

# Same or different? Comparing Facebook data with survey answers on music taste ~~and leisure time~~ ~~habits~~

Workshop on the Concept of Quality for Big Data  
*Budapest*  
*2020.02.25*

Zoltán Kmetty PhD  
Renáta Németh PhD

**Eötvös Loránd University, Faculty of Social Sciences**



# Content

- Focus
  - measurement validity
- Topic
  - music preferences
- Data
  - Social media + survey data

# Research problem I

- Internal validity of data depends on the accuracy of the measurement, on the fact whether we have actually measured what we supposed to measure.
- (some) known problems with surveys
  - Self reported data
  - Recall bias
  - Context effect
  - interviewer-related measurement error
  - survey questions are simplistic, incapable of grabbing the complexity of social reality
  - Social desirability

# Research problem II

- (some) known problems with social media data sources
  - shifts in the algorithm
  - Platform effect (affordance)
  - Unstructured data sources
  - Social desirability
  - (problems with representativity)
- **Combination of surveys and social media data offers an opportunity to cross-validate the different methods**

# Music preferences

- Central role in cultural sociology
- It does not depend heavily on the financial situation of the household
- Measurement
  - Simple survey questions
  - Listening and evaluating songs
  - Digital traces
    - Twitter, LastFm, Facebook, Spotify
- There are **no** studies on the relationship between digital trace data and survey data regarding music preference

# Data

- Novel joint data source of combined Facebook and survey data
- 150 participants (non-probability quota sample )
- Fieldwork: 2019 between April and September
- Respondents download their own FB archives and shared it with the researchers
  - posts, comments, likes and reactions, pages, friends, profile, and ads data.
  - Data was fully anonymized
- Survey with the same respondents: politics, media usage, self-representation, mental health, spare-time activities, and music preferences

# Measurement

- Survey
  - 1-7 scale
- Facebook
  - Page likes (digitally expressed interest)
    - 3803 pages categorized as music pages (by FB)
      - Manually categorized by BA students into categories (same genres as in the survey)
    - Number of music pages likes
    - Dummy variable of music page like (more than 1 pages coded to 1)
  - Ads interest data (inferred interest)
    - algorithmic classification of users by FB
    - No category for “Mulatós” music genre (special Hungarian genre)

# Results - Digitally expressed interest

11% percent had no music page likes  
**why don't they have any page likes?**

## EXP1

We calculated the number of really liked genres based on self-reported interest data (6 or 7 on the scale) and analyzed the average of this variable within those who had music page like and within those who didn't – no significant relationship

## EXP2

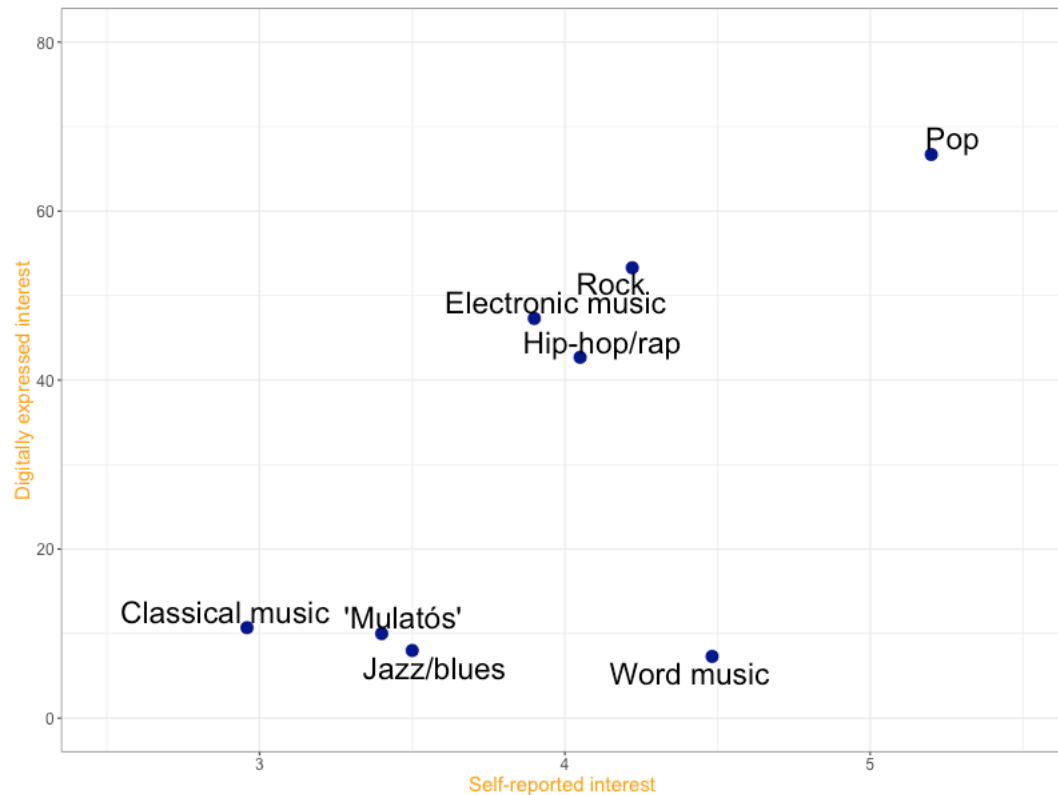
Seventy-five percent of the whole sample uses Facebook more than once a day. This proportion is 70 percent within those who don't follow any music-related page, and 76 percent within those who follow at least one page

## EXP3:

We calculated the mean number of all page likes in the two groups. The mean value was 55 within the group with no music page like, and 620 in the other group and this difference **was significant** ( $p=0.00$ ) even in this small sample.



# Results - Digitally expressed interest

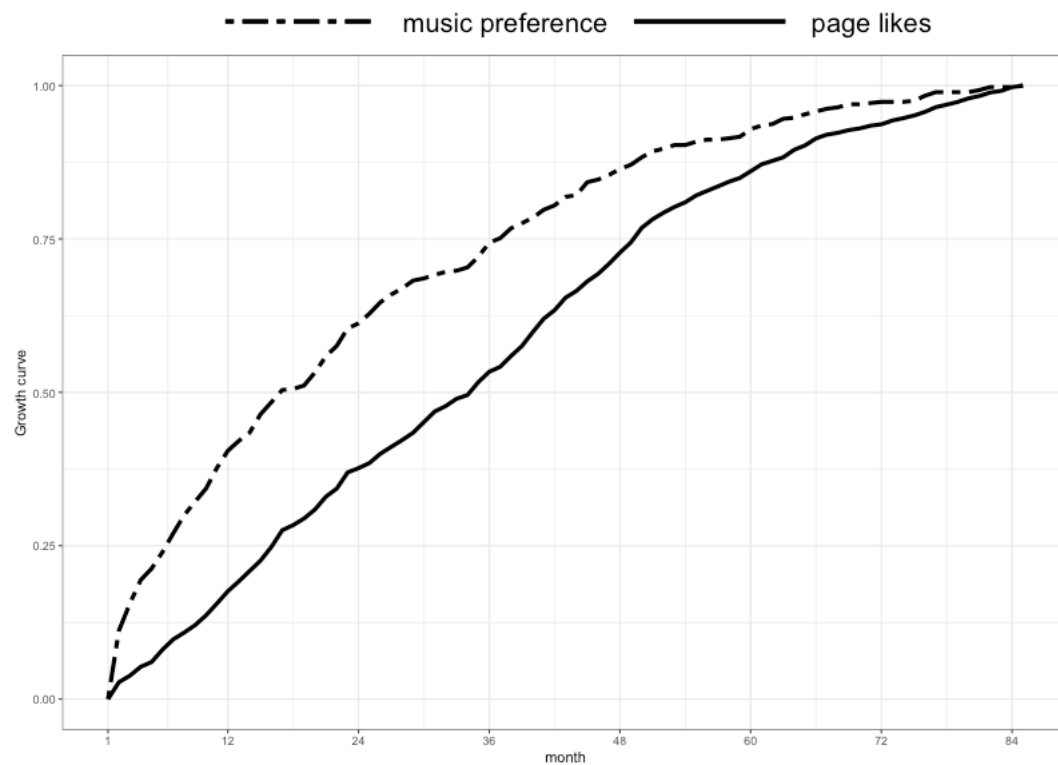


Comparing self-reported interest and digitally expressed interest

	Classical music	Electronic music	Hip-hop/rap	Jazz/blues	'Mulatós'	Pop	Rock	World-music
Classical music	.26**	-.09	-.21*	.04	.12	-.16*	.11	.06
Electronic music		.27**	.20*	-.13	-.15	.07	-.12	.20*
Hip-hop/rap			.44**	.05	.32**	.26**	-.09	.22**
Jazz/blues				.15	-.02	-.04	.11	-.05
'Mulatós'					.23**	.24**	.01	-.08
Pop						.21*	-.10	.14
Rock							.38**	.20*
World-music								-.02

Correlation of self-reported interest data (row) and digitally expressed interest data (column)

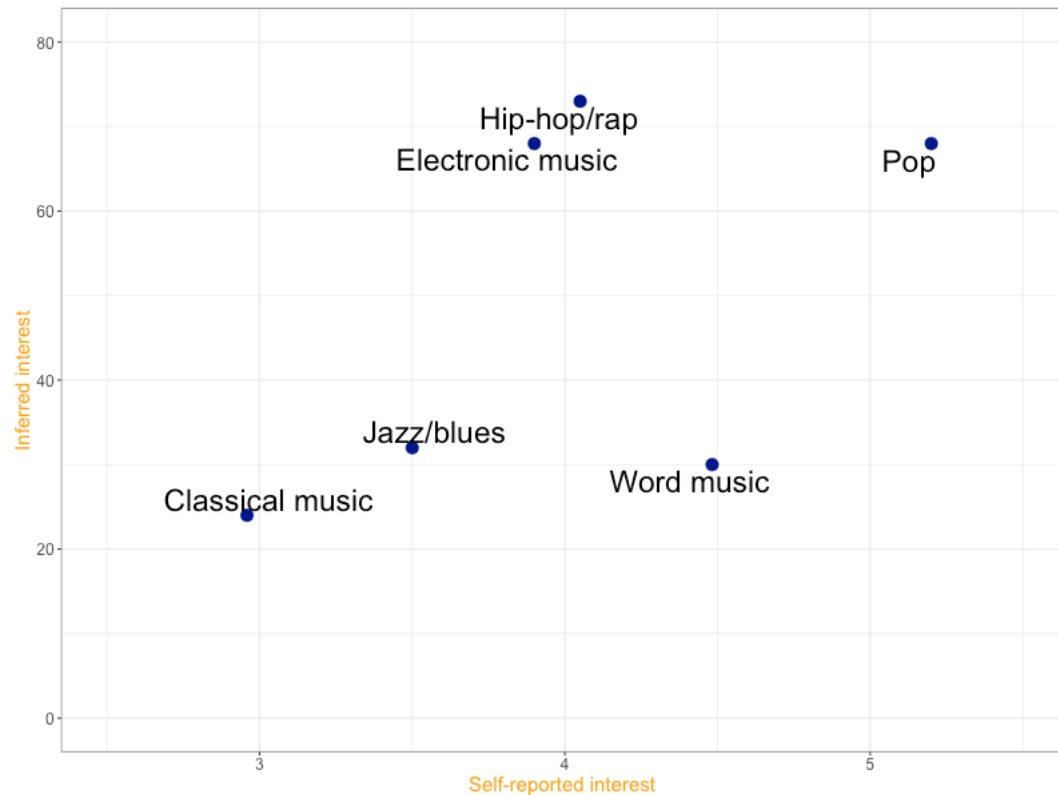
# Results - Evolution of music taste



The Cumulative growth of Facebook page likes and ratio of preferred music genres

- No kick-off increase for page likes
  - For the first five year linear trend, but after that the growth rate is starting to decrease
    - Age effect of the sample?
- After two years of Facebook usage, we can map 61 percent of the genres a user likes, after 3 years, 74 percent, and after 4 years, 86 percent.
- ! If we use the last two years of FB activity the correlation between survey and FB data become much lower

# Results - Inferred interest



Comparing self-reported interest and inferred interest

- Eighty-one percent of our respondents were linked to rock music, and even the least popular classical music was linked to every fourth of our respondents.
- The correlation with the survey was lower compared with the digitally expressed data

# Discussion

- Our paper is the first attempt to compare self-reported music preference with Facebook-based music preference classification.
- Word music measured differently
  - What does this category means?
  - Underrepresented in FB?
- Blurry category boundaries
  - Rap/hip-hop?
- Different genres could have different representation in FB

# Discussion

- How to create the variables from social media data?
  - No gold standards
  - Lot of operationalization related questions
- Ads interest is a black-box
  - No „mulatós” category in ads interest data
- The combination of survey and social media data could help us to understand sociological phenomenon like social desirability

# Thank you for your attention!

- Paper is available here:  
<https://arxiv.org/pdf/2002.00501.pdf>
- This research was supported by National Research, Development and Innovation Office of Hungary (NKFIH), grant number FK-128981.
- [kmetty.zoltan@tatk.elte.hu](mailto:kmetty.zoltan@tatk.elte.hu)
- [nemeth.renata@tatk.elte.hu](mailto:nemeth.renata@tatk.elte.hu)

