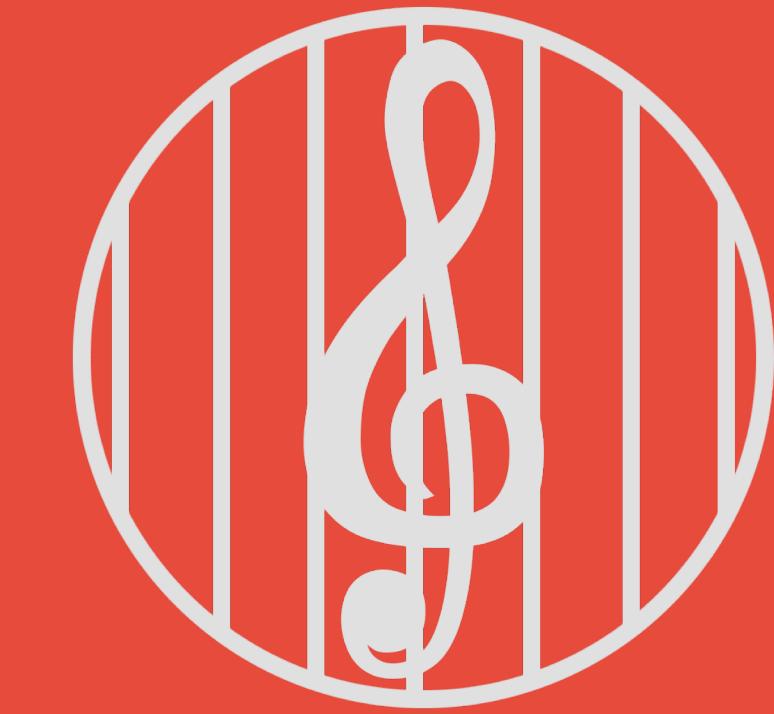




# Sound of Censorship

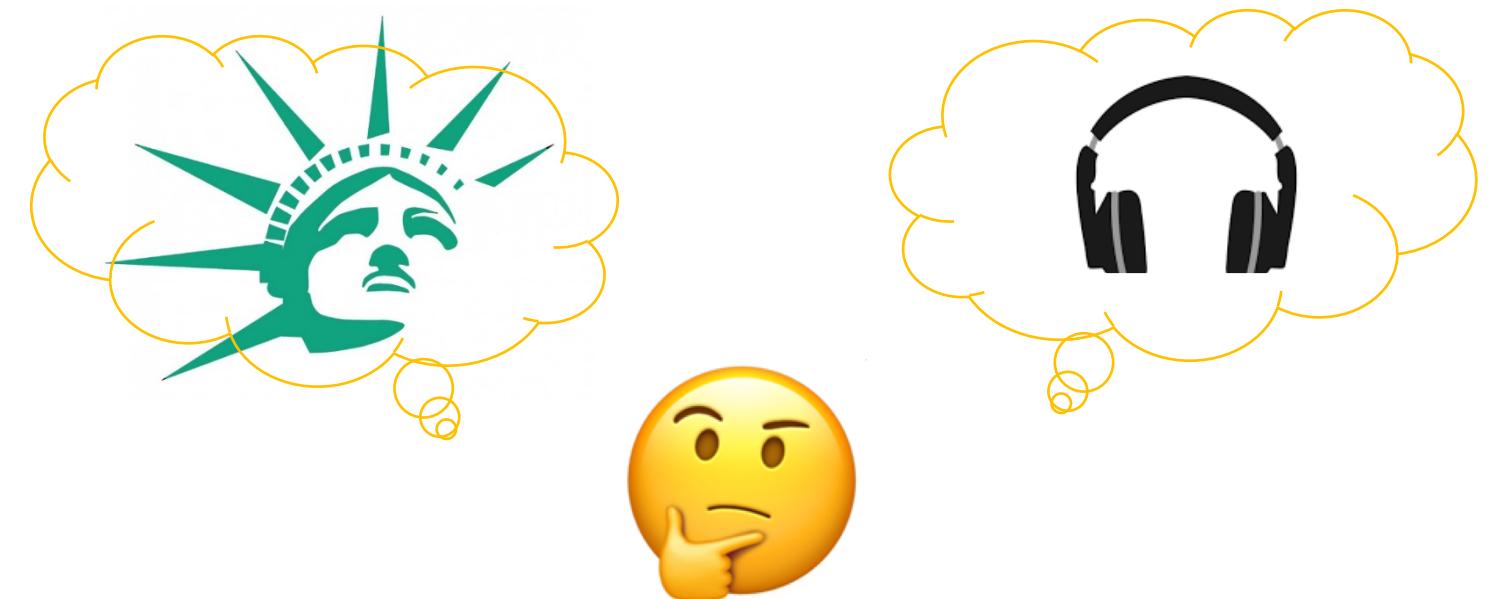
Nikhil Pant, Deniz Bayazit, Athiwat Thoopthong, Cancan Huang  
 Brown University, Providence, Rhode Island 02912, USA



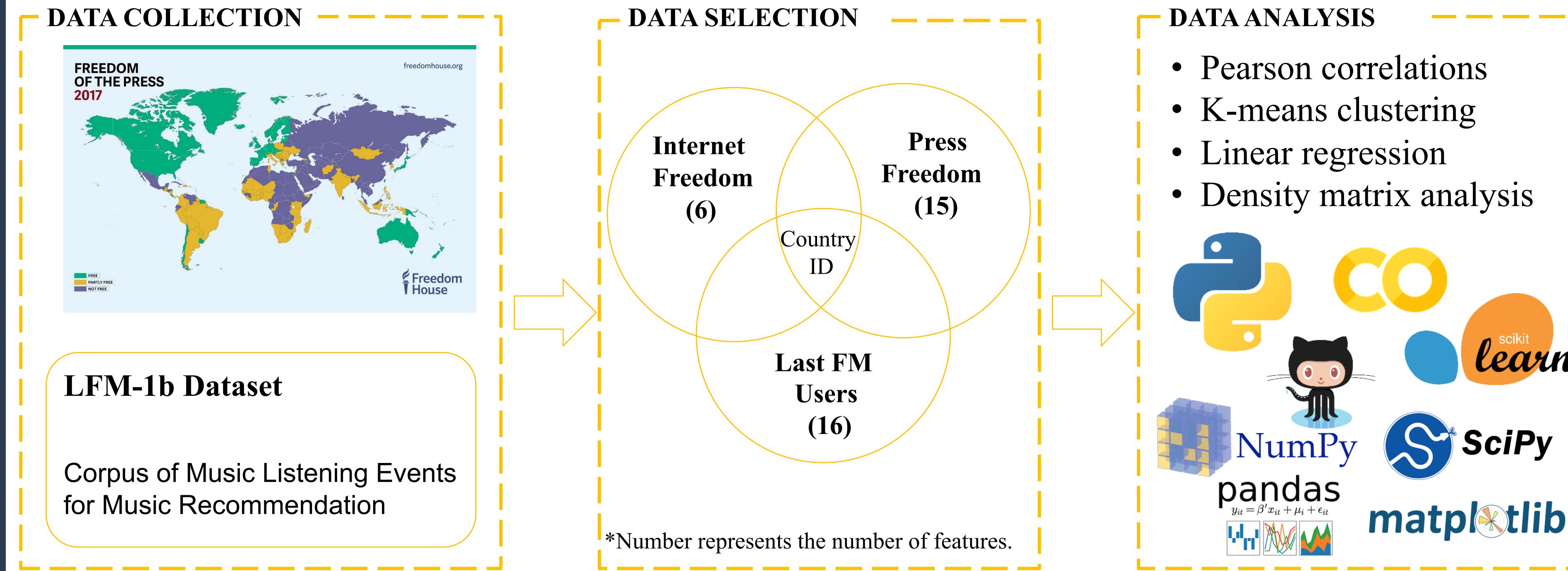
DATA SCIENCE

## Abstract

Correlations between user features (country, age, gender, etc.) and a country's freedom of expression features (press freedom ranking, internet freedom status, etc.) have been investigated. Our findings show that based upon our data, there isn't a significant relation between a country's freedom expression and citizen listening habits. The lack of significant results can be explained by the lack of data, which is due to heavy censorship in less free countries.



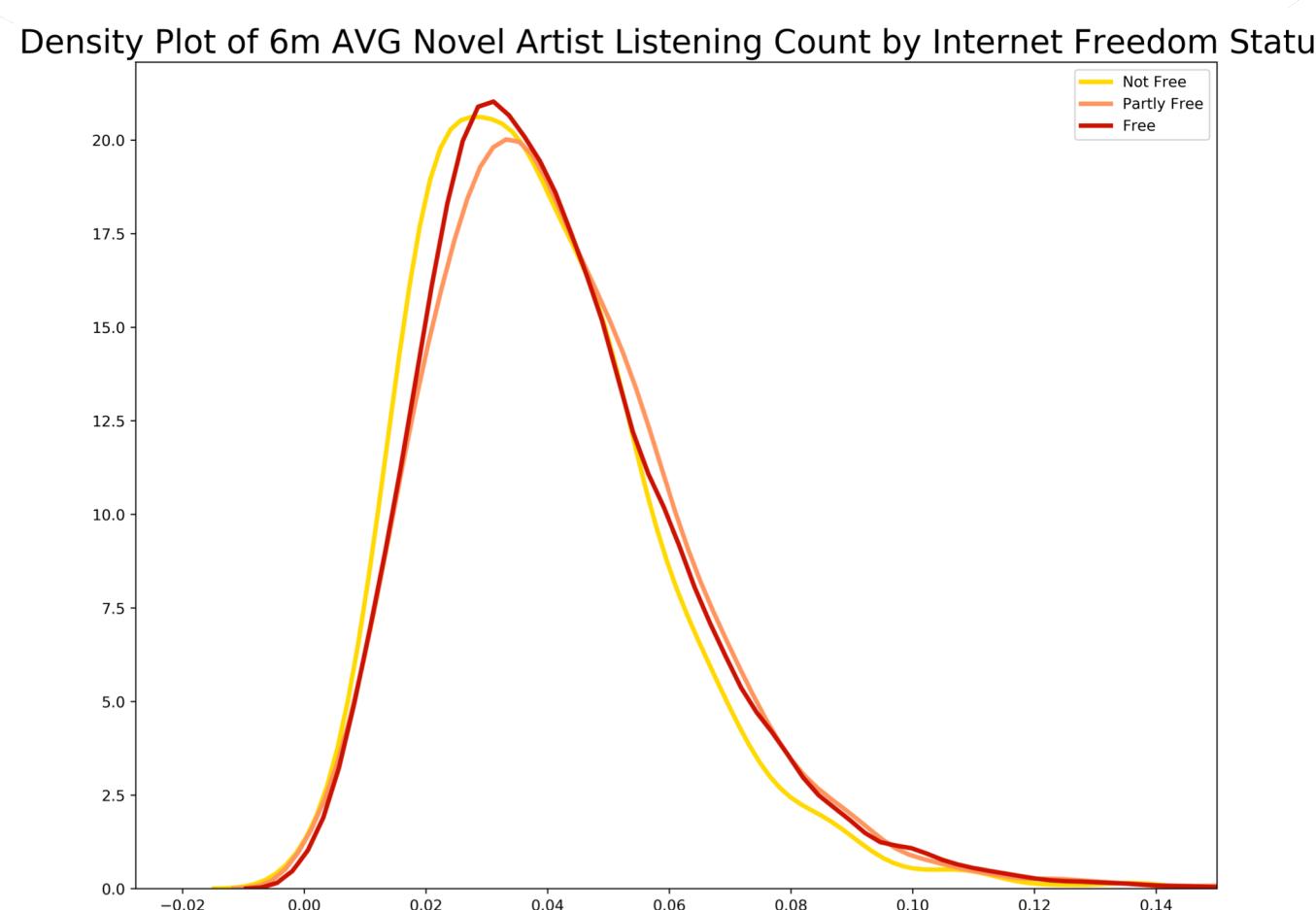
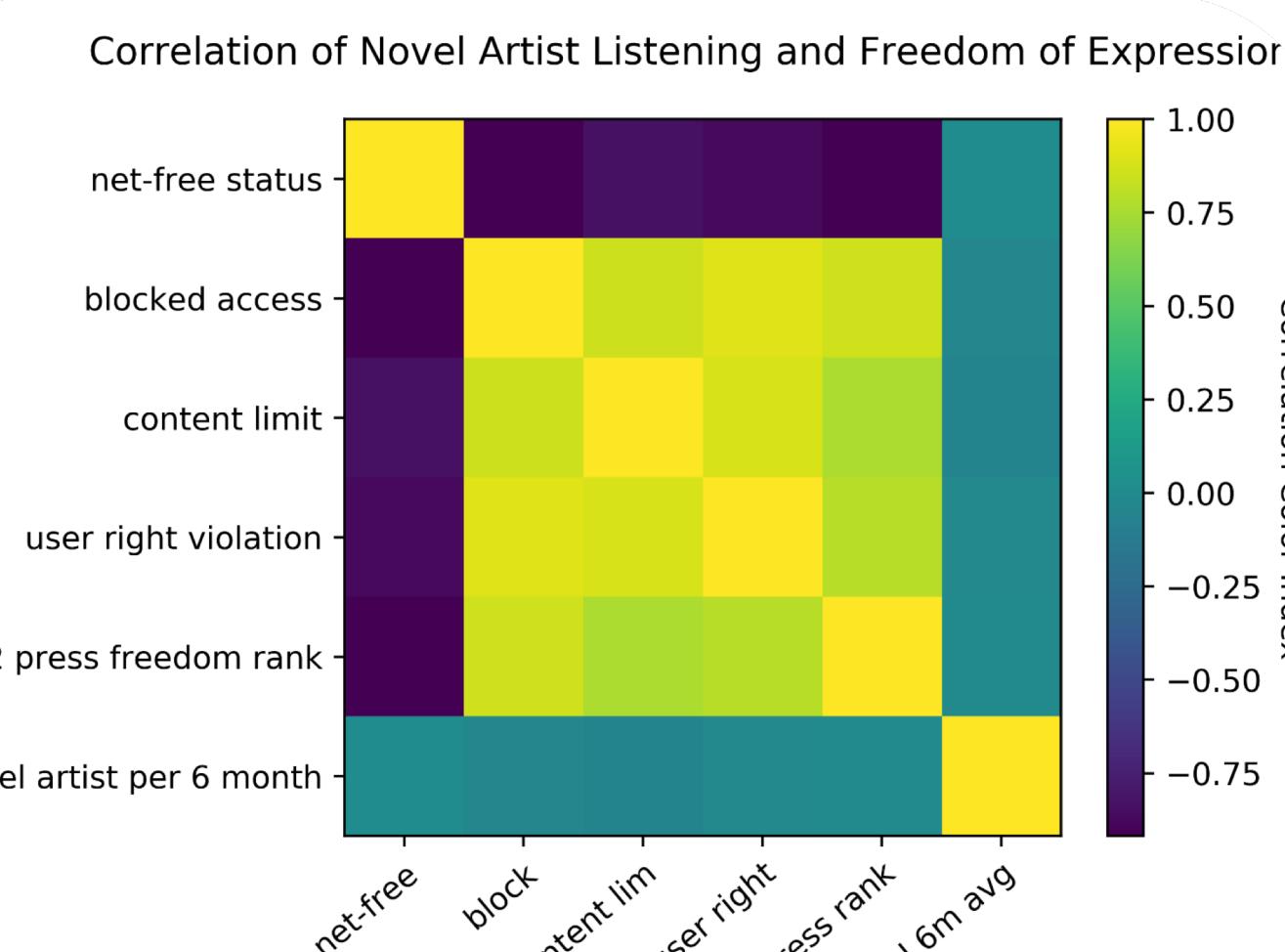
## Methodology



## Hypotheses & Results

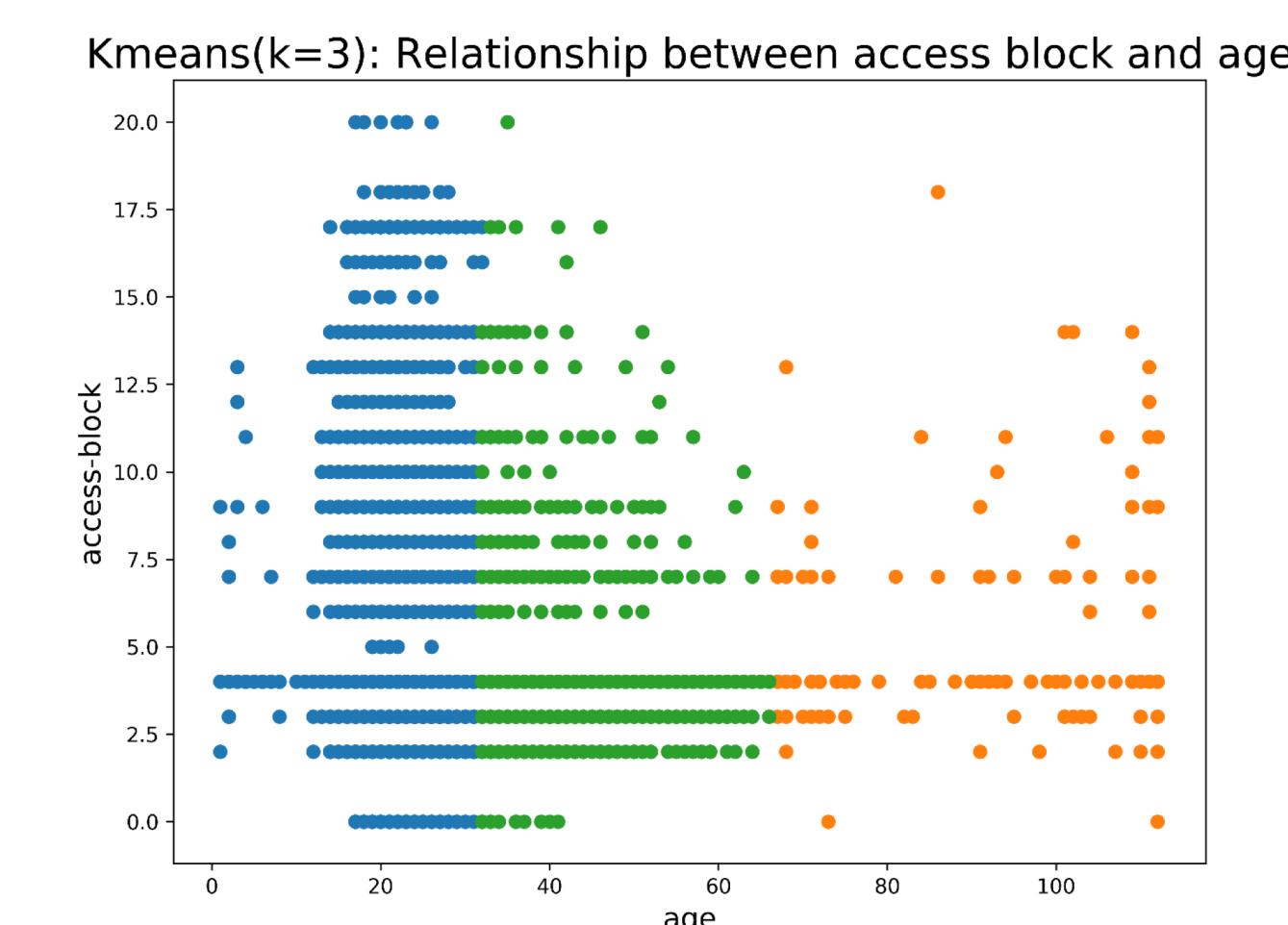
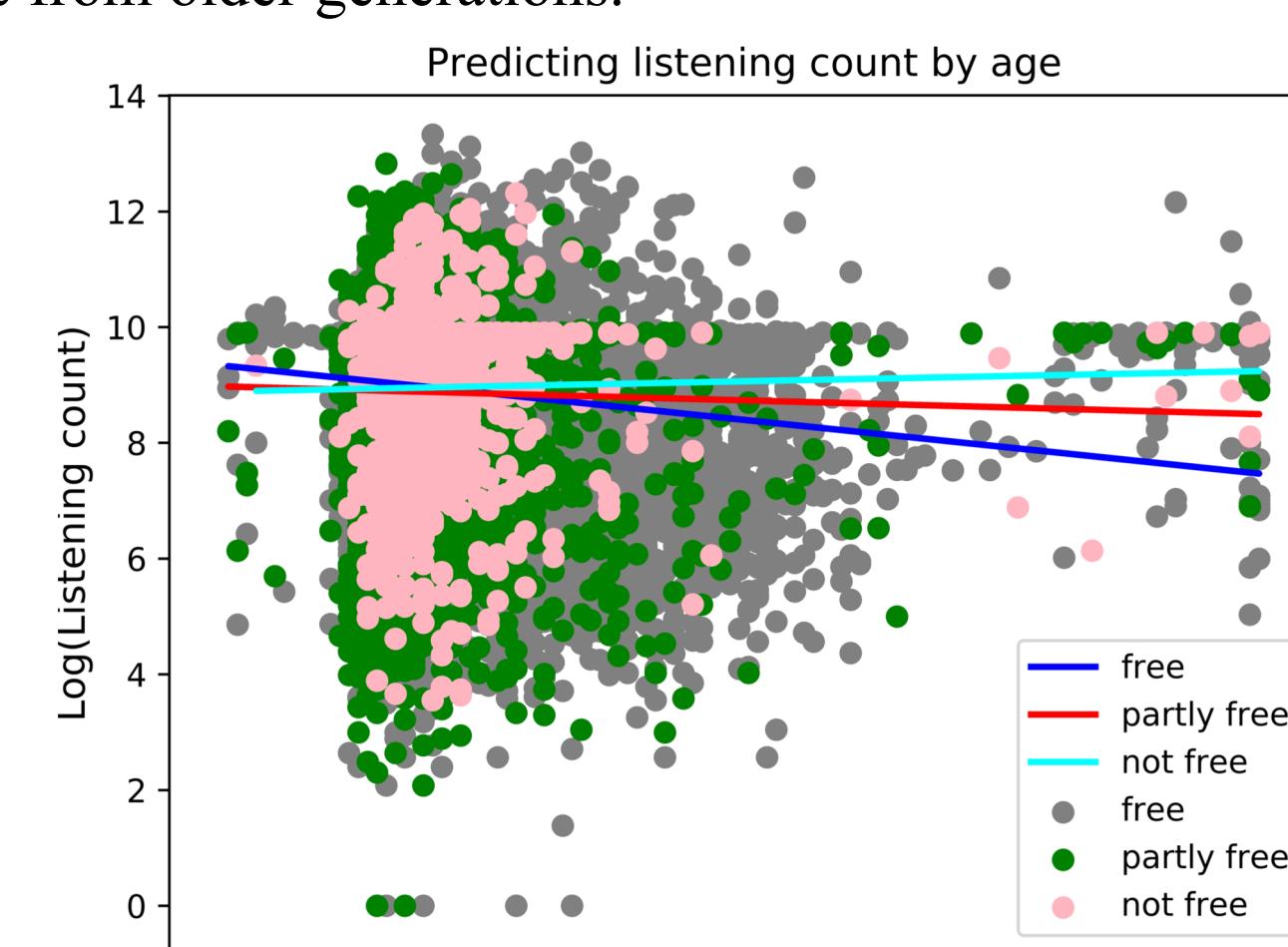
### Hypothesis I

**Relation between artist novelty and Internet freedom features:** We expect a correlation between the variety of musicians listened and the Internet freedom ranking of the country where the user is located. A country with higher ranking in freedom of Internet may have more medium to spread information about other (novel to the user) music.



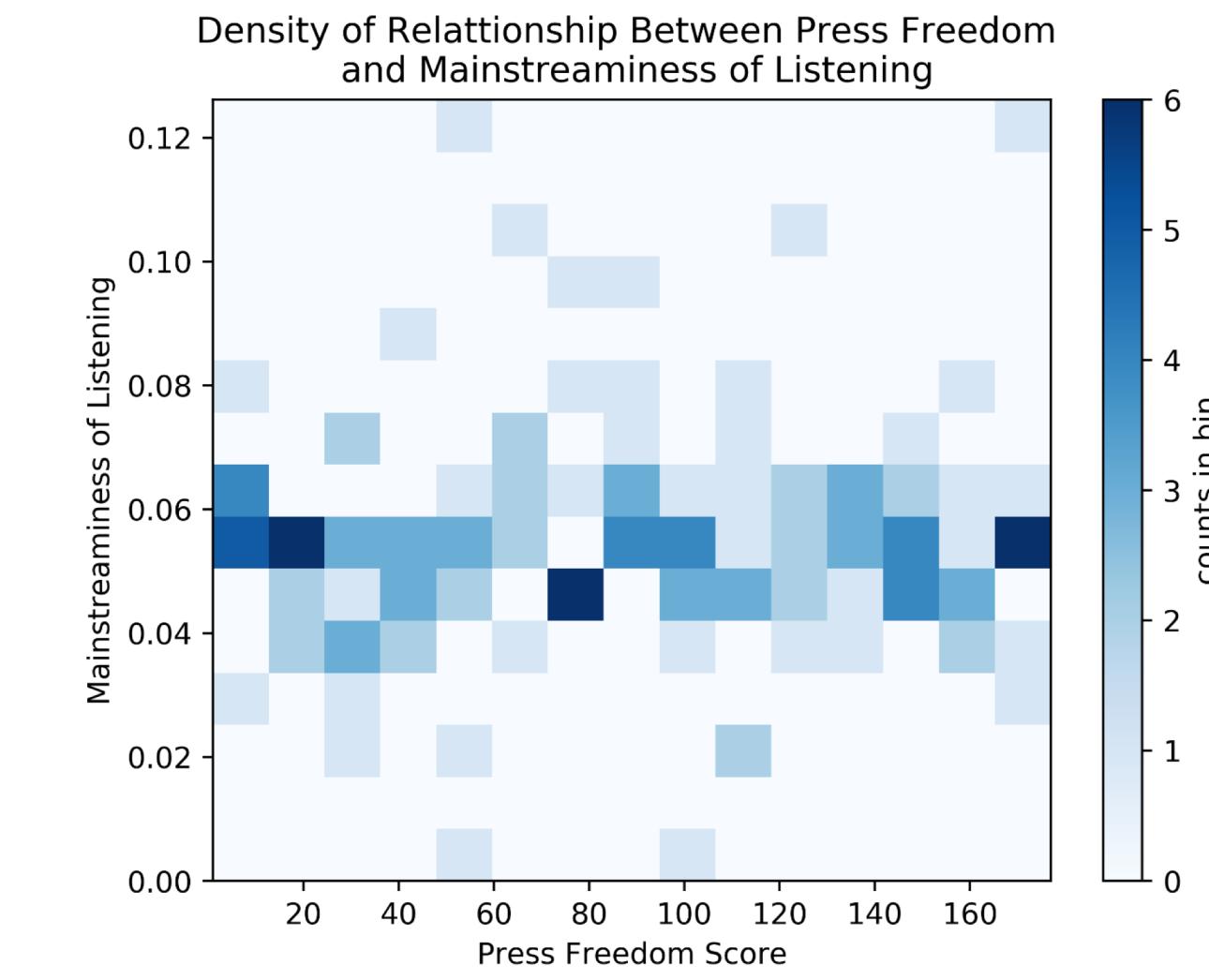
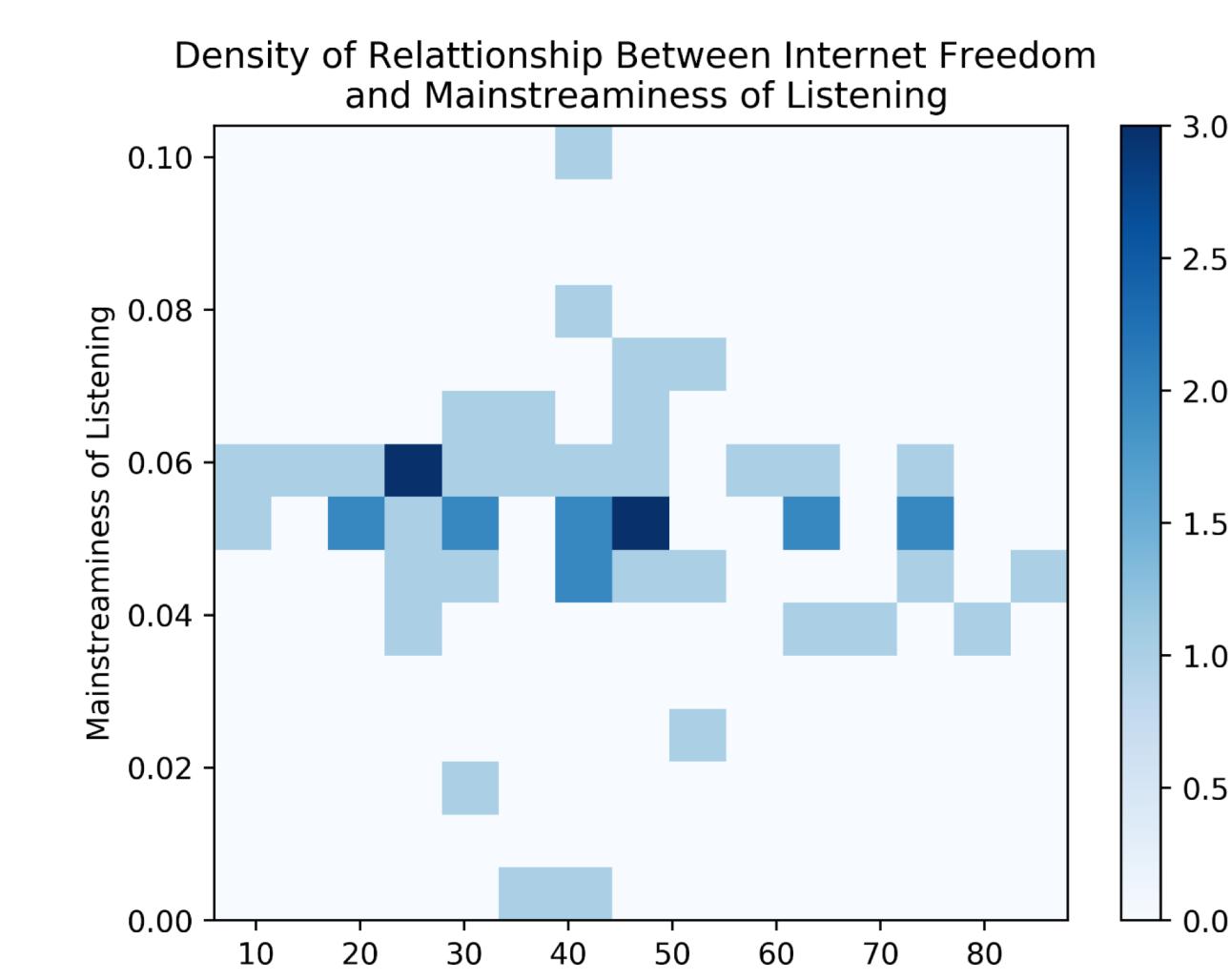
### Hypothesis II

**Relation between age /country Internet freedom features and the count of listening events:** We expect that younger generations (age 0 - 35) would find the obstacles put by the government less of a problem (ex: ease at using VPN) and have listening habits that are similar to the ones in countries. This also means that we expect the opposite from older generations.



### Hypothesis III

**Relation between mainstreamness of listening and freedom of expression features:** We expect a correlation between the number of musicians listened and the Internet freedom ranking of the country where the user is located. The motivation behind this is similar to the hypothesis above.



## Conclusions



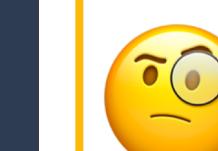
Unfortunately, no null hypotheses can be ruled out.

**Hypothesis I:** The average novel artist listening count in 6 months has no correlation with any of the freedom of expression features. The means and the variances of the distributions for each internet freedom are plotted as well, but they are extremely similar, showing no correlation, either.

**Hypothesis II:** The correlation between number of listening events and age showed slightly difference between countries with different freedom. The number of access block has no relationship with age of citizens, indicating the young and the old are influenced by censorship in same content.

**Hypothesis III:** Our results showed a relatively constant distribution of listening habits, independent of censorship levels. The source of this uncertainty is the lack of variation along the axis of country freedom.

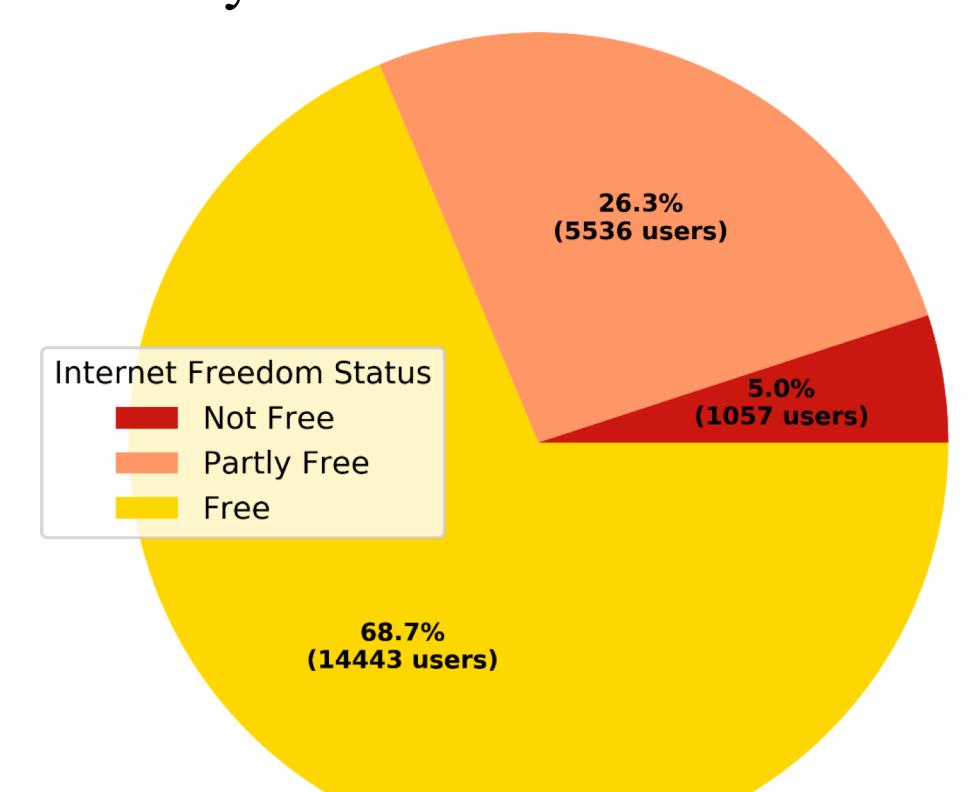
**Conclusions:** Therefore, since we couldn't refute any of our null hypotheses, our analysis shows there isn't a significant relation between a country's freedom expression and citizen listening habits.



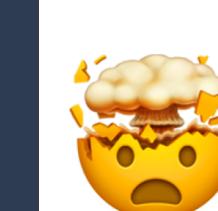
### Reasons & Limitations

- Lack of users from "NOT FREE" countries
- Year mismatch in datasets

Variety of Country Internet Freedom Status in User Dataset

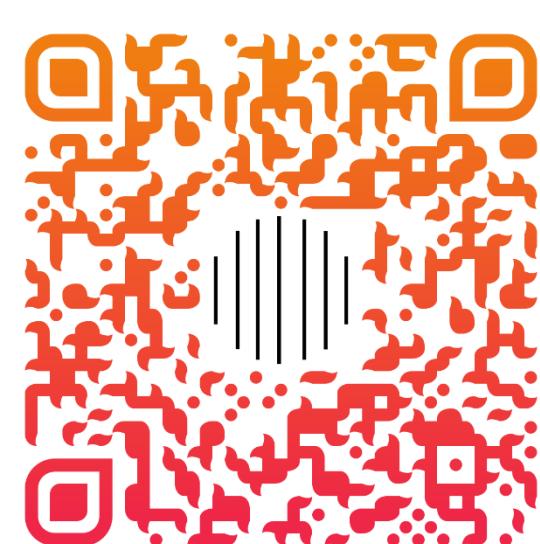


## Further Information



I. We have tried to use Spotify. We want to predict music taste with freedom expression. However, the data from Spotify didn't specify the user country information.

II. We have tried to predict PISA scores of a country with respect to freedom of expression features.



## References

- The LFM-1b Dataset for Music Retrieval and Recommendation Schedl, M. Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR 2016), New York, USA, April 2016.
- Internet freedom (<https://freedomhouse.org/report/freedom-press/freedom-press-2017>)
- Press freedom ([https://tcdatadata360.worldbank.org/indicators/h3f86901f?country=BRA&indicator=32416&viz=line\\_chart&years=2001,2015](https://tcdatadata360.worldbank.org/indicators/h3f86901f?country=BRA&indicator=32416&viz=line_chart&years=2001,2015))