



*Semester project*



27 club: Do musicians who die young  
become more famous?



Student: Sofia Kypraiou  
Supervisor: Tiziano Piccardi  
dlab - Spring Semester 2020



dlab

# 27 Club: Fame and popularity








# Research Questions

- ★ RQ1: Does the 27 club exist?
- ★ RQ2: What is the impact of dying young to long-term popularity?

# Dataset

- ★ **Wikipedia** for personal and work details of a person
  - 227,205 people - 15,734 musicians
- ★ **Pageviews from the English Wikipedia** for popularity
  - From 2017 - 2020

# Dataset

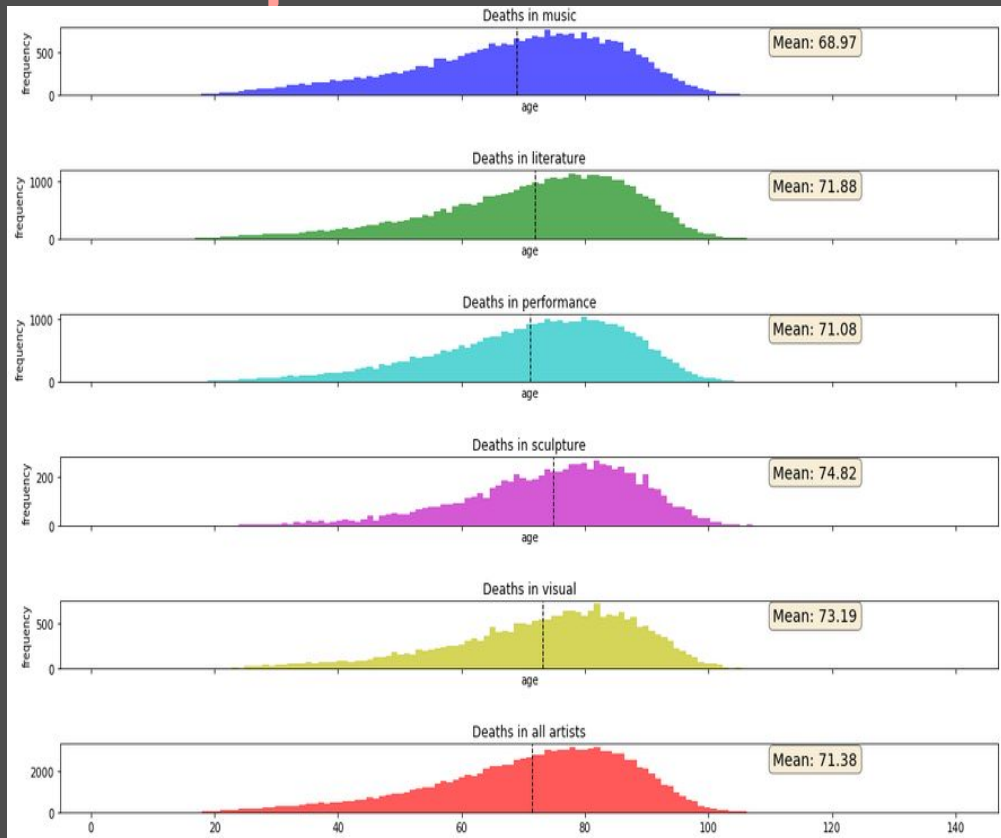
- ★ **Artist:** A person with > 50% artistic occupations:
  -  Music
  -  Visual
  -  Performance
  -  Literature
  -  Sculpture

# RQ1: Does the 27 club exist?

- ★ Is there any difference in life span based on different occupations? (Artist vs non-artist)
- ★ Is there any difference in mortality rate at different ages? Do artists tend to die more in their 20s compared to other people?




# RQ1: Lifespan on occupations

🎵 have the lowest mean age of death among all the other artists




# RQ1: Balancing the dataset

Exact matching on:

	Birth info	<b>Year of birth, country of birth</b>
	Personal info	<b>Gender, citizenship</b>
	Death info	<b>Country of death</b>

Different on:

	Work info	<b>Occupation</b>
---	-----------	-------------------



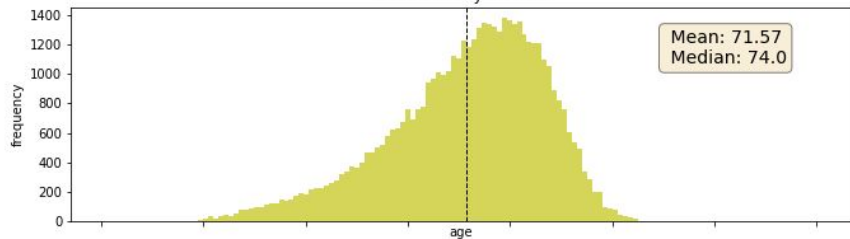
# RQ1: Lifespan on occupations

Artists 🎵🎨 vs Non-artists 🧑

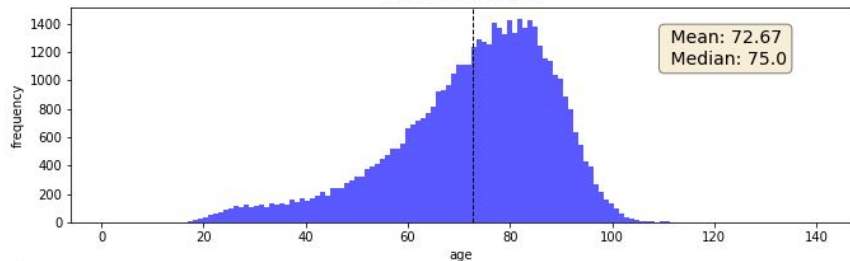
🎵 vs 🧑

Artists vs Non-artists (48303 pairs)

Deaths in any art field



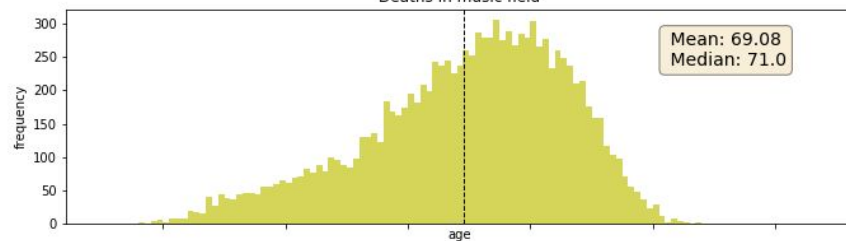
Deaths in None field



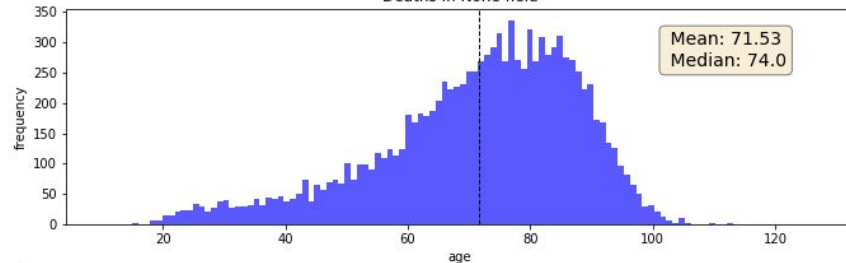
p-value=0.00000  
Reject H0: Sample distributions are not equal

Musicians vs people (11285 pairs)

Deaths in music field

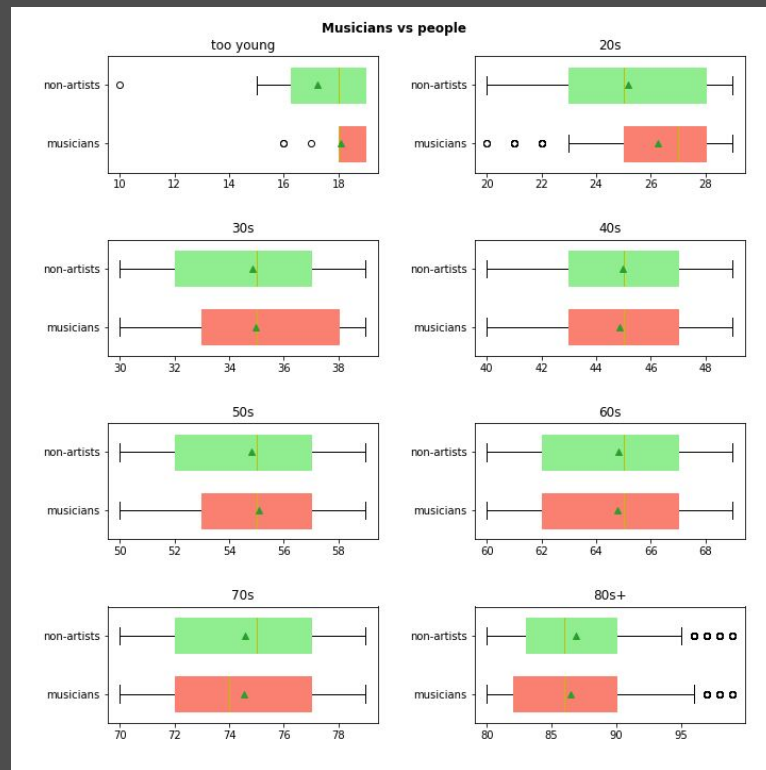
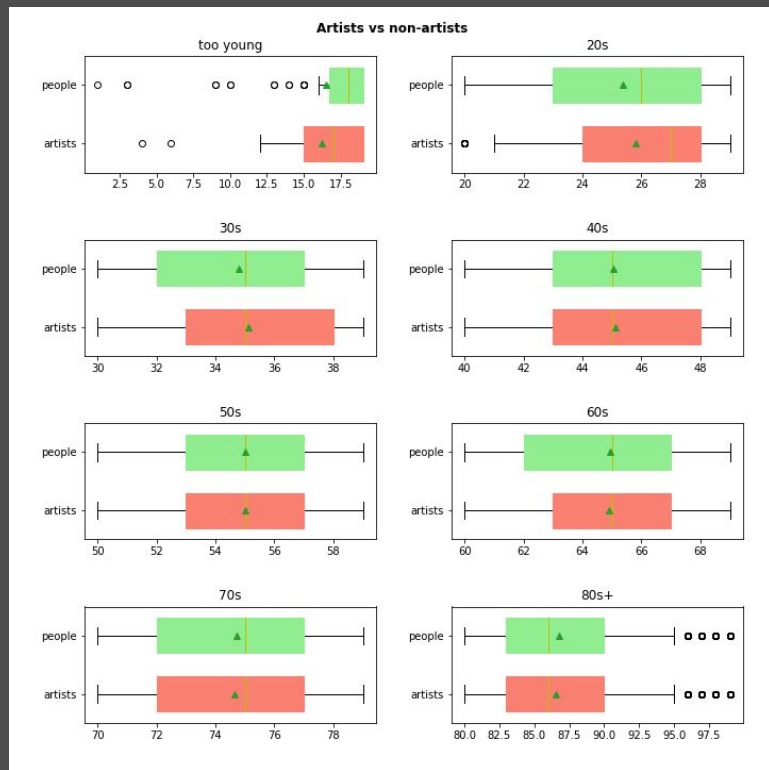


Deaths in None field

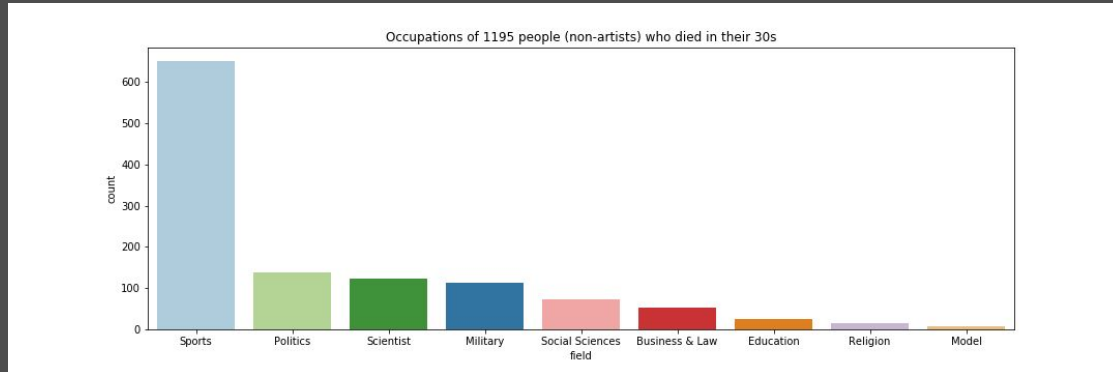
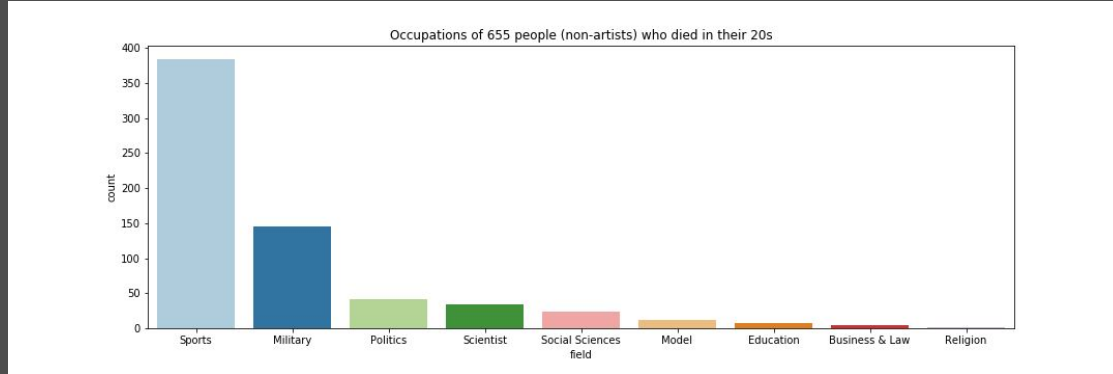


p-value=0.00000  
Reject H0: Sample distributions are not equal

# RQ1: Mortality rate at different ages



# RQ1: Occupations of non-artists



# RQ2: Do musicians who die young become more famous?

- ★ Investigate the causal effect of dying young on the long term popularity.

# RQ2: Exact matching



Emma

👤: 1990



Nora

👤: 2010



Same:

- Year of birth
- Country of birth
- Country of death
- Citizenship
- Gender
- Genre
- Awards received



Nora died 20 years after Emma

# RQ2: Balanced Pairs

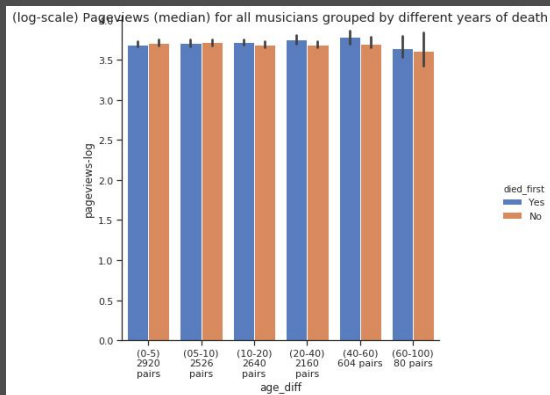
- ★ Analysis with:
  - All data: 24360 pairs
  - Musicians dead before 1950: 395 pairs
  - Musicians dead after 1950: 21982 pairs
  
- ★ Maximal matching on the balanced pairs,  
Grouped on how many years apart they died

## RQ2: Balanced Pairs (matched date of death)

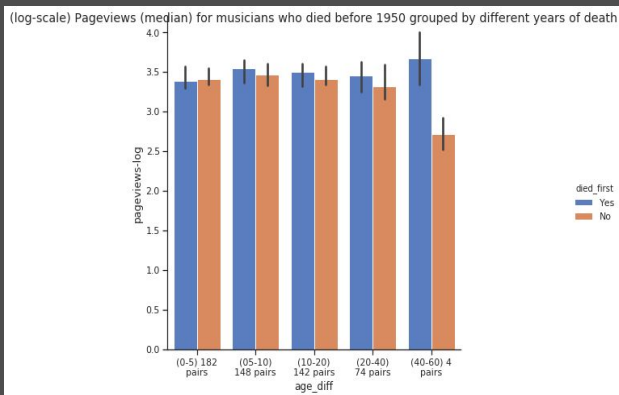
- ★ Analysis with:
  - All data: 22978 pairs
  - Musicians dead before 1950: 563 pairs
  - Musicians dead after 1950: 22415 pairs
  
- ★ Maximal matching on the balanced pairs,  
Grouped on how many years apart they died

# RQ2: Popularity - pageviews

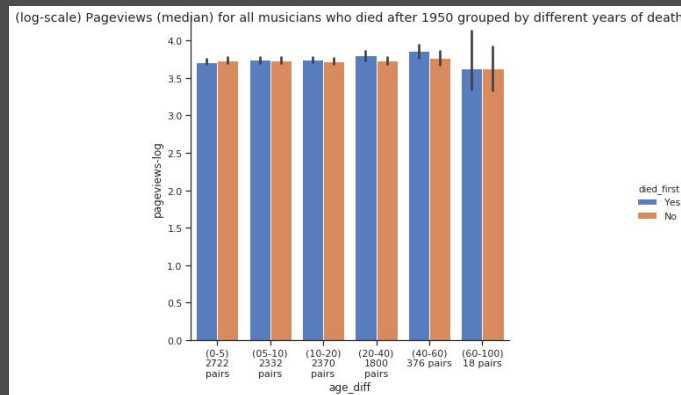
All 🎵



🎵 🦴 < 1950



🎵 🦴 > 1950



★ Significant for:  
(20-40), (40-60)

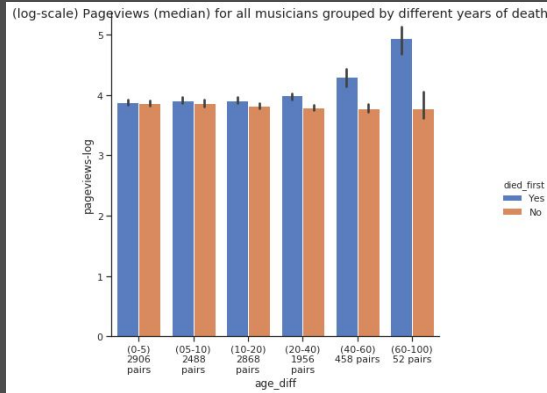
★ Significant for:  
nothing!

★ Significant for:  
(20-40), (40-60)

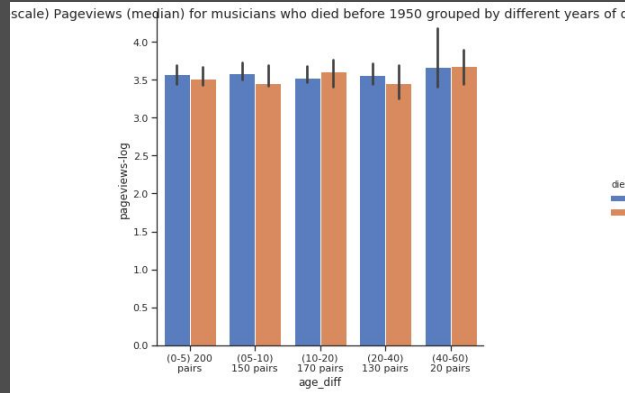


# RQ2: Popularity – pageviews (matched date of death)

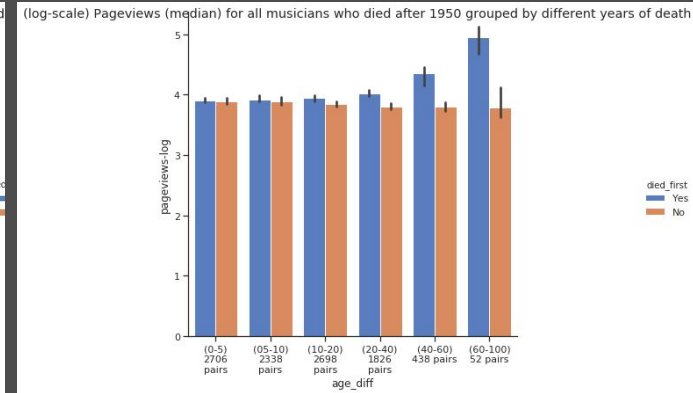
All 🎵



🎵 🦴 < 1950



🎵 🦴 > 1950



★ Significant for:  
(10-20), (20-40),  
(40-60), (60-100)

★ Significant for:  
nothing!

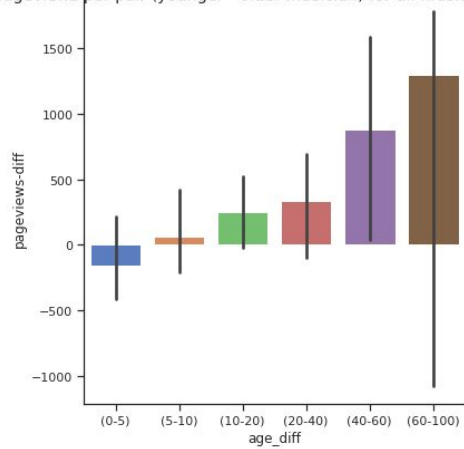
★ Significant for:  
(10-20), (20-40),  
(40-60), (60-100)

# RQ2: Popularity - pageviews

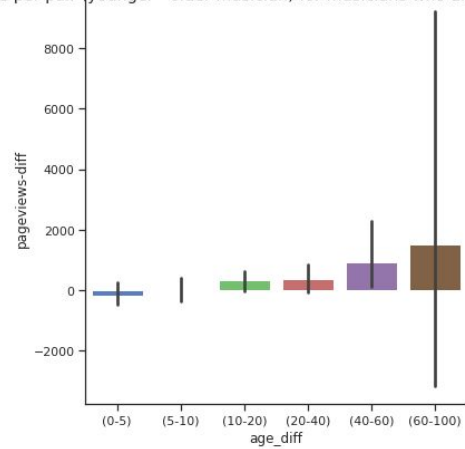
All 🎵

🎵 🦴 > 1950

Median difference in #pageviews per pair (younger - older musician) for all musicians per age\_diff bracket



Median difference in #pageviews per pair (younger - older musician) for musicians who died after 1950 per age\_diff bracket

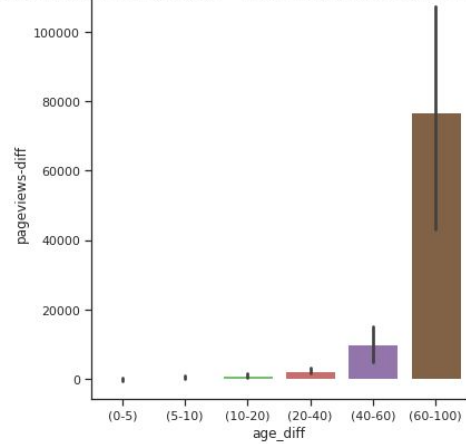


# RQ2: Popularity - pageviews (matched date of death)

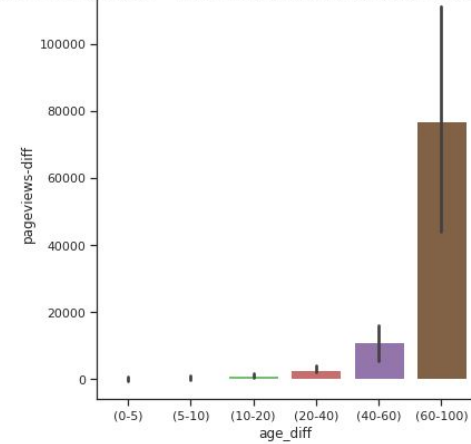
All 🎵

🎵 🦴 > 1950

Median difference in #pageviews per pair (younger - older musician) for all musicians per age\_diff bracket

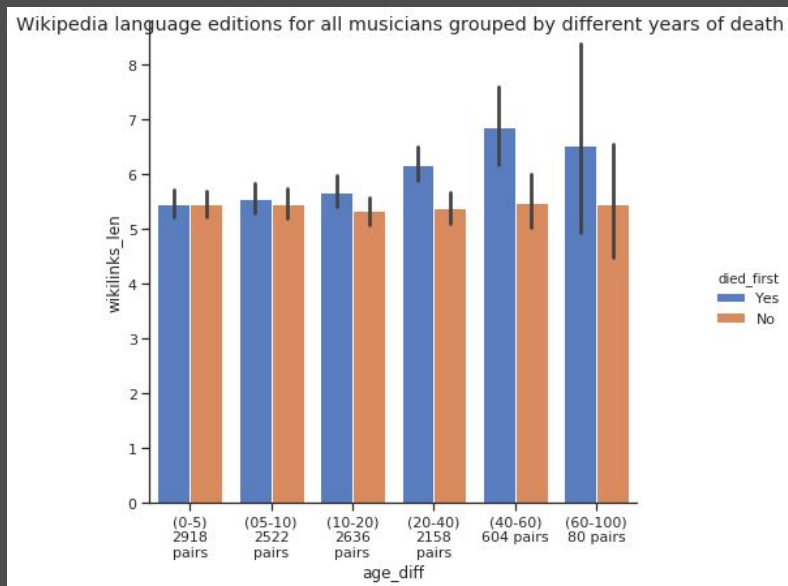


Median difference in #pageviews per pair (younger - older musician) for musicians who died after 1950 per age\_diff bracket

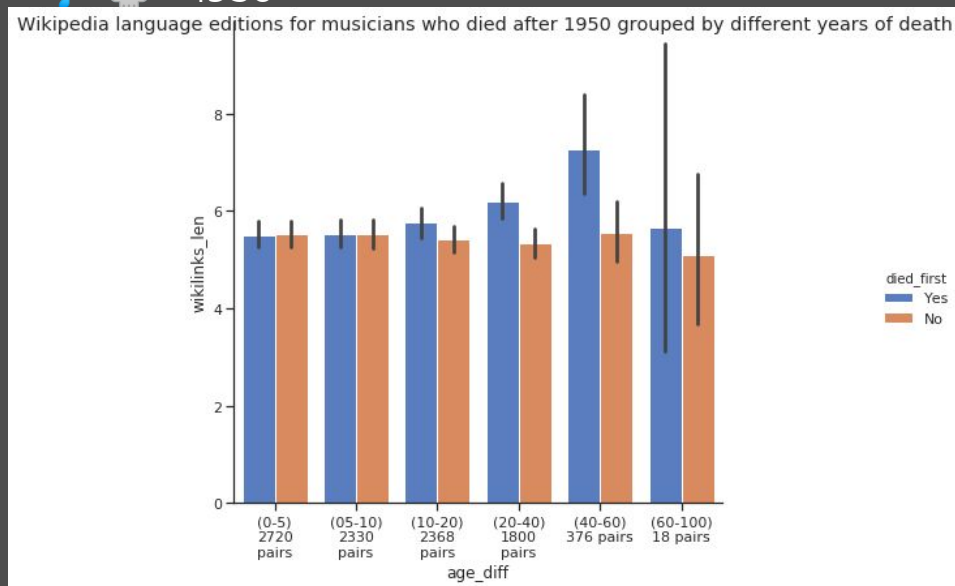


# RQ2: Popularity - language versions

All 🎵



🎵 🦴 > 1950

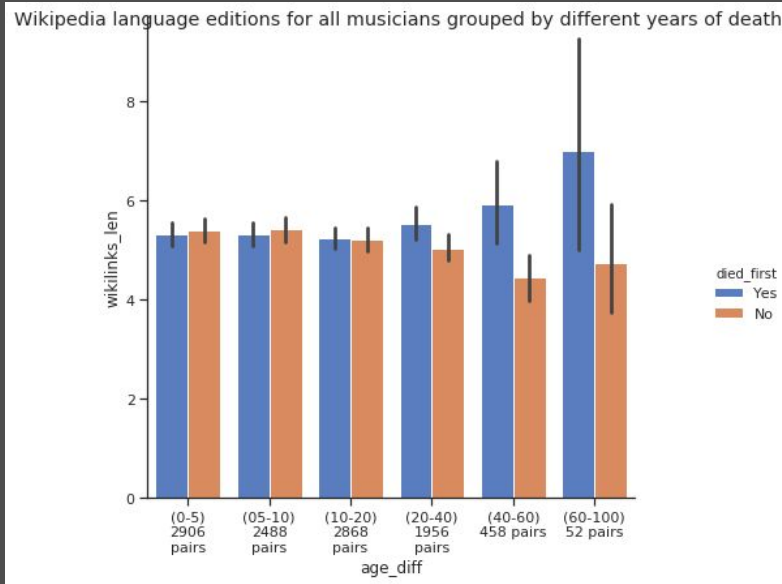


★ Significant for:  
(20-40), (40-60)

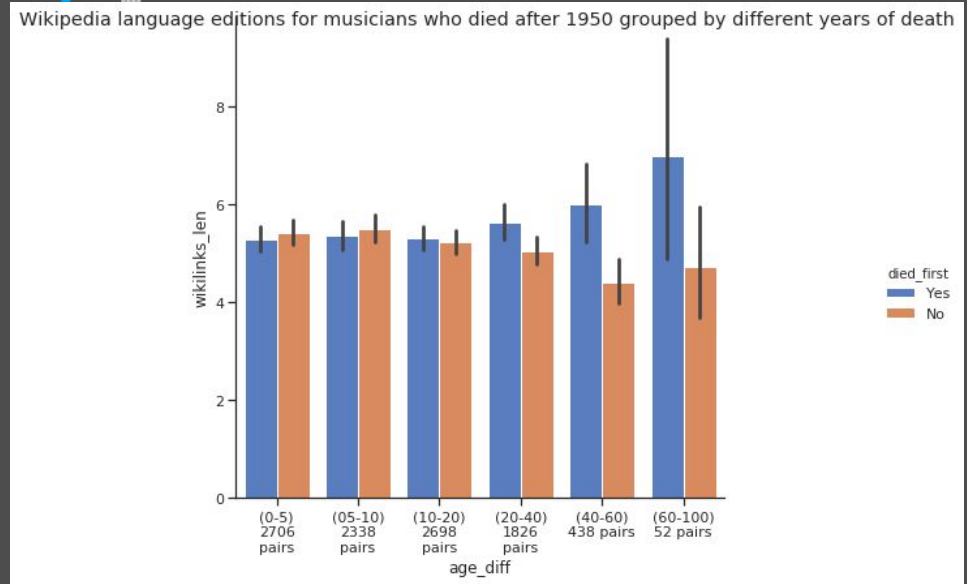
★ Significant for:  
(20-40), (40-60)

# RQ2: Popularity - language versions (matched date of death)

All 🎵



🎵 🦴 > 1950



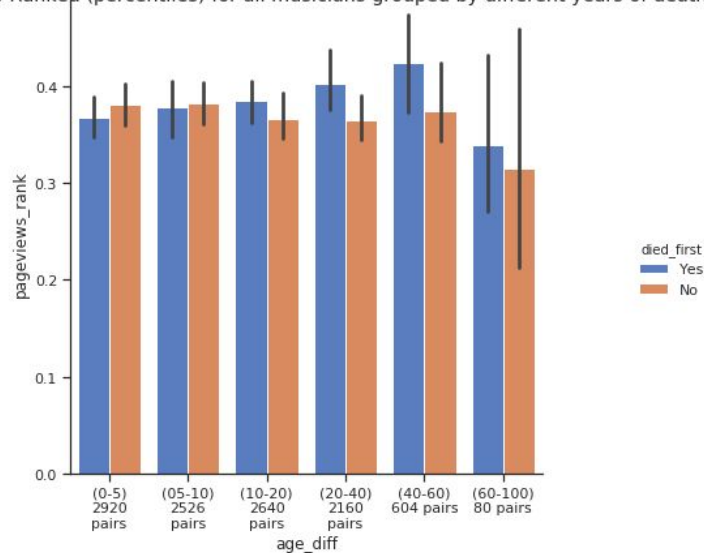
★ Significant for:  
(20-40), (40-60)

★ Significant for:  
(20-40), (40-60)

# RQ2: Popularity - percentiles

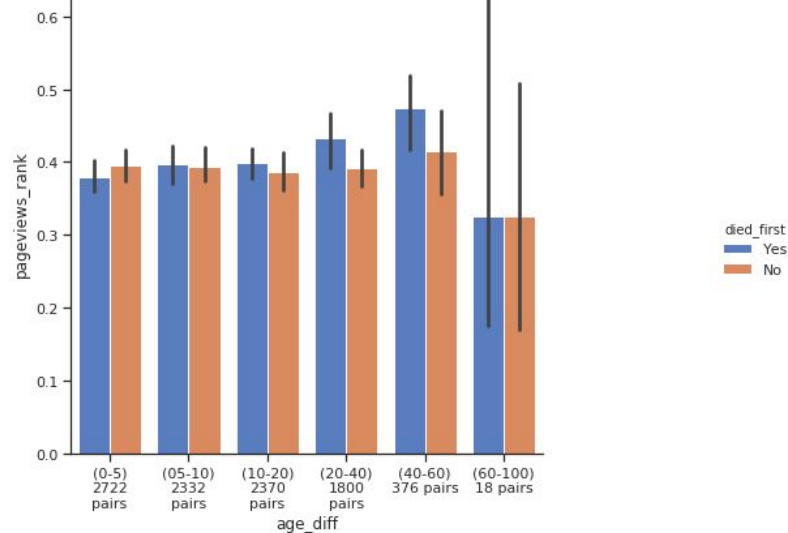
All 🎵

Pageviews-Ranked (percentiles) for all musicians grouped by different years of death



🎵 🦴 > 1950

Pageviews-Ranked (percentiles) for musicians who died after 1950 grouped by different years of death



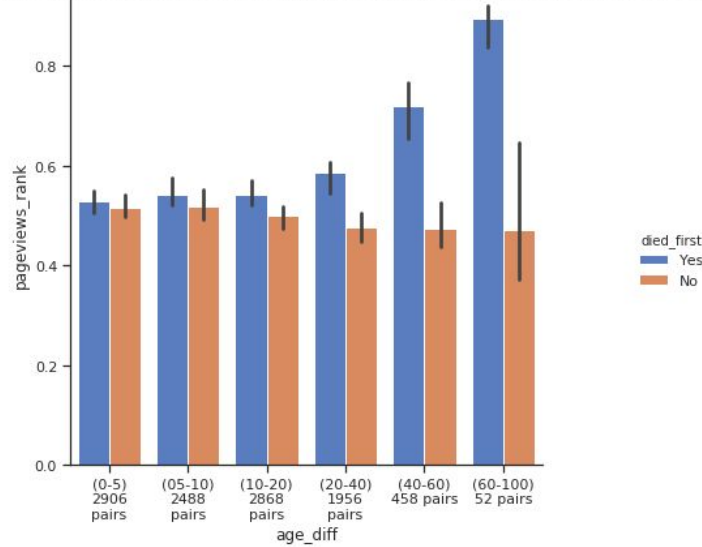
★ Significant for:  
(20-40)

★ Significant for:  
nothing

# RQ2: Popularity - percentiles (matched date of death)

All 🎵

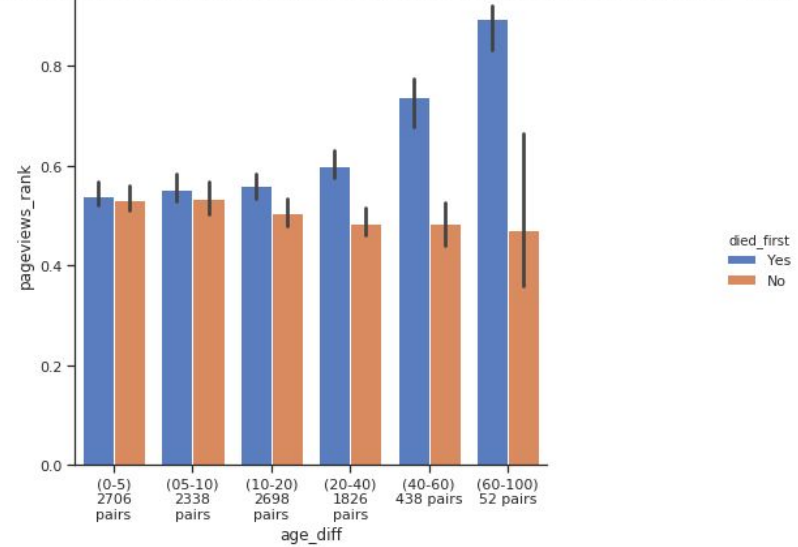
Pageviews-Ranked (percentiles) for all musicians grouped by different years of death



★ Significant for:  
(10-20),(20-40), (40-60), (60-100)

🎵 🦴 > 1950

Pageviews-Ranked (percentiles) for musicians who died after 1950 grouped by different years of death



★ Significant for:  
(10-20),(20-40), (40-60), (60-100)

# Conclusion

- ★ Club 27 is still a myth!
- ★ Evidence that musicians who died younger are more popular than musicians who died later



# Limitations

- ★ Popularity at the age of death
- ★ Wikipedia's recency bias
  - If Nora died in 2010 where Wikipedia was already created, then it's more likely to have an article than Emma who died in 1990
  - Older people are more noteworthy?

# Limitations

- ★ Validate quality of the matched pairs
- ★ Define the uncertainty of our results
- ★ Argue for the validity of our modeling assumptions.

*Thank you!*