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

Sofia Kypraiou
École Polytechnique Fédérale de Lausanne (EPFL)
sofia.kypraiou@epfl.ch

Tiziano Piccardi
École Polytechnique Fédérale de Lausanne (EPFL)
tiziano.piccardi@epfl.ch

Robert West
École Polytechnique Fédérale de Lausanne (EPFL)
robert.west@epfl.ch

ABSTRACT

Famous people and their deaths have always fascinated pop culture. Especially when they die at a young age, it creates a myth around them, with the most famous being the '27 club', an age when a lot of talented musicians at the height of their fame have died. Our approach is to verify if there is enough statistical evidence that the 27 club exists. We also explore the effect that dying at a younger age has to a musician's reputation. We use data from Wikipedia, the largest encyclopedia ever built, to collect the information about the musicians and answerer these questions. We design an observational study to explore the effects of dying young in their long-term popularity. We find that statistically there is no evidence that the 27 club is real. However, we find evidence that musicians who died younger are more popular than musicians who died 40 and 60 years later than them.

KEYWORDS

27 club, observation study

1 INTRODUCTION

Famous musicians who died suddenly have been exciting to study and have triggered media interest. The 27 club - a list of popular musicians and artists in general - includes almost 75 people [1]. Among them are Jimi Hendrix, Janis Joplin, and Jim Morrison who all died at the age of 27 between 1969 and 1971. However, it wasn't Kurt Cobain's 1994 death, at age 27, that the existence of that club began to catch on in public perception. The later death of Amy Winehouse in 2011 reopened the public talk on the existence of this club.

This myth has risen many questions and still exists as a cultural phenomenon. Why 27? Is something special about this age or the 27 club exists by chance, where people focus on results that support their hypothesis? Could it be that dying young increased their fame and popularity? Or is it true only in a few cases?

Several approaches to verifying the myth have been implemented in the past (Sec. 2). One of the limitations of previously proposed solutions is the amount of data they used to answer these questions.

Nowadays, we have access to a much bigger amount of data. Using Wikipedia, we find information about the life and death of many famous - and not so famous - musicians and people. Besides the information, we can verify someone's current popularity, by looking at the pageviews of the biography in Wikipedia of that person.

The main idea to our approach is to compare similar musicians where the one died at some point in time and the other musician died several years later. We find pairs of people who were as similar as possible when they lived, had almost the same level of popularity when the first one of the pair died, while the other musician had more years on being productive.

To address the above problems, we ask the following research questions:

- **RQ1** Does 27 Club exist? (Sec. 4)
- **RQ2** What is the impact of the age of death on popularity? (Sec. 5)

In order to answer these questions, we collect a large dataset including all the people in Wikidata. By analyzing this dataset [1] we make the following main contributions:

- We find that there is no statistical evidence that the 27 club exists. We create a balanced dataset to compare musicians with non-artists and musicians with other artist occupations. We find that non-artists die younger than musicians in their 20s and 30s.
- We show that musicians who died younger are more popular than musicians than the ones who died 20 and 40 years later. More specifically, when we focus on musicians who died after 1950 the effects are even stronger

We can conclude that although the 27 club doesn't exist, there is some evidence indicating that dying younger increase popularity, however further analysis is needed to verify that claim.

2 RELATED WORK

There have been various statistical studies that prove that musicians are not more likely to die in 27 than any other age.

In their study "Is 27 really a dangerous age for famous musicians?" [4] the researches use a sample of 1046 musicians to explore the myth. They select musicians (solo artists and band members) who had a number one album in the UK between 1956 and 2007. They find that although the sampled musicians faced a two- to three-fold increase in the risk of death in their 20s and 30s compared to the general UK population, this was not limited to the age of 27. They conclude that fame may increase the risk of death among musicians, but this risk is not limited to age 27.

Dianna Theadora Kenny, professor in the University of Sidney, also provided statistical evidence that popular musicians are not more likely to die at the age of 27 [3]. The analysis was performed in a population of dead musicians spanning seven decades from 1950 to 2010 ($n = 11,054$). They found that age 56 had the highest frequency of deaths and that the actual number of deaths at 27 is statistically significantly less than the number that would need to die at 27 if the 27 Club hypothesis were correct based on numbers alone.

A 2013 study tries to go beyond the music industry and focuses on fame and death "Study: People Who Are Famous and Successful Have Shorter Lives" [2]. The researches analyze 1000 consecutive

[1] Notebooks with code at https://github.com/epfl-dlab/artist-popularity

obituaries published in NYT over the period 2009–11 in terms of gender, occupation and terminal disease, as attributed. "Success," by their measure, was defined as having lived a life that merited an obit in the paper of record. People who were both successful and famous died earliest. The average age at death of performers and athletes, 77.2 years was younger than those who had achieved success in other fields.

3 DATA COLLECTION

To study the myth of Club 27 and the effect of death to popularity, we collected data from the Wikidata.

We collected data for all people from Wikidata accessed in February 2020. Since we are researching death and popularity, only dead people appear in the dataset. The dataset also contains people who were born after 1850, to include the year the first member of the club 27 was born [2].

Each person is then characterized by a set of features that corresponds to personal and work details. That includes information related to place and year of birth, place and year of death, citizenship, gender and occupation. In the case of artists, there is also the genre they belong, the instruments they might play and the awards they received.

To identify the artists among all the people, we built a dataset to match the occupation code to the art field. We categorize the artistic occupations in 5 main fields.

Literature the art of narrative (ex. writers, poets)
Visual the art of seeing (ex. painters, designers)
Sculpture the art of shapes and in 3-dimensions (ex. architects)
Performance the art of movement and motions (ex. actors, dancers, directors)
Music the art of audio (ex. musicians, singers)

For a person to be considered as an artist, she needs to have at least half of her occupations fall in any of the above artistic fields.

Also, to measure the popularity, we collected the pageviews from Wikipedia from 2017 to the beginning of 2020. To prevent the trends in the pageviews [5], as recently deceased people will have more pageviews, we removed from the dataset people who died in the past 5 years.

At the end of the data cleaning, the dataset contains $227,205$ people, of which $15,734$ are musicians (0.07%)

4 RQ1: DOES 27 CLUB EXIST

In this section, we perform a statistical investigation on the myth with data from Wikipedia. To answer the research question, we break it down to the following subquestions.

- 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? I.e. Do artists tend to die more in their 20s compared to other people?

[2] https://en.wikipedia.org/wiki/27_Club#Identified_members

4.1 Life span of different occupations

To see if musicians have a shorter life span than other artistic occupations, we start by exploring the mean age of death in the different artistic occupations.

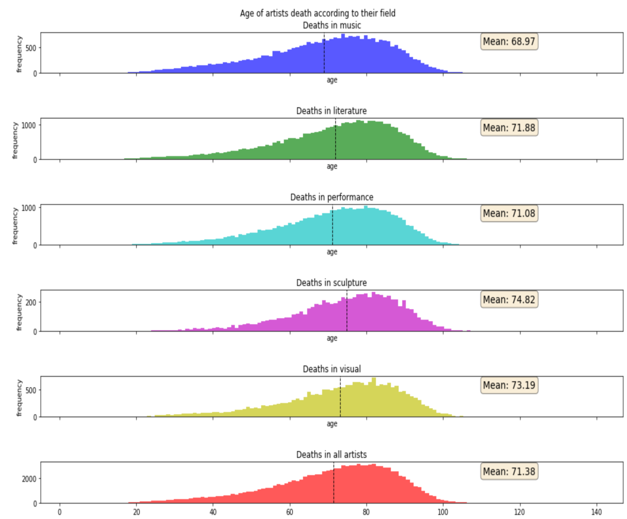


Figure 1: Histograms showing the distribution of the age of death for artists and for each artistic field

As seen in figure 1, the musicians have the lowest mean among all the other artistic fields.

This is a good indicator that musicians die younger, but we need to validate the results.

To compare the results, we need to balance the dataset. To address this we perform a matched observational study.

Extact matching We balance the dataset using exact matching. We compare people who have a different occupation (artistic or not) but have the same personal information. We control for the properties of *year* and *country of birth*, *citizenship*, *gender* and *country of death*.

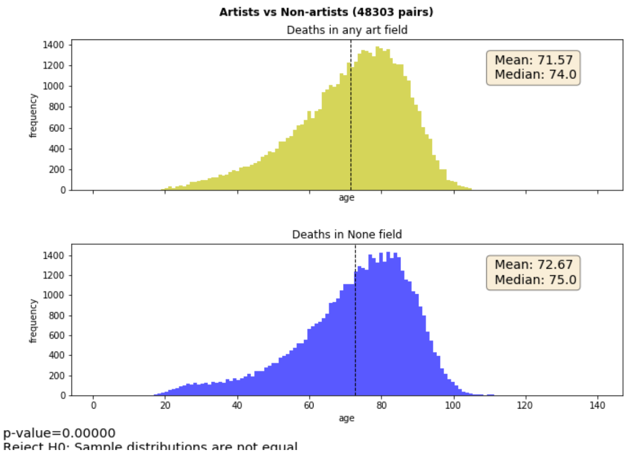
In total, we get 16061 matches between artists. Then we compare artists with non-artists, musicians with non-artists and musicians with the other artistic fields.

Results Figure 2 shows the histogram with the balanced dataset. In all the cases explored, musicians have a smaller mean and median age of death.

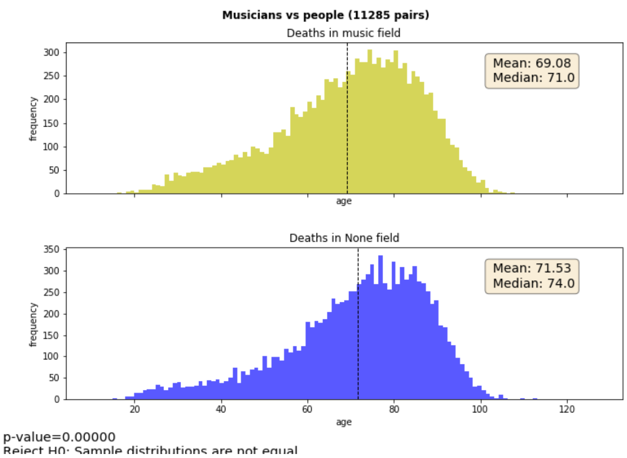
4.2 Mortality rate at different ages

To come closer to understanding if there is an impact, we proceed by stratifying the balanced groups by age (Figure 3). We test if the difference in the age of death is significant by using the Mann–Whitney U test.

We find that people with non-artistic occupations die younger in their 20s and 30s than artists or musicians ($p << 0.003$). This is opposite of the plots we saw before. To investigate why this happens we analyse the occupations of the non-artists. From Figure 4 we see that there are a lot of militaries, which explains why this group died younger.



(a)



(b)

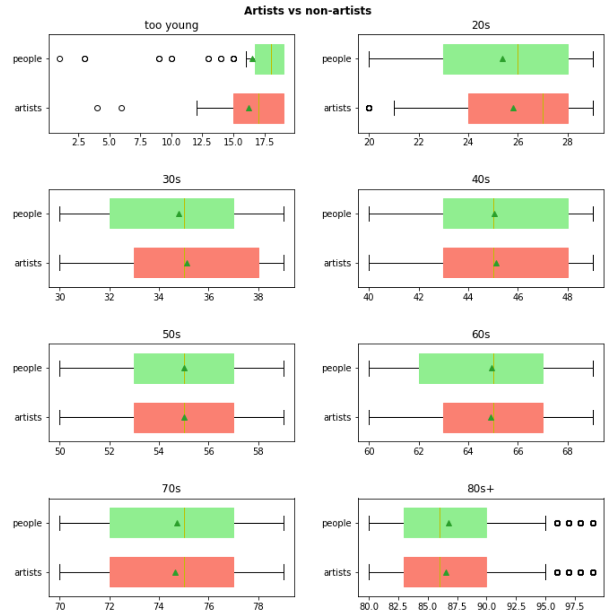
Figure 2: Histograms showing the distribution of the age of death for (a) artists and non-artists (b) musicians and non-artists.

5 RQ2: WHAT IS THE IMPACT OF THE AGE OF DEATH ON POPULARITY?

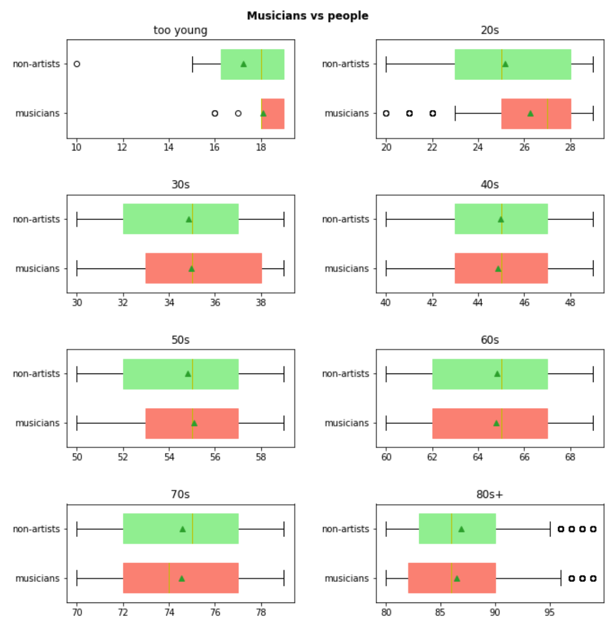
In this section, we proceed to our second research question, which investigates the causal effect of dying young on the long term popularity.

To understand the causal impact of how dying young on musician's popularity, we perform a matched observational study. The ideal goal would be to compare the popularity of musicians who lived the same life, but one member died while the other one continued to be productive and creative for more years. In this ideal goal, where the musicians would have the same success before death, an example would be to compare members of the same band.

Dataset. In addition to the previous features, we also add the awards received to measure the popularity before death. One important note is that we removed the awards that were received after the one person of the pair died. (maybe mention here that we removed the dataset)



(a)

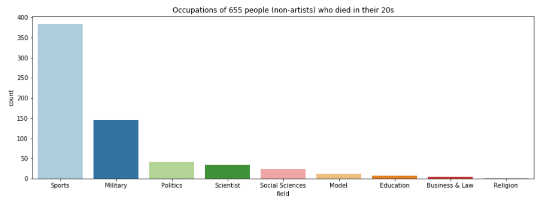


(b)

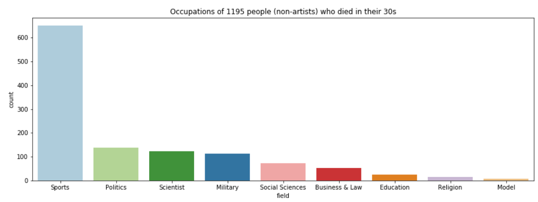
Figure 3: Box-plots with statification of the age of death for (a) artists and non-artists (b) musicians and non-artists.

In our setting, we define the pageviews as the treatment variable and all the other features as observed variables.

Exact Matching. We controlled for the properties *country* and *year of birth*, *citizenship*, *gender*, *genre* and *awards received.* We split the analysis into 3 categories: first with all the data that we have



(a)



(b)

Figure 4: Occupations for non-artists who died in their (a) 20s (b) 30s.

(people born after 1850 and died before 2014) and then into two subcategories, with people who lived between 1850 and 1950, and people who lived between 1950 and 2014. In the first case, we have 1085 groups of people with the same properties, 142 groups for people who died before 1950 and 943 groups for the others. The groups don't add up because some of the matches in the previous plots included pairs where the one died before 1950 and the other after.

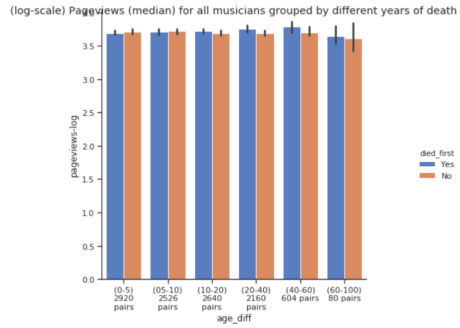
Balanced Pairs. For each of these groups, we calculate how many years more a person of the group lived. That way, we get 24360 pairs for all of the data. To find the balanced pairs, we use the maximal matching algorithm. Finally, we use these pairs and we group them now by the years more than the one person lived. For each pair, we have the person who died first and the person who died x years later, grouped by x.

Results. Figure 5 visualizes the results of the balanced pairs using pageviews as the long-term popularity metric. We observe a pattern that the person who died first has more page views than the person who died x years later.

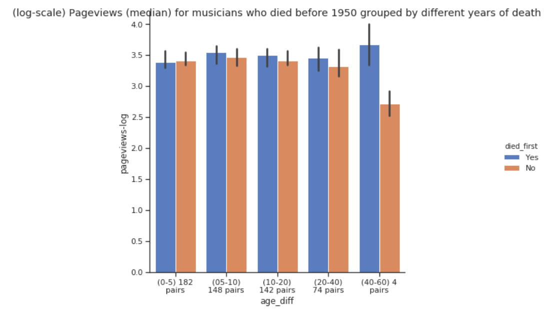
To verify that there is a significant statistical difference we use the Wilcoxon signed-rank. The popularity is statistically significant with $p < 0.05$ for the age groups (20-40) and (40-60) when using all the musicians (Figure 5a). For 5b we don't find any statistical evidence. Lastly, for musicians who died after 1950 (Figure 5c) we find that musicians who died first in the age groups (20-40) and (40-60) are more popular than their match.

The above results can also be seen in the complementary figure 6. We take the difference in the number of pageviews per pair (between the musician who died first and the one who died x years later) and then compute the median of these differences.

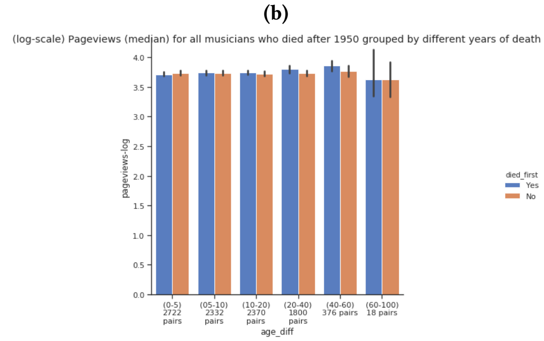
The same analysis 7 is repeated with the number of Wikipedia language editions each biography has. Using all the samples 7a and for musicians who died after 1950 7c we get a significant difference for the age groups (20-40) and (40-60). We don't find anything



(a)



(b)

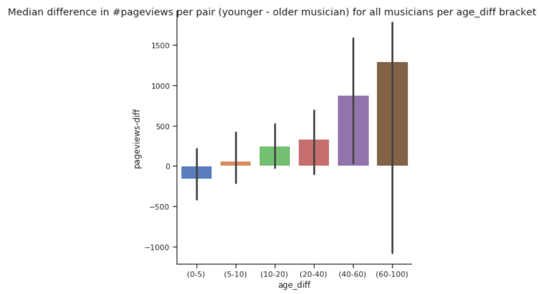


(c)

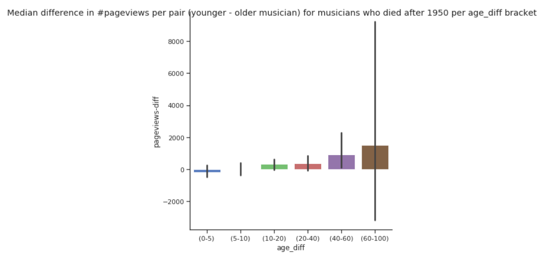
Figure 5: Median pageviews (in log scale) grouped by difference in year of death for (a) all the dataset (b) musicians who died before 1950 (c) musicians who died after 1950.

significant for musicians who died before 1950 7b (as this sample is much smaller).

Another metric for popularity is using the relative rank instead of raw pageviews numbers 8. The most popular musician is assigned to 1 and the least popular musician is assigned to 0. In our dataset, the most popular musician is Freddie Mercury. Using again the Wilcoxon test, we find using all the samples 8a and for musicians who died after 1950 8c we get a significant difference for the age groups (20-40). We still don't find anything significant for musicians who died before 1950 8b (as this sample is much smaller).



(a)



(b)

Figure 6: The median of differences in pageviews for each pair of musicians (a) all the dataset (b) musicians who died after 1950.

6 DISCUSSION AND CONCLUSION

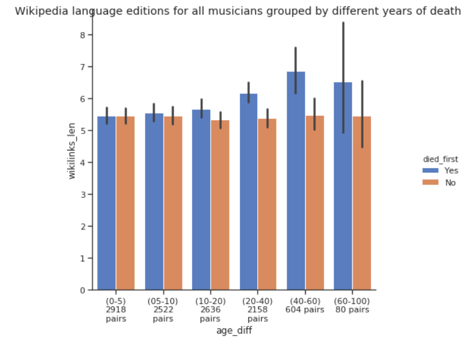
We analyzed people and more especially musicians, using the data from Wikipedia, found the mortality rates on musicians and compared them with long-term popularity. We find out that:

- There isn't any evidence that the 27 club exists. It might seem that musicians die younger, however, when comparing to the general population there is evidence that other occupations are more dangerous with a higher mortality rate at their 20s and 30s
- Musicians who died younger, especially the ones who died after 1950 are more popular than musicians who died later

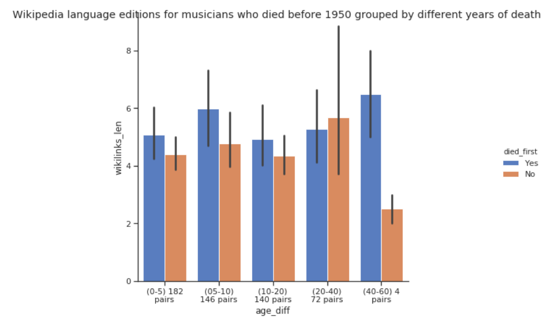
Limitations and future work The main pinpoints is the long-term popularity metric and the popularity of the musicians at the time of death. As we go further in the past, finding data becomes more difficult to find. A suggestion would be to use an external dataset to define the popularity of musicians at the time of death, such as using the sales-records.

Besides, Wikipedia has a recency bias: For example, two people, A and B, of identical (latent) "noteworthiness" (at the time the first person died) and born in the same year (say, 1940). A died in 1960, B in 2010. Wikipedia is more likely to have an article about B than about A, since deaths trigger the creation of Wikipedia articles, and A died before Wikipedia started to exist, whereas B died when Wikipedia already existed. So older people in matched pairs are on average more noteworthy. That means that people who died later are less popular simply because they had a higher likelihood of being included in Wikipedia.

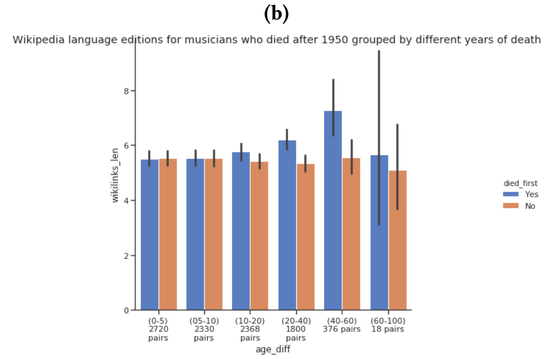
We are continuing our work to address these issues by checking the quality of the matched pairs and defining the uncertainty of our results [3]. We are making the causal question stronger [4] and working on the minimal adjustment set of the properties that affect the long-term popularity



(a)



(b)



(c)

Figure 7: Plots using the number of Wikipedia language editions (mean) for each biography grouped by difference in year of death for (a) all the dataset (b) musicians who died before 1950 (c) musicians who died after 1950.

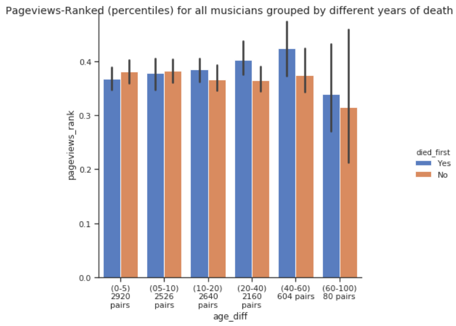
ACKNOWLEDGMENTS

We thank our pets for helping us survive the quarantine

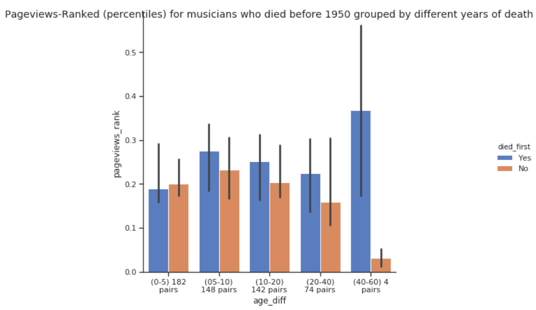
REFERENCES

[1] 2020. 27 Club. https://en.wikipedia.org/wiki/27_Club#Identified_members

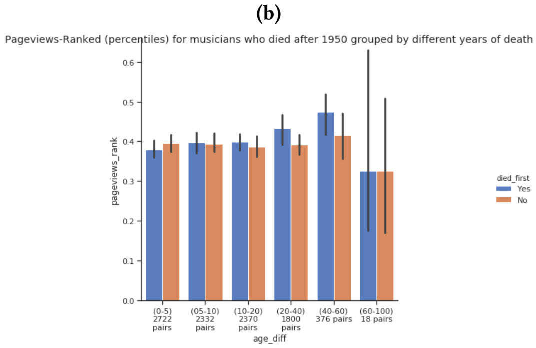
[3] Google document with attack senarios
[4] Google document with the causal question



(a)



(b)



(c)

Figure 8: Plots using the relative rank (where 1 is the most popular musician) grouped by difference in year of death for (a) all the dataset (b) musicians who died before 1950 (c) musicians who died after 1950.

[2] C.R. Epstein and R.J. Epstein. 2013. Death in The New York Times: the price of fame is a faster flame. *QJM: An International Journal of Medicine* 106, 6 (04 2013), 517–521. https://doi.org/10.1093/qjmed/hct077 arXiv:https://academic.oup.com/qjmed/article-pdf/106/6/517/17495911/hct077.pdf

[3] Dianna Theadora Kenny. 2014. *The 27 Club is a myth: 56 is the bum note for musicians.* https://theconversation.com/the-27-club-is-a-myth-56-is-the-bum-note-for-musicians-33586

[4] Martin Wolkewitz, Arthur Allignol, Nicholas Graves, and Adrian G Barnett. 2011. Is 27 really a dangerous age for famous musicians? Retrospective cohort study. *BMJ* 343 (2011). https://doi.org/10.1136/bmj.d7799 arXiv:https://www.bmj.com/content/343/bmj.d7799.full.pdf

[5] Mitsuo Yoshida, Yuki Arase, Takaaki Tsunoda, and Mikio Yamamoto. 2015. Wikipedia Page View Reflects Web Search Trend. In *Proceedings of the ACM Web Science Conference (WebSci '15)*. Association for Computing Machinery, New York, NY, USA, Article 65, 2 pages. https://doi.org/10.1145/2786451.2786495