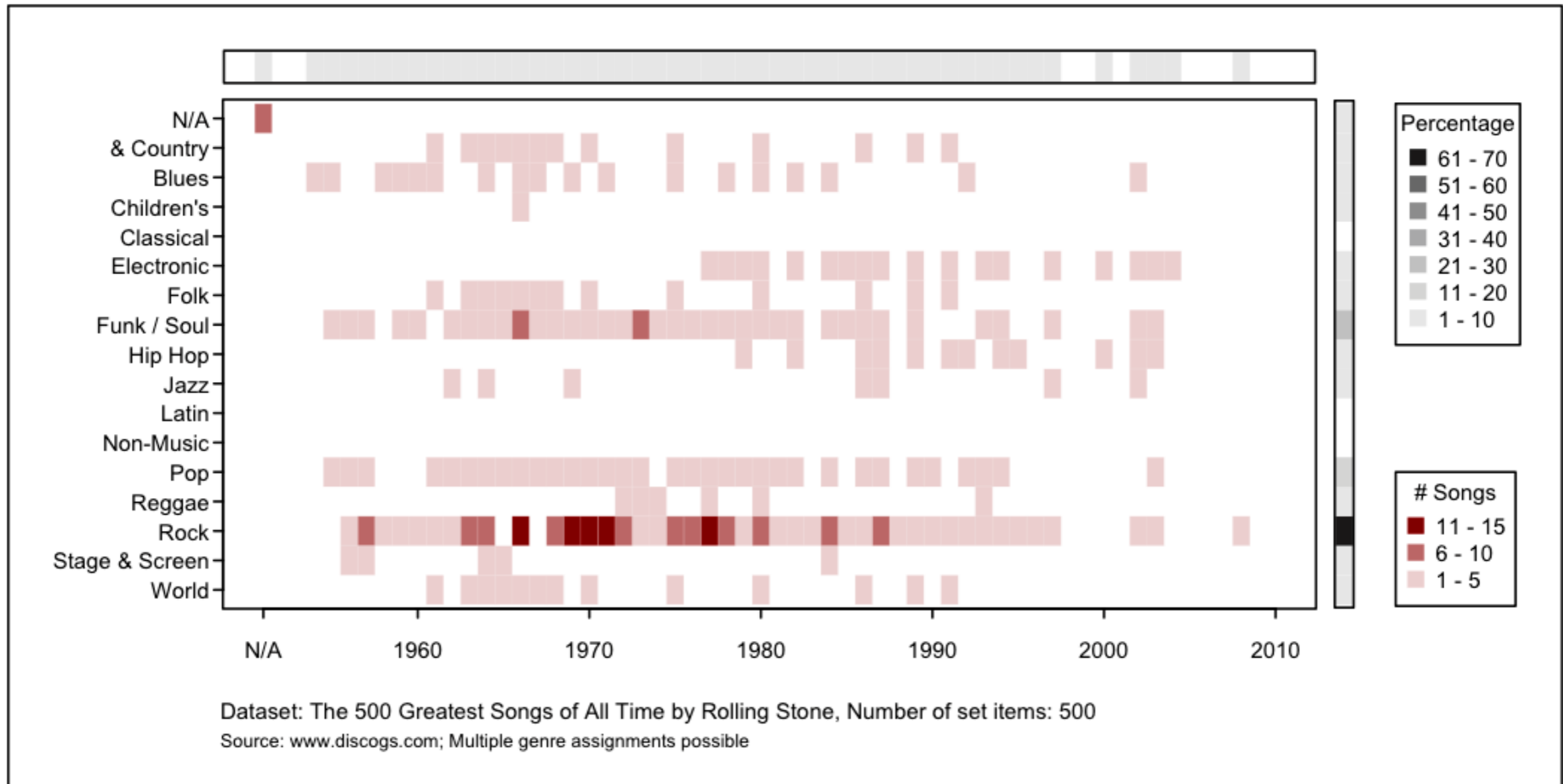
# Popular Music

## Selected Music Blogs



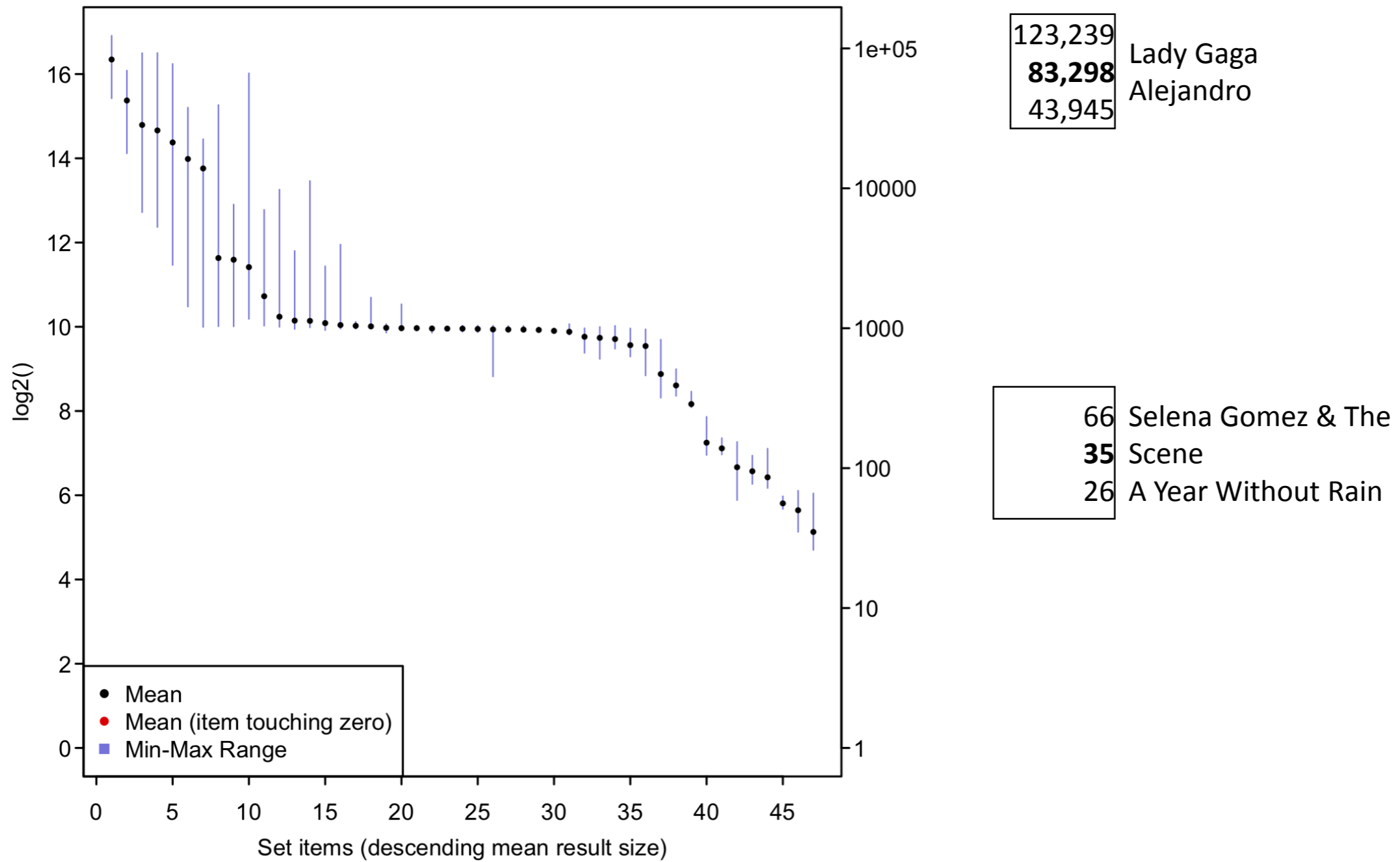
# Popular Music

## The 500 Greatest Songs of all Time

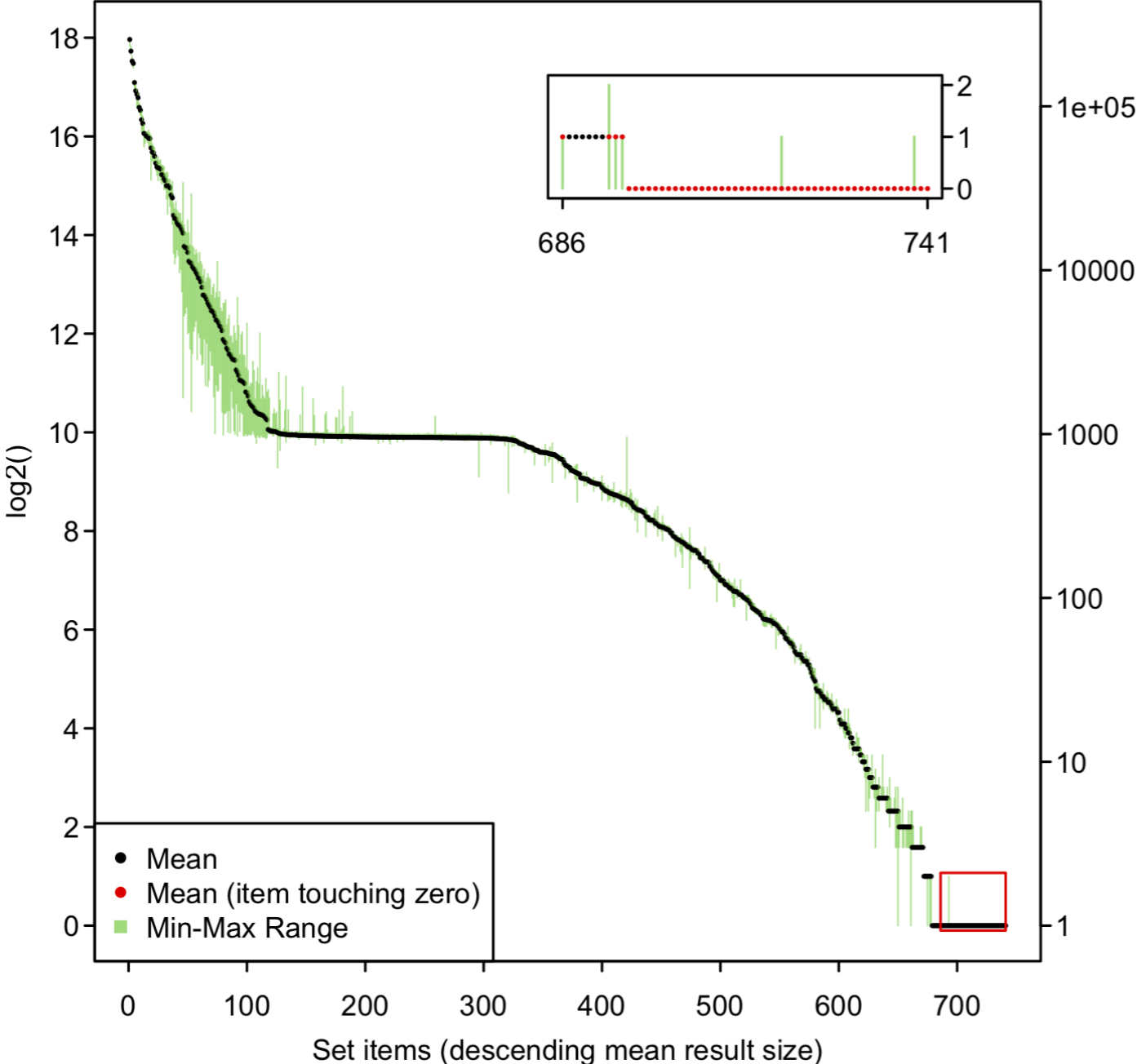


# Total Result Size Range

## US Top 40 Singles Charts of 9/25/10



# Total Result Size Range Selected Music Blogs



264,753  
**256,205**  
 232,936

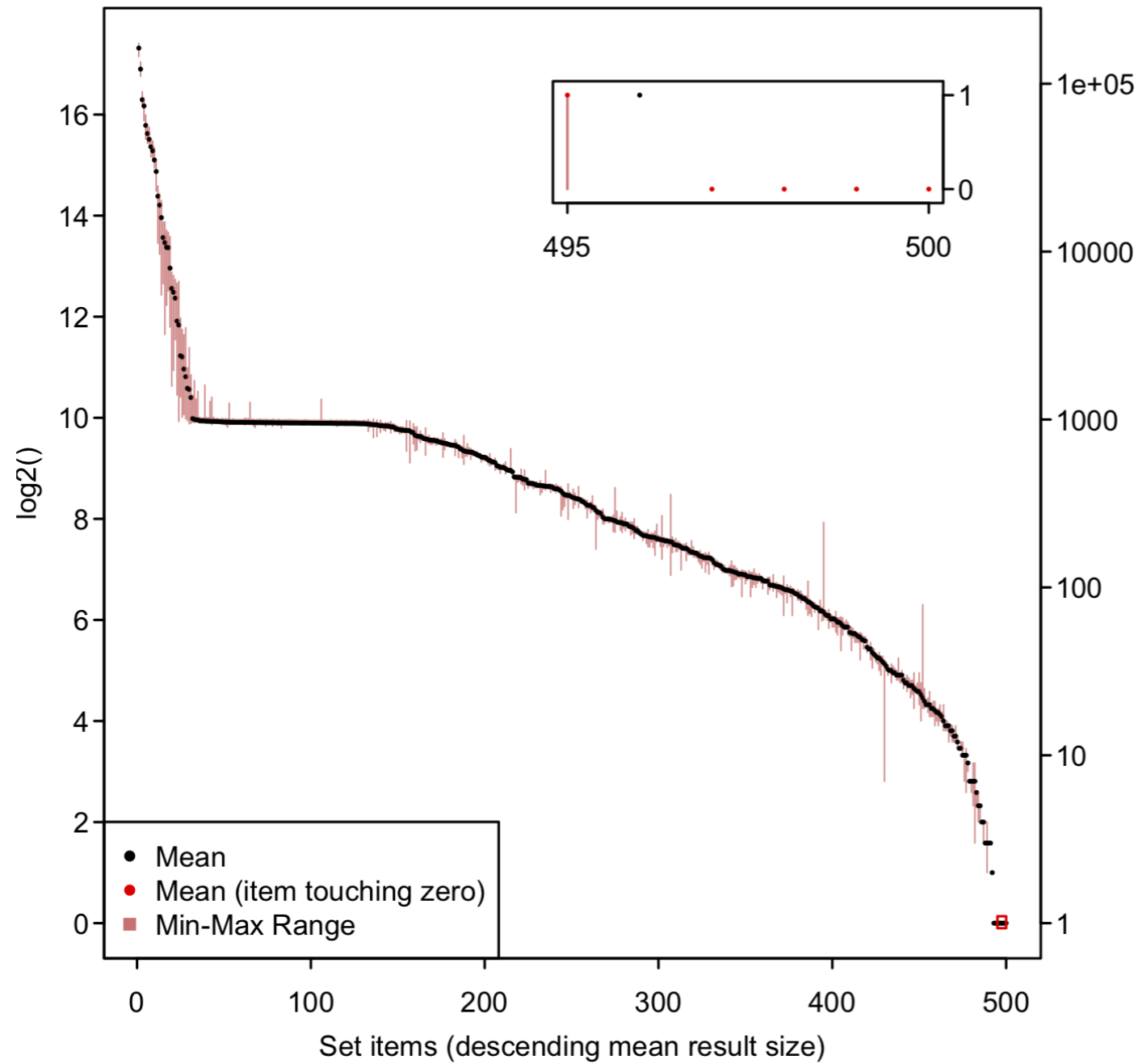
Lady Gaga  
 Bad Romance

0  
 0  
 0

Mariah Carey featuring  
 Juelz Santana & Bone  
 Thugs-n-Harmony  
 Don't Forget About Us

# Total Result Size Range

## The 500 Greatest Songs of all Time



174,088  
162,937  
145,076

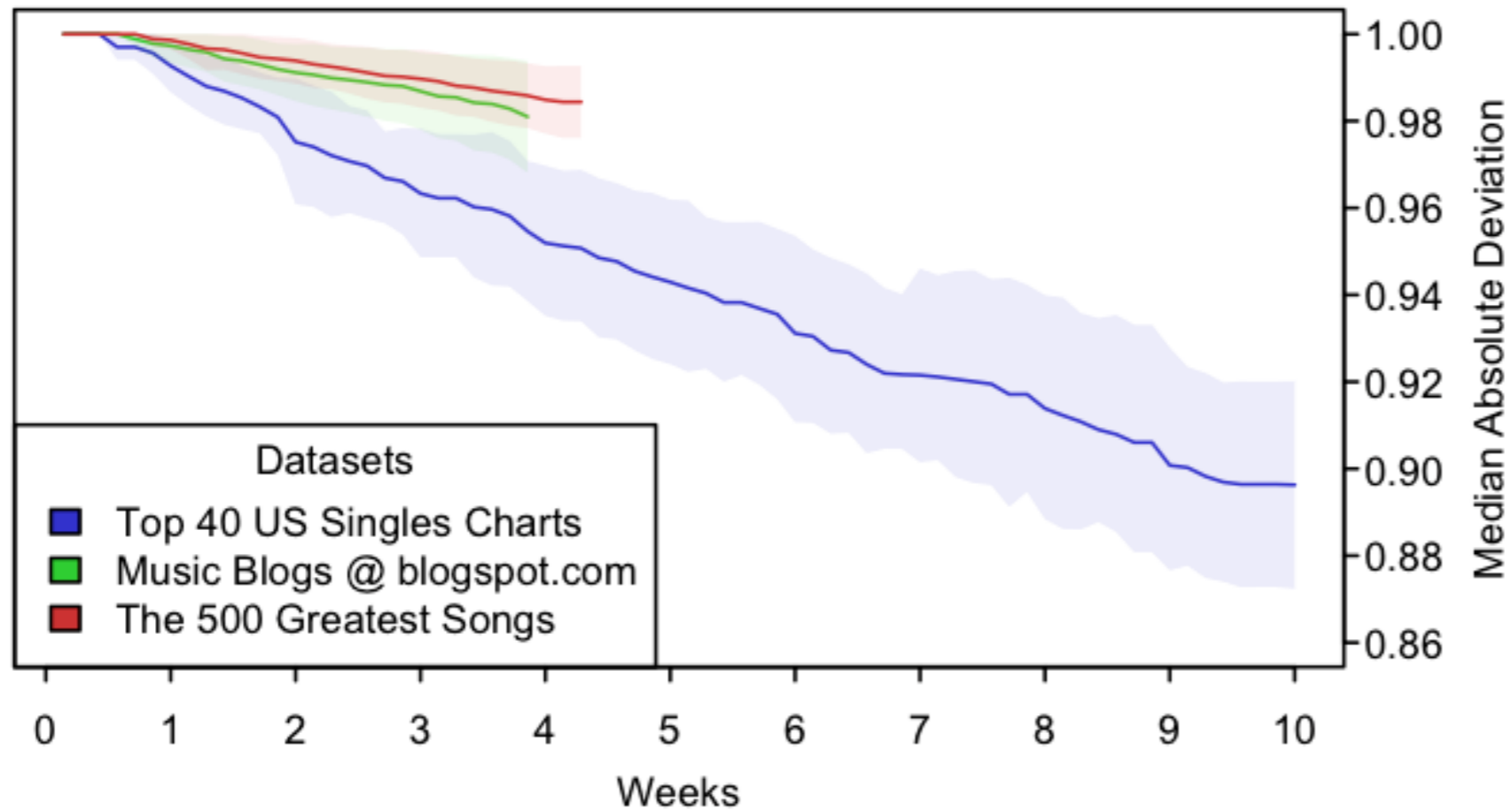
Michael Jackson  
Billie Jean

0  
0  
0

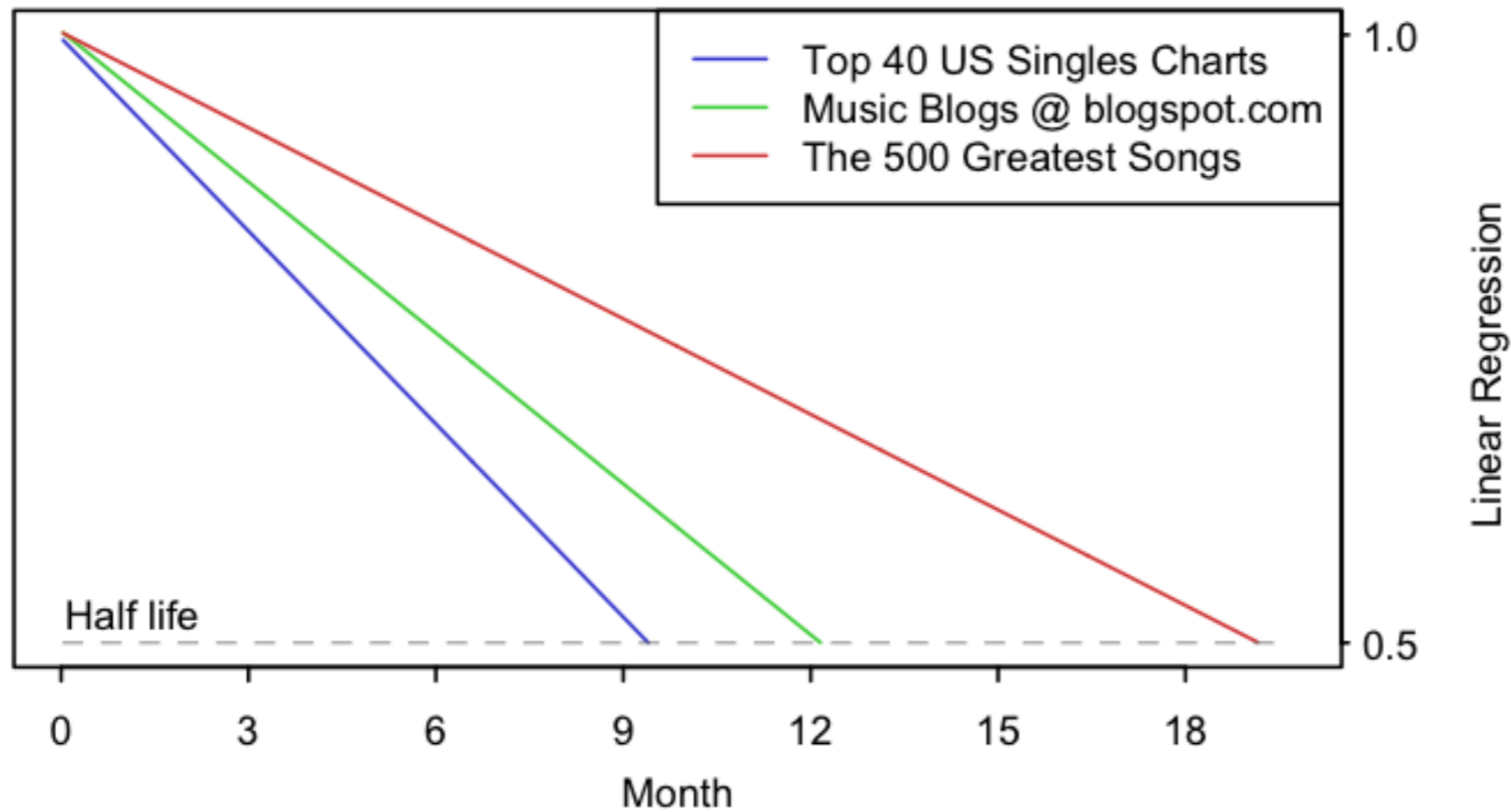
The Isley Brothers  
That Lady (Part 1 and 2)

# URI Unavailability

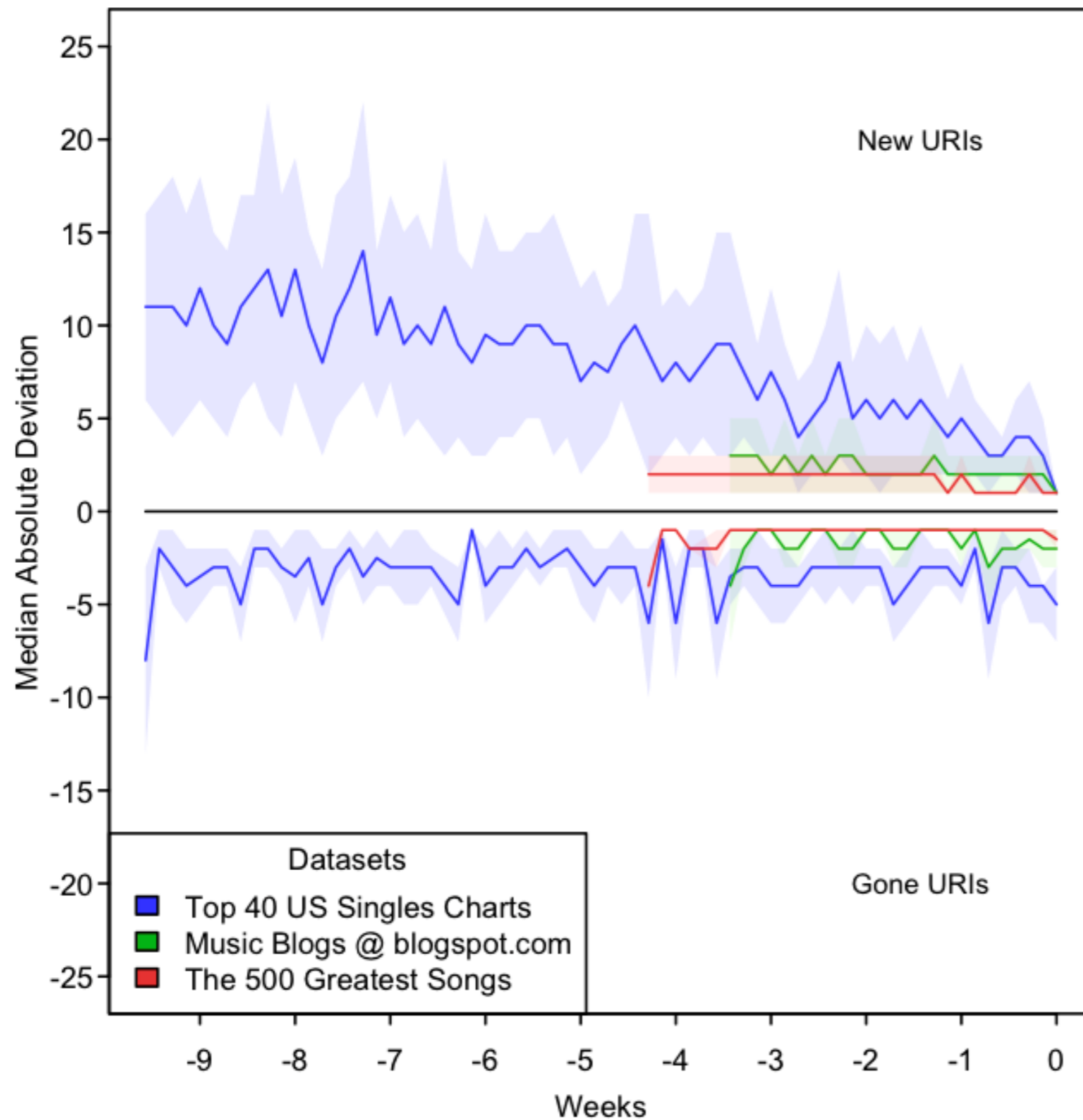
## Rooted from a selected collection



# URI Unavailability Expected Half-life

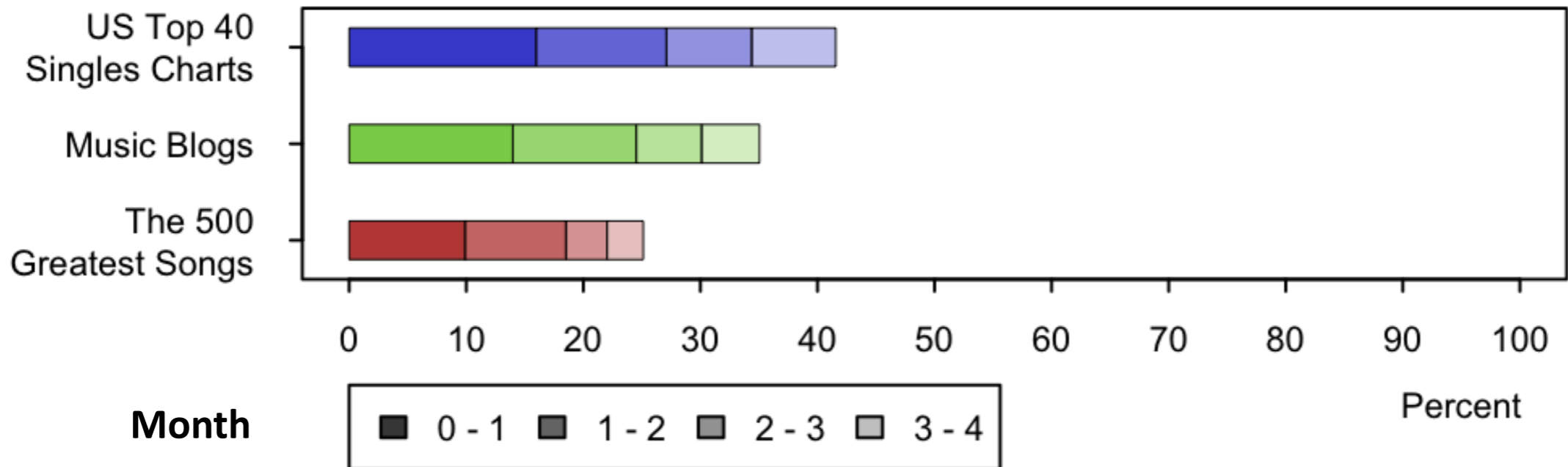
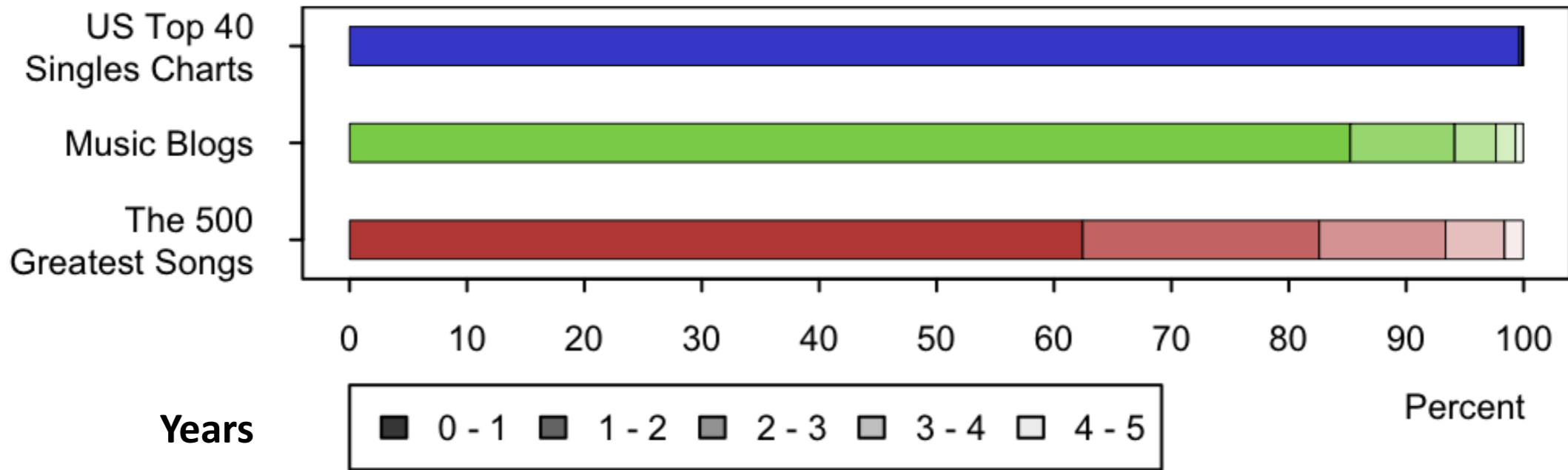


# URI Publication and Removal Rate





# Lifetimes of unavailable videos



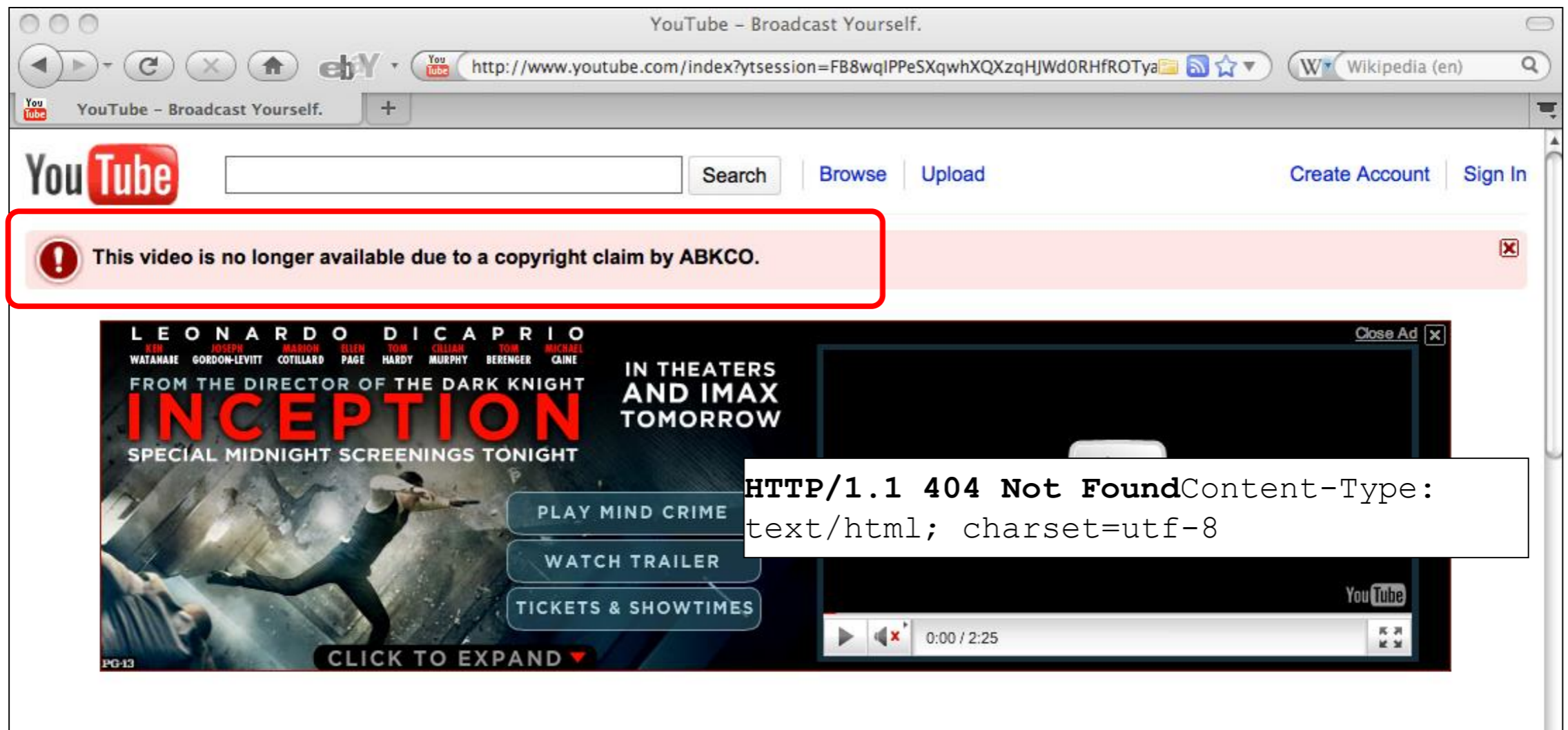
## Reasons for no unavailable videos

Gone Reason	Percent
<b>Third-party triggered</b>	<b>48.8</b>
This video contains content from [copyright owner], who has blocked it (in your country) on copyright grounds.	26.0
This video is no longer available because the YouTube account associated with this video has been terminated due to multiple third-party notifications of copyright infringement.	18.3
This video is no longer available due to a copyright claim by [copyright owner]	4.5
<b>YouTube triggered</b>	<b>23.7</b>
This video is no longer available because the YouTube account associated with this video has been terminated.	13.5
This video has been removed as a violation of YouTube's [...] policy	3.1
This video has been removed because its content violated YouTube's Terms of Service.	2.8
This video contains content from [copyright owner]. It is not available (in your country).	2.4
This video is no longer available because the YouTube account associated with this video has been terminated due to repeated copyright infringements.	1.9

<b>User triggered</b>	<b>13.2</b>
This video is no longer available because the uploader has closed their YouTube account.	12.3
This video has been deleted.	0.9
<b>Other/Errors</b>	<b>14.3</b>
Authentication required (see 2.2 reasons)	7.8
N/A	3.3
Verification required (see 2.2 reasons)	2.3
Upgrade to Flash Player 10	0.7
The video you have requested is not available	<0.1
This video is not available in your country	<0.1
The video is a duplicate copy of a previously uploaded video	<0.1

# When a YouTube video disappears

- ▶ video title The Rolling Stones - Satisfaction
- ▶ url <http://www.youtube.com/watch?v=214szPQBUYc>
- ▶ Published 2009-06-13 13:44                      Removed                      2010-04-09 (300 days online)



# Metadata purged from YouTube Databases

- ▶ **Video feed**

```
curl -I "http://gdata.youtube.com/feeds/api/videos/214szPQBUYc"
```

```
HTTP/1.1 404 Not Found
```

```
Content-Type: text/html; charset=UTF-8
```

```
Private video
```

- ▶ **Related videos**

```
curl -I
```

```
"http://gdata.youtube.com/feeds/api/videos/214szPQBUYc/related"
```

```
HTTP/1.1 404 Not Found
```

```
Content-Type: text/html; charset=UTF-8
```

```
Parent Video not found
```

- ▶ **Video comments**

```
curl -I
```

```
"http://gdata.youtube.com/feeds/api/videos/214szPQBUYc/comments"
```

```
HTTP/1.1 200 OK Content-Type: application/atom+xml; charset=UTF-8
```

# Metadata Normalization

The screenshot shows a YouTube video player for 'Michael Jackson - Billie Jean (With Lyrics + HQ Sound)'. The video is at 4:54 / 4:54. The page includes a search bar with 'Michael Jackson', a 'Share' button, and a 'Replay' button. A 'Next in Michael Jackson Mix' section shows 'Michael Jackson Medley' with 11,723,412 views. A Bing advertisement is visible. The video title 'Michael Jackson - Billie Jean (With Lyrics + HQ Sound)' is highlighted with a red box. Below the video, the artist 'TheJukeboxRock' and date 'December 28, 2008' are shown. The view count is '6,420,997 views'. At the bottom, there are links for 'Michael Jackson - Billie Jean' (highlighted with a red box) and 'Download This Song: iTunes AmazonMP3' (with 'AmazonMP3' highlighted with a red box).

YouTube - Michael Jackson - Billie Jean (With Lyrics + HQ Sound)

http://www.youtube.com/watch?v=75sx7U6dAB4

Michael Jackson

Michael Jackson - Billie Jean (With Lyrics + HQ Sound)

TheJukeboxRock 83 videos

Share Replay

Michael Jackson - Black or White W/LYRICS  
From: supermileyM  
Views: 100255

Michael Jackson - Smooth Criminal Lyrics  
From: MJKingofPop  
Views: 5819117

Billie Jean with lyrics  
From: kim8789

STOP SEARCHING AND START EXPLORING WITH BING.

Artist: Michael Jackson  
Title: Billie Jean (Single Version)

TheJukeboxRock | December 28, 2008  
6,420,997 views

Michael Jackson - Billie Jean  
Download This Song: iTunes AmazonMP3

# Availability of music-related metadata

- ▶ parsed out only at the first time a URI showed up in the result list for the first time
  - ▶ YouTube crawling restrictions

**Table 1: Availability of additional music-related metadata**

Dataset	YouTube	Amazon
US Single Charts Top 40 Total URI count: 85,831	42.0% 36,040	41.4% 35,491
Music Blogs Total URI count: 582,068	11.9% 69,068	11.6% 67,483
The 500 greatest songs of all time Total URI count: 313,542	30.0% 94,275	28.2% 88,290

- ▶ Remaining portion
  - ▶ query video title against music related services via search engines
    - ▶ Google/Yahoo! with site parameter [www.last.fm/music](http://www.last.fm/music)

# Retrieving and preserving a video's metadata

- ▶ **Active preservation**  
attempt once a video copy is available
  - ▶ Parse HTML out for structured music-related metadata
    - ▶ YouTube generated meta data
    - ▶ AmazonMP3 affiliate link
    - ▶ search engines with free-form video title against music-related websites
- ▶ **Preserving** metadata into the public web infrastructure



```
1 {
2   "uri":"http://www.youtube.com/watch?v=214szPQBUYc",
3   "title":"The Rolling Stones - Satisfaction",
4   "dur":"344",
5   "music":{
6     "artist":"The Rolling Stones",
7     "song":"Satisfaction"
8   }
9 }
```

# Preservation Prototype

The screenshot shows the YouTubeResolve web application interface. At the top, the browser address bar displays the URL: `http://webspace.cs.odu.edu/~mprellwi/app/ytres/preserve?url=http%3A%2F%2Fwww.youtube.com%2Fwatch%3Fv%3D%3A3a7cHPy04s8`. The main content area is divided into several sections:

- URL:** `http://www.youtube.com/watch?v=3a7cHPy04s8` with a "query" button and the text "and please be patient".
- General Video data:** Video Title: Satisfaction-Rolling Stones; Duration: 243 sec (4:03 min).
- Step 2: Select Metadata to Preserve:**
  - URI Related:** Canolized Video Watch URI. Options: regular (selected) `http://www.youtube.com/watch?v=3a7cHPy04s8`, shortened `http://youtu.be/3a7cHPy04s8`.
  - Video Related:** Video Title (selected) and Duration (selected). Options: title `Satisfaction-Rolling Stones` (checkbox), Duration `243 sec` (radio).
  - Music Related:** Artist and Title. Options: Youtube Data `The Rolling Stones` (radio), `(I Can't Get No) Satisfaction` (radio); Your Data (empty input fields).
- Step 3: Publish Metadata:** A JSON snippet is shown in a text area: 

```
#3a7cHPy04s8{
  "u": "http://youtu.be/3a7cHPy04s8",
  "ma": "The Rolling Stones",
  "ms": "(I Can't Get No) Satisfaction",
  "d": "243"
}
```

 To the right, it indicates "119/140 char (adjusted without indents, spaces, and line breaks as shown)" and includes a "Tweet" button.
- A note: "Please make sure to save your share as public note/bookmark. So our resolver can use search engines to retrieve the information afterwards, and you and others can benefit." Below this is another JSON snippet: 

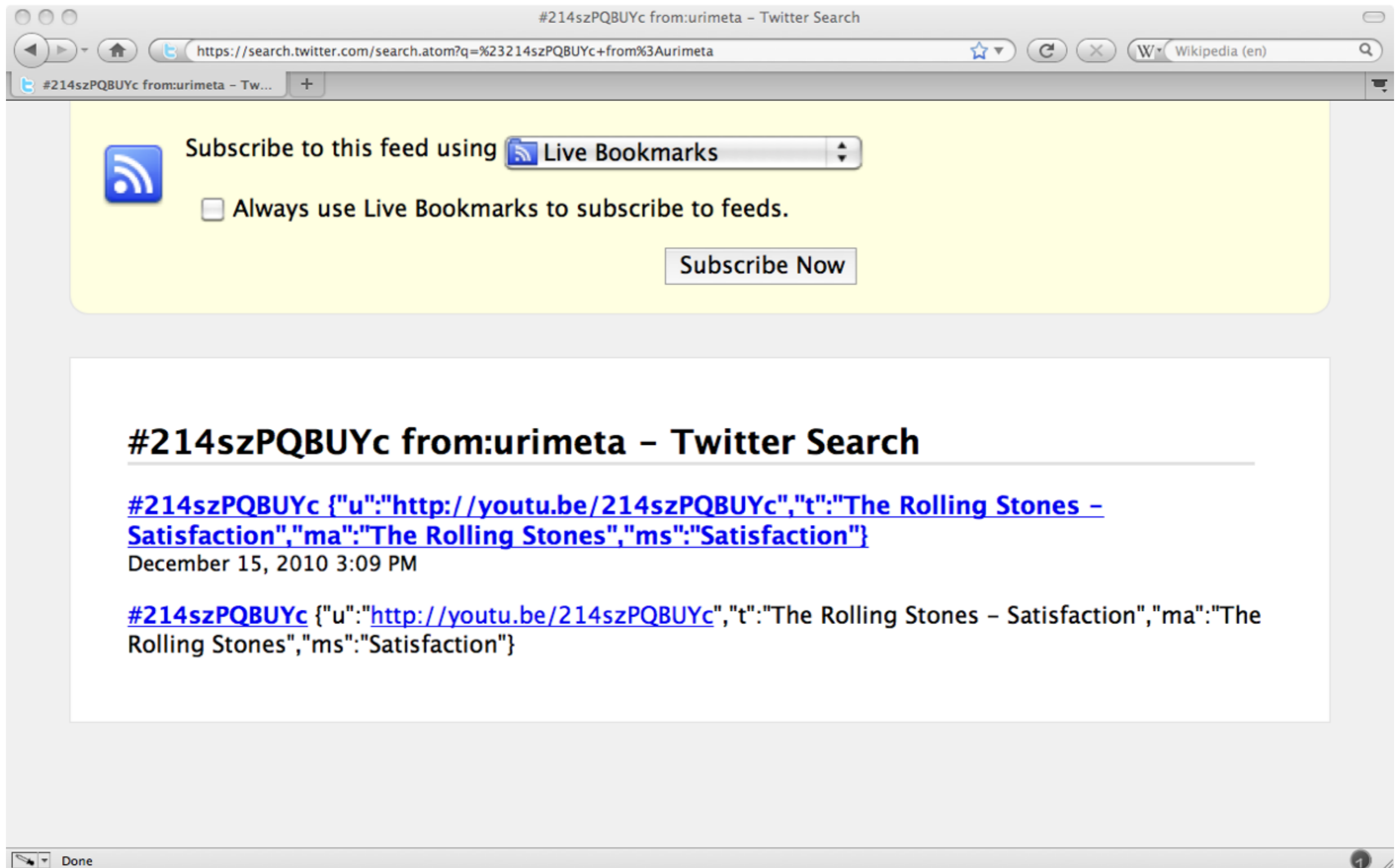
```
{
  "uri": "http://www.youtube.com/watch?v=3a7cHPy04s8",
  "dur": "243",
  "music": {
    "artist": "The Rolling Stones",
    "song": "(I Can't Get No) Satisfaction"
  }
}
```

 Social sharing buttons for "tumblr." and "Buzz" are visible.



# Metadata preservation

## Example: twitter



The screenshot shows a browser window with the address bar containing the URL `https://search.twitter.com/search.atom?q=%23214szPQBUYc+from%3Aurimeta`. The page title is "#214szPQBUYc from:urimeta - Twitter Search". Below the address bar, there is a yellow banner with a RSS icon and the text "Subscribe to this feed using Live Bookmarks". A checkbox labeled "Always use Live Bookmarks to subscribe to feeds." is present, and a "Subscribe Now" button is at the bottom right of the banner.

The main content area displays the search results for "#214szPQBUYc from:urimeta - Twitter Search". The first result is a blue link: `#214szPQBUYc {"u":"http://youtu.be/214szPQBUYc","t":"The Rolling Stones - Satisfaction","ma":"The Rolling Stones","ms":"Satisfaction"}`. Below the link is the text "December 15, 2010 3:09 PM". The second result is a black link: `#214szPQBUYc {"u":"http://youtu.be/214szPQBUYc","t":"The Rolling Stones - Satisfaction","ma":"The Rolling Stones","ms":"Satisfaction"}`.

The browser's status bar at the bottom shows "Done" and a page count of "1".

# Pointing to a Resolver service

- ▶ `http://ytresolve.cs.odu.edu/r/http://www.youtube.com/watch?v=214szPQBUYc/`
- ▶ Author-side approach
  - ▶ content creator points directly to a resolver service
- ▶ Server-side approach
  - ▶ Plugin/Renderer class automatically rewrites YouTube video watch URIs to resolver service
- ▶ Client-side approach
  - ▶ Web-Browser plugin intercepts click on Youtube video watch URIs and redirects to resolver service



# YouTube Resolver service

```
http://www.youtube.com/watch?v=214szPQBUYc
http://www.youtube.com/v/214szPQBUYc
http://www.youtu.be/214szPQBUYc
http://www.youtube.com/user/WEASELxLOVER#p/a/u/2/214szPQBUYc
```

HTTP/1.1 404 Not Found

HTTP/1.1 200 OK  
HTTP/1.1 303 See Others \*

HTTP  
Status  
Code

*redirect*



search for preserved metadata  
▶ in list of designated accounts

exact best available granularity

query YouTube API with those

Provided (and evaluate)  
alternative copies

\*)

```
http://www.youtube.com/verify_controversy...
http://www.youtube.com/verify\_age...
https://www.google.com/accounts/ServiceLogin...
http://www.youtube.com/das\_captcha..
```

# Future Work

- ▶ Evaluation of preservation and **retrieval quality** of chosen services
  - ▶ exchange services
- ▶ additional **automation** of preservation process
  - ▶ once YT URI was passed for resolving
- ▶ Evaluation of retrieved available **copies**
  - ▶ redirect to best copy instead of returning a list to choose
- ▶ Consider **international requesters**
  - ▶ taking requester's location (country) into account

# Summary

- ▶ Pointing to a specific YouTube video copy by its URI has a risk of disappearance
  - ▶ alternative copies over time available
  - ▶ YouTube URIs unlikely to be cached once gone
  - ▶ YouTube metadata only reliable for available URIs
    - ▶ active preservation attempt
- ▶ Introducing a level of indirection: Resolver service
  - ▶ check URI status and location header
  - ▶ search the public web for injected metadata
    - ▶ query for alternative copies