

---

### Summary

For Problem One, we build a sub-family from influencer to suitor, with different musical genres as a unit. Based on the number of direct suitors in three areas, affecting the time range, the number of indicator of the music genre, establish a comprehensive indicator of the gray association of the impact, and select the influencer with the highest impact score in each music genre. For the selected influencers, draw a tree chart to represent the sub-series of different music genre and analyze them. The representative character in POP/ROCK is The Beatles, and the rest of the results are shown in the body.

For Problem two, we model the similarity of music and determine which artists in the genre are more similar to the artists in the genre. First, we use the main ingredient analysis method toa by artist 14 indicators in the data analysis to 5 main ingredients. Secondly, we use cosine similarity to analyze the correlation of musicians. Compare the similarities between artists within a genre and between genres, where the average similarity of 6 genres is greater than the average similarity between genres.

For Problem three, we compare similarities and influences between and within schools and reveal how genres change over time. First of all, using the data of Problem two and the similarity fever try to find out the comparative analysis between genres and the influential musicians within the genre, the specific content can be found in the body. "Second, we plot the changes in the prevailing genre over time, and finally we use Sen's trend estimation to determine the change in the overall trend." Among them, the overall trend of the number of people in the POP/Rock category is on the rise, and the upward trend is greater.

For Problem 4, first of all, the extent of the impact can be reflected in the similarity between the indicators, and we select a representative of the country genre, Hank Williams, for analysis. The cosine similarity analysis was carried out on the representative character and the suitor using the five main ingredients in Problem two, and it was found that it was not directly affected, and the cosine similarity between the indicators was calculated. Second, we can calculate the correlation coefficient between the remaining indicators and the Spearman rank of popularity, and select the high correlation with the popularity indicator. Indicators of infectious power: popularity, valence, tempo, loudness, acousticalness, duration. Finally, we chose the country genre subnet as the primary study object. Make a comprehensive analysis of the gray correlation of the selected indicators and find out the outstanding figures of the comprehensive indicators. Analyze the influence of musicians with a high composite index. The results showed that the infectious effect index had little to do with a musician becoming an influencer.

For Problem 5, we identify the characteristics of the major leaps in the evolution of music and identify the influencers of major changes. First, the indicators data by year d-d system are de-dimensionally analyzed by the main ingredient analysis method, and 4 main ingredients are obtained. Use the Mann-Kendall trend test for the four main ingredients to identify the year in which the mutation occurred, 1923 and 2017. Finally, the representative figures of the sudden years are analyzed. In 1923, the representative of the vicinity was: Hank Williams, Billie Holiday. The nearby representative for 2017 is Frank Ocean, Flume.

For Problem 6 We analyze the impact of music evolution over time in the Blues genre, identify indicators that reveal dynamic influencers, and explain how artists change over time. First of all, we plot the indicators in the genre about the age of the line chart to observe the dynamic changes of the indicators. Second, we will select 3 indicators with a large absolute inclination as dynamic indicators for each indicator that changes over time. Finally, select to analyze The Beatles band's relevant data, the specific analysis of the body.

**Keywords**    **principal component**    **gray analysis**    **sen's trend**    **Mann-Kendall**

# Digital Music based on Principal Component Analysis

## Contents

- 1.Introduction
    - 1.1 background
    - 1.2 The Task at Hand
    - 1.3 Assumption
    - 1.4 Notations
  - 2.Problem 1
    - 2.1 Explanation of nouns
    - 2.2 Developing a comprehensive evaluation model of influencers' influence
    - 2.3 Representative figure-centered relationship diagram
    - 2.4 Result Analysis
  - 3.Problem 2
    - 3.1 Main Ingredient Analysis
    - 3.2 Correlation analysis
    - 3.3 Similarities among musicians within genres
    - 3.4 Similarity of artists between genres
    - 3.5 Comparison between two types of similarity data
  - 4.Problem 3
    - 4.1 Relationship between genres and within genres
    - 4.2 The prevailing genres have evolved over time
    - 4.3 Trends in the number of people in the Pop/Rock genre
  - 5.Problem 4
    - 5.1 Influencer influences the pursuer relationship model
    - 5.2 Spearman Rank Correlation - Select an infectious indicator
    - 5.3 Comprehensive evaluation of the infectious index model
  - 6.Problem 5
    - 6.1 Analysis of the main ingredients
    - 6.2 Mann-Kendall Trend Test
    - 6.3 Influencers of major changes
  - 7.Problem 6
    - 7.1 Dynamic changes in the Blues genre
    - 7.2 Dynamic changes in the Beatles band
  - 8.Problem 7
  - 9.Strengths and Weaknesses
  - 10.Reference
  - 11.One-page Document
- Appendix

# 1. Introduction

## 1.1 Background

With the development of society and the changes of the times, music has become an indispensable part of human daily life. Throughout the history of music development, we can find that the development of music is inseparable from the development of human society. Artists who create music are naturally influenced by their environment, and their music style is also closely related to social customs, culture, and social political events in the same period. In addition, it has a lot to do with the talents of artists.

In order to better understand the influence and development of the music, we have studied genres and artists in different periods. Besides we use diversified musical characteristics such as rhythm, melody, and lyrics to summarize the genres of music and consider the diversified influence relationship between genres and artists to understand the evolution of music.

## 1.2 The Task at Hand

- ✧ Through the given data to establish a directional network from influencer to follower, and describe the manifestation of music influence in the network.
- ✧ Using the given data to develop a measure of music similarity, and use this to measure the relationship of the artists within genre and between genres.
- ✧ Comparing the relations between different genres and understand its relationship with other musical characteristics and their own evolution.
- ✧ Judging whether the influencer will have an actual influence on the follower based on the given data and whether it is related to the appeal of the genre.
- ✧ Judging from the data whether there are might signify revolutions in the development of music, and finding artists who represent revolutionaries in the network model we have established.
- ✧ Studying the evolution of a genre of music, and explain the relationship between influencers and followers to explain the evolution of artists and genres over time.
- ✧ Explain how the model expresses the relationship between music and the environment, culture, politics, even society.

## 1.3 Assumption

1. Assuming that the similarity calculation of music is limited to the data given in the attachment.
2. Assuming that the contagion of music tends to popularity.
3. Assuming that all music feature data show a linear relationship between adjacent years.
4. When considering the mutual influence within genres, assume that he is only affected by the representative of the genre and not affected by other influencers.
5. Assuming that the number of artists in genres in different eras can reflect the music trends and trends of the era.

## 1.4 Notations

Symbol	Definition
$N$	The number of direct followers
$T$	The influencing duration of influencer
$P$	The number of influenced music genre
$Z$	the comprehensive index
$X$	Data by artist's indicators
$h_j$	energy content
$cc_k$	cumulative energy content

$\delta_i$	Cosine distance
$y_i$	main ingredient
$UF_k$	Positive time series
$UB_k$	Time reverse sequence

---

## 2. Problem1

### Problem Analysis

Problem one requires us to build a sub-family from influencer to suitor. Analyze influence\_data data in a data source. First, we look at the data to find that the pursuers of influencers include both direct and indirect suitors. Among them, the direct pursuer is the direct influence of the influencer and thus becomes the pursuer, and the indirect pursuer is the pursuer influenced by the direct pursuer.

Second, the gene of music has increased with the progress of the times, resulting in a new gene. Among them, music gene refers to the style and characteristics of music, which can essentially reflect the times, nationalities, musicians' personal ideas, aesthetic ideals, spiritual temperament and other internal characteristics of the external imprint. Therefore, we use the gene of music as an important label to divide musicians, to divide musicians.

Again, we identify representative influencers from the music gene and create a comprehensive indicator of the degree of impact. From the table data, it can be seen that the influencer reflects the influence degree from three aspects, namely, the number of direct suitors, the time interval in which the pursuers are located, and the number of music genes affected. "For suitors, we consider only direct suitors because the indirect suitor may follow the influencer because of the personal style of the influencer at the next level, not because of the top influencer." Because the relationship between these three factors is not only interrelated, but also has its own characteristics, so we choose the gray correlation method for comprehensive evaluation, to get a comprehensive index of impact score. Calculate the influence degree score of the influencers in each music category and select the representative influencers.

Finally, for the selected influencers, a tree map is drawn to represent the sub-lines of different musical genes and analyzed.

### The establishment of the Model

#### 2.1 Explanation of nouns

- ✧ Direct follower: People who are directly influenced by influencer
- ✧ Indirect follower: People who are directly influenced by direct follower

#### 2.2 Developing a comprehensive evaluation model of influencers' influence

##### *Step1 The Establishment of Classification Standards*

According to the analysis of the data type in *influence data*, we find that the attributes of musicians include artists name, number ID, music genre, influencer active start. Analyzing the above data types, we regard the genre of music as an important label for dividing artists to divide the artists. Because music genre refers to the style and characteristics of music, which can essentially reflect the internal characteristics of the era, nation, artists' personal ideas, aesthetic ideals and spiritual temperament.

##### *Step2 The Selection of Indicators*

In order to reflect the influence of a musician, we analyze from the three aspects

**The number of direct follower(N):** The number of direct followers can directly reflect the popularity of the influencer, and it can also reflect the influence of the influencer. Since indirect followers may follow the influencer because of the personal style of the upper-level influencer, not because of the top influencer, so here we only consider the number of direct followers.

**The influencing duration of influencer(T):** The range of the followers' time reflects the influencing time range of the influencer. The larger the scope indicates that the influencer has far-reaching influence (to describe time), the smaller the scope indicates that the influencer is only a flash in the pan, which is

a temporary trend.

$$T = t_1 - t_2 \quad (1)$$

Among them,  $t_1$  represent the largest data in followers active start,  $t_2$  represents the smallest data in follower active start.

**The number of influenced music genre(P):** Although the genre of each musician is different, the range of genre influenced by the influencer can also reflect the influence of the influencer. The larger the scope of influence, the deeper the influencer's knowledge of music, which can penetrate from the details to all levels.

Due to the music genre increases with the progress of the times, we perform a weighted summation based on the number of music genre in the age of the followers when calculating this index. The specific formula is as follows:

$$P = N \sum_{i=1}^n \frac{1}{a_i} \quad (2)$$

In which  $a_i$  represents the total number of music genre existing in the age of the followers.

### **Step3 The Setting of Comprehensive Evaluating Model**

Based on the above description, we have selected three indicators: the number of direct followers, the influencing time range of the influencer and the music genre of influence.

#### ● **Selecting Top Indicators**

Since the number of direct followers intuitively reflects the degree of influence of the influencer, the number of direct followers is used as the parent indicator.

#### ● **Calculating Correlation Coefficient**

Define the correlation coefficient of the top index N and P, T at the kth point is  $y_i(k)$

$$y_i(k) = \frac{a + b\rho}{\Delta_i(k) + b\rho} \quad (3)$$

The calculating formula of  $\Delta_i(k)$  in (3) is:

$$\Delta_i(k) = |x_{ki} - x_{k0}|, i = 1, 2, \dots, m, k = 1, 2, \dots, n \quad (4)$$

Among them,  $a, b$  is:

$$a = \min_{1 \leq k \leq n} \min_{1 \leq i \leq m} \Delta_i(k) \quad (5)$$

$$b = \max_{1 \leq k \leq n} \max_{1 \leq i \leq m} \Delta_i(k) \quad (6)$$

The value range of  $\rho$  is (0,1), here we take  $\rho = 0.5$

#### ● **Calculate the Degree of Relevance**

$$r_i = \frac{1}{n} \sum_{k=1}^n y_i(k), i = 1, 2, \dots, m \quad (7)$$

#### ● **Setting the Comprehensive Evaluating Model**

Through the correlation obtained in the previous step, calculating the weight of each indicator

$$r'_j = \frac{r_j}{r_1 + r_2 + r_3} \quad (8)$$

As the weight of each index, to get the comprehensive index Z.

$$Z = r'_1 x_{k1} + r'_2 x_{k2} + r'_3 x_{k3} \quad (9)$$

Indicator	The number of direct follower(N)	the influencing time range of the influencer (T)	the music genre of influence (T)
Weights	0.963244	0.915657	0.8207

Table 1

Finally, the comprehensive index is the score of influence, the higher the score, the greater the influence.

#### **Step4 Selecting the Representative Character in Each Music Genre**

According to the above evaluation model, the comprehensive score of each influencer's influence is calculated, and the person with the highest score in each field is selected as the representative character of the music genre. The information of the 19 selected representatives is shown in the table below:

Artist name	Artist id	genre	T	N	P
Terry Riley	750519	Avant-Garde	40	36	5
Muddy Waters	608701	Blues	60	113	6
Alvin & the Chipmunks	744969	Children's	40	3	2
Steve Reich	37658	Classical	40	24	5
Spike Jones	744548	Comedy/Spoken	60	20	9
Hank Williams	549797	Country	80	184	6
Henry Mancini	678009	Easy Listening	50	19	6
Kraftwerk	104714	Electronic	40	108	4
Woody Guthrie	577531	Folk	80	80	7
Fela Kuti	138833	International	50	23	5
Miles Davis	423829	Jazz	60	114	12
Tito Puente	607283	Latin	50	21	6
Mannheim Steamroller	675495	New Age	40	9	5
The Beatles	754032	Pop/Rock	50	615	15
Marvin Gaye	316834	R&B;	60	169	7
Desmond Dekker	819344	Reggae	30	21	2
Amy Grant	24944	Religious	40	16	5
Ennio Morricone	798662	Stage & Screen	50	30	3
Billie Holiday	79016	Vocal	80	106	9

Table 2

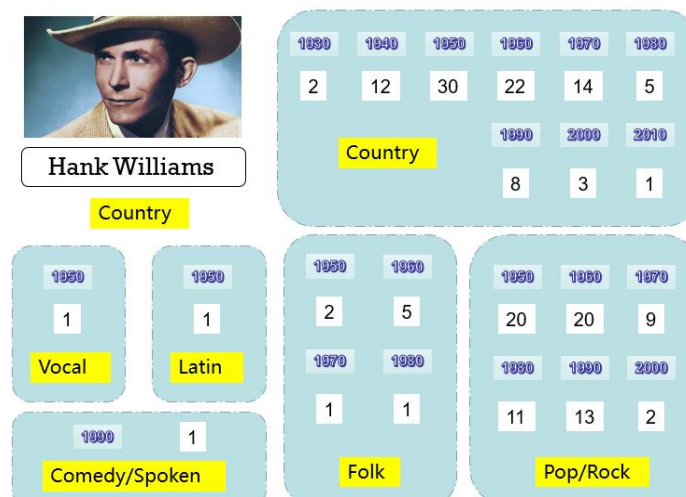
### **2.3 Representative figure-centered relationship diagram**

Drawing a network diagram of the representative characters in each music genre, including a tree diagram of direct and indirect followers, and indicate the age and music genre attributes of the followers. Due to the complexity of the data, we chose The Beatles and Hank Williams for visual subnet analysis. We use tabular data to display the remaining major influencers.

#### **2.2 Result Analysis**

Shown below are the direct influencers of Hank Williams, the influence category and the year.

Figure 1: The genre of country



Hank Williams, an early representative of Country genre, has a great influence. Many artists with greater influence in the 1960s and 1970s are its direct followers, but due to the age that he belonged to is a long time ago, it has led to his musical style gradually faded out of the stage in the wheel of history, with less and less influence, and a clear downward trend in direct followers. He has influenced 6 musical genres, and the music genre of his direct followers is mainly Country and Pop/Rock genres.

Shown below are the direct influencers of The Beatles, the influence category and the year.

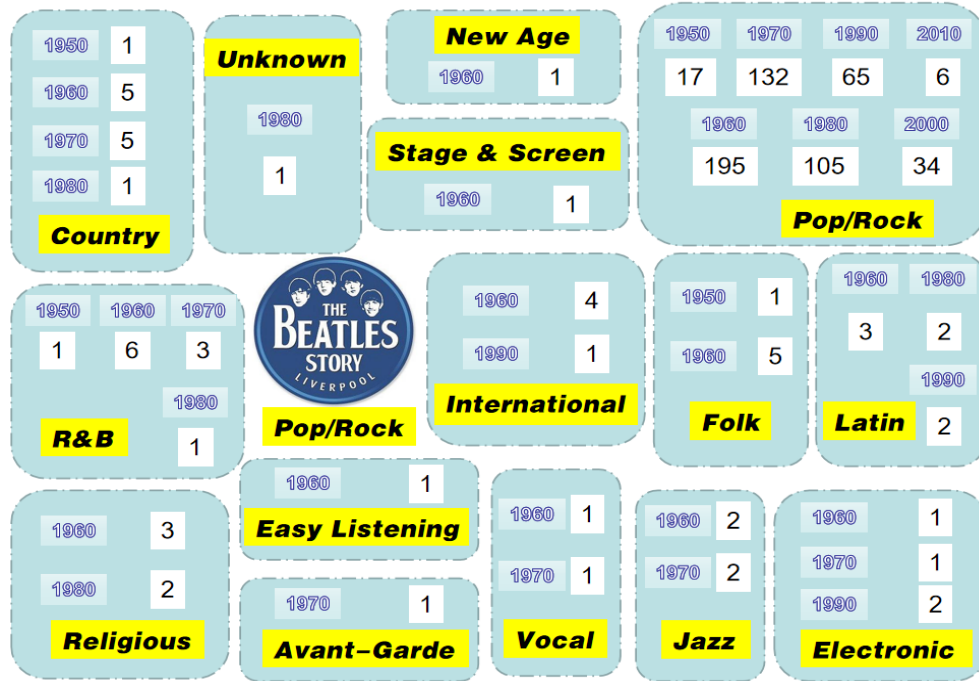


Figure 2: The genre of Pop/Rock

The Beatles are a pop/rock style band. They have a large number of suitors and a long influence. Since 1960, they have had a large number of followers in the world, especially in the 60s, 70s and 80s. 200 The range of music genres it influences is very wide, ranging from their own pop music/rock music to country, folk, R&B, Latin and other genres all their direct influencers, but their main influence is still in Pop. On /Rock, more than 90% of direct followers are from people of different ages in Pop / Rock genres.

The two music genre visualization subnets are made above. Due to the large amount of representative information, the remaining representative information is displayed by the following image:

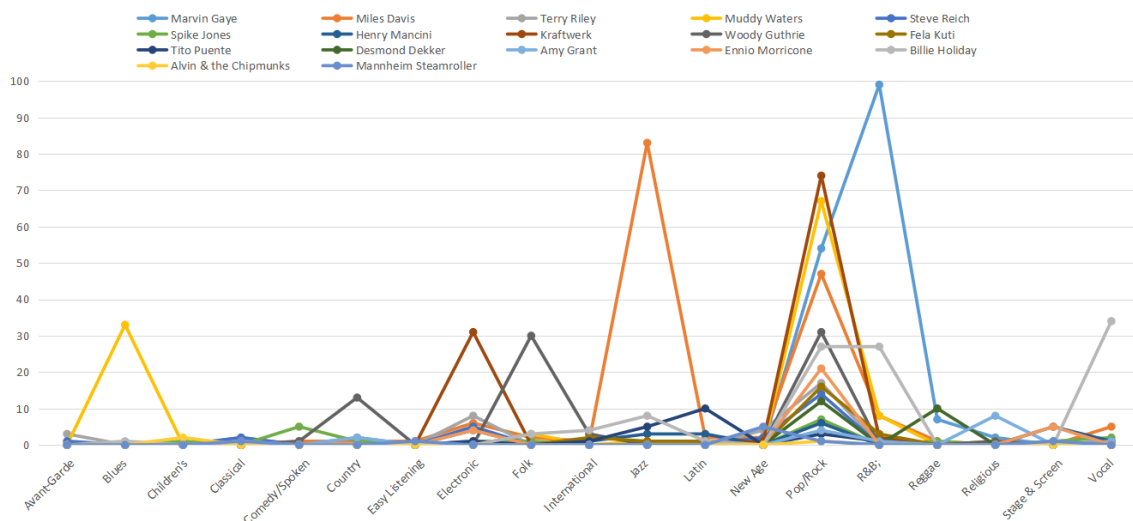


Figure 3: Different genres

Each broken line in the above picture represents an artist who has a great influence in the music genre he belongs to. Observing the above picture, we find that each polyline has a peak, and this peak basically appears in the music genre where they are located; secondly, every artist has a greater influence and more in the Pop/Rock genre from this judgment, Pop/Rock genre is relatively popular and can be widely accepted. In addition, Miles Davis of Jazz genre and Marvin Gaye of R&B genre have more direct followers in their respective genre than other influencers. From this point, it can be seen that their songs are highly recognized and conform to the trend of the times.

### 3. Problem2

#### Problem Analysis

Problem two requires us to model the similarity of music and to determine which artists within the genre are more similar to the artists between the genres. First of all, by looking at the data of data by artist, we observed 14 indicators such as artist name, artist id, energy, many of which are more similar, reflecting a certain degree of overlap in information, in the study of correlation variables too many increase the complexity of the problem, so we choose the main ingredient analysis method to reduce the index dimensional, with fewer indicators to reflect the information. The information contribution rate and cumulative contribution rate of the indicator are calculated by the main ingredient analysis method, and the indicators are sorted, the main ingredients of the indicator with the contribution rate of more than 90% are selected, and the weighted and obtained by the main ingredients is the comprehensive index of the musician to reflect its characteristics.

Second, we analyze the relevance of musicians. Because each musician's feature information is in the same dimension, we use cosine similarity to calculate the relevance of each type of song.

Finally, we analyze the correlation between songs of different music genre, and compare the correlation with the average correlation between the same genre musicians by selecting the works of the sub-groups made in Problem I.

#### The establishment of the Model

##### 3.1 Main Ingredient Analysis

###### *Step1: The selection of indicators*

We use the following 14 indicators for main ingredient analysis:

danceability, energy, valance, tempo, loudness, mode, key, acousticness, instrumentalness, liveness, speechiness, explicit, duration ms, popularity, count.

###### *Step2: The main ingredient analysis*

Since the 14 indicators given by the data can reflect the main characteristics of the song in multiple directions, but there is a certain correlation between the indicators, the information reflecting the problem will have a little bit overlap.

Selecting the main ingredient analysis method to turn 14 indicators into a few more comprehensive indicators, calculating the energy and cumulative energy content of each indicator, and using the energy content of the main ingredient as the weight to develop a comprehensive evaluation model of the main ingredient.

The sequence of 14 indicator data is  $X = X_1, X_2, \dots, X_{14}$

- Normalizing the sequence of 14 indicators after data preprocessing

$$\bar{x} = \frac{x_{ij} - \frac{1}{n} \sum_{i=1}^n x_{ij}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \frac{1}{n} \sum_{i=1}^n x_{ij})^2}}, (n = 5854) \quad (10)$$

- Calculating the correlation coefficient matrix



$$r_{ij} = \frac{\sum_{k=1}^n \widetilde{x}_{ik} \widetilde{x}_{kj}}{n-1}, (n = 5854) \quad (11)$$

■ Calculating eigenvalues and eigenvectors

Counting the eigenvalues of the correlation matrix R  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{14} > 0$  and corresponding normalized feature vector  $\mu_1, \mu_2, \dots, \mu_{15}$

$$\begin{cases} x_1 = \mu_{1,1}\widetilde{X}_1 + \mu_{1,2}\widetilde{X}_2 + \dots + \mu_{1,14}\widetilde{X}_{14} \\ \dots \\ x_{14} = \mu_{14,1}\widetilde{X}_1 + \mu_{14,2}\widetilde{X}_2 + \dots + \mu_{14,14}\widetilde{X}_{14} \end{cases} \quad (12)$$

■ Figuring out the energy content and cumulative energy content

$$\begin{cases} h_j = \frac{\lambda_j}{\sum_{i=1}^{14} \lambda_i} \\ cc_k = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^{14} \lambda_i} \end{cases} \quad (13)$$

The energy content and cumulative energy content of the 14 indicators are as follows:

indicators	energy content	cumulative energy content	indicators	energy content	cumulative energy content
1	0.227378746	0.227378746	8	0.06040273	0.810419479
2	0.122550481	0.349929227	9	0.053311569	0.863731048
3	0.100385594	0.450314821	10	0.046942923	0.910673971
4	0.091267136	0.541581957	11	0.039310617	0.949984588
5	0.073334632	0.614916589	12	0.02339942	0.973384008
6	0.0702334	0.685149989	13	0.020041915	0.993425923
7	0.06486676	0.750016749	14	0.006574077	1

Table 3

Selecting the top x main ingredients according to the above table, so that the cumulative energy content of the feature vector exceeds 90%. The resulting data is large and is not presented in the text.

Use the weights as matrix A, reduce the dimensions of 14 indicators to 5 main ingredients, and use the following formula to calculate:

$$y_i = \sum_{j=1}^{14} a_j \widetilde{X}_j \quad (14)$$

Due to the large amount of data after calculation, it will not be shown in the text.

### 3.2 Correlation analysis

#### Step1

The partial data of the five main ingredients of all musicians calculated above are shown in the following table:

#### Step2

Calculate the cosine similarity between the five main ingredients of the two musicians. Record the sequence of the musician's main ingredients as  $Y_i$ . The calculation formula for the cosine similarity is as follows:

$$\delta = \cos\theta \quad (15)$$

The partial data of the similarity matrix is as follows:

	Frank	Vladimir	Johnny	Billie	Bob	Rolling
Frank Sinatra	1	0.956	0.938	0.964	0.976	0.979
Vladimir Horowitz		1	0.967	0.99	0.979	0.979
Johnny Cash			1	0.988	0.979	0.981
Billie Holiday				1	0.990	0.991
Bob Dylan					1	0.999
The Rolling Stones						1

Table 5

The closer the cosine value is to 1, the closer the angle is to 0 degrees, that is, the more similar the two vectors, and the higher the correlation between the two musicians.

### 3.3 Similarities among musicians within genres

When analyzing the musicians in the genre, we mainly analyzed the two music genres, Pop/Rock and R&B. Find out the correlation data between the musicians in the two music genres of Pop/Rock and R&B in the above correlation matrix, recreate the two correlation matrices, and use the correlation information of the two music genres through cluster analysis and thermodynamics the image is represented visually.

Pop/Rock genre cluster analysis is shown below. Since we have already obtained the correlation matrix, we choose R-type cluster analysis. Among them, we use numbers to represent the names of musicians, and the corresponding relationships are shown in the following table:

1	The Beatles	8	Casting Crowns	15	Be Bop Deluxe
2	Soundgarden	9	Teenage Fanclub	16	Klaatu
3	System of a Down	10	The Apples in Stereo	17	BarlowGirl
4	T. Rex	11	The Beach Boys	18	Don McLean
5	Big Star	12	Blind Melon	19	John Fogerty
6	Sloan	13	Badfinger	20	Bob Weir
7	Toto	14	Christine McVie		

Table 6:

The cluster analysis of Pop/Rock genre and the heat map of Pop/Rock genre correlation are as follows:

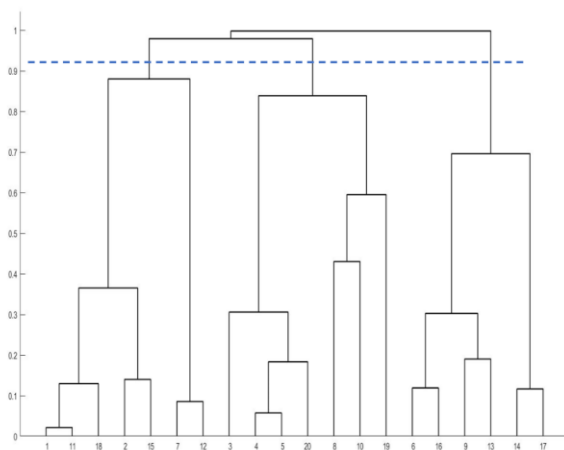


Figure 4: The cluster map

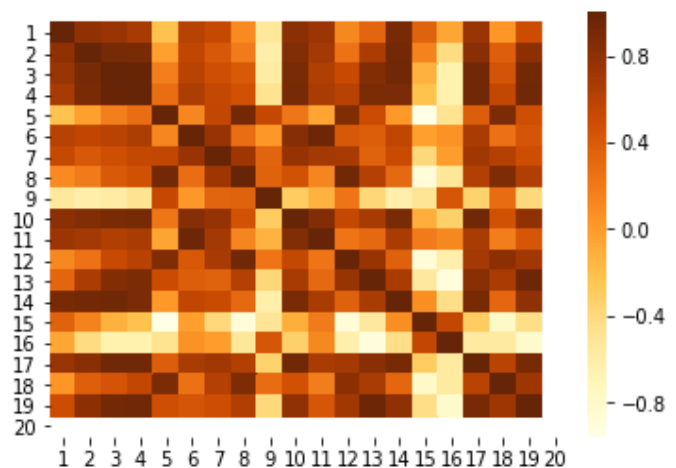


Figure 5: The heat map

From the figure 4, we can see that the clustering results number the musicians as (1,11,18,2,15,7,12), (3,4,5,20,8,10,19), (6, 16, 9, 13, 14, 17) are grouped into 3 categories. Musicians in the same category are more similar.

From the figure 5, we can intuitively see the similarity between musicians. The darker the color, the higher the similarity. Regarding the R&B genre, the correspondence between numbers and musician names is shown in the following table:

1	Prince	8	D'Angelo	15	Al Green
2	Patti LaBelle	9	Lisa Stansfield	16	Don Covay
3	Trey Songz	10	Ready for the World	17	The Flamingos
4	Morris Day	11	Chaka Khan	18	The Mills Brothers
5	The-Dream	12	Dionne Warwick	19	The Stylistics
6	The Weekend	13	Commodores	20	Lamont Dozier
7	TLC	14	Dinah Washington		

Table 7:

The cluster analysis of R&B genres and the heat map of R&B genre relevance are as follows:

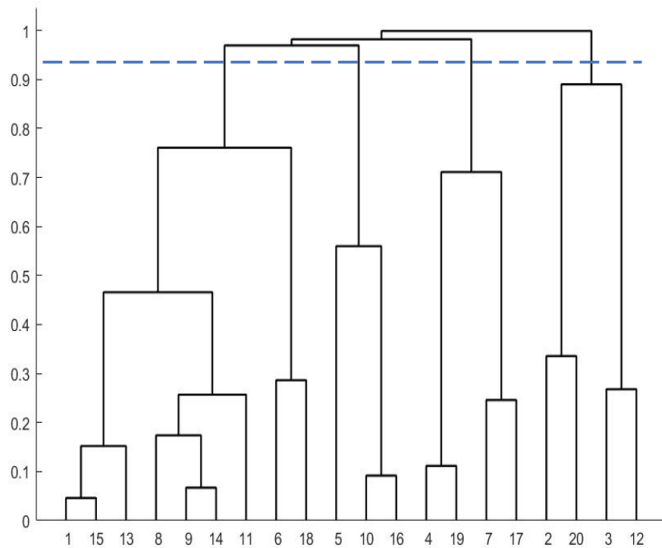


Figure 6: The cluster map

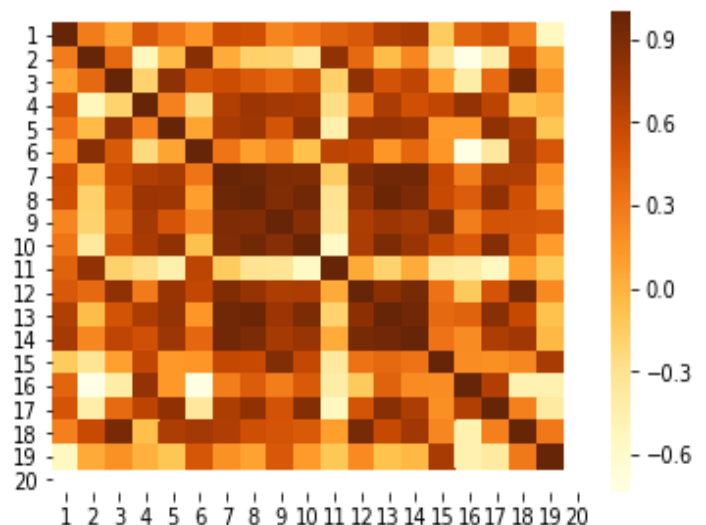


Figure 7: The heat map

From the figure 6, we can see that the clustering results number the musicians as (1,15,13,8,9,14,11,6,18), (5,10,16), (4,19,7,17), (2,20,3,12) are clustered into 4 categories.

### 3.4 Similarity of artists between genres

For musicians between genres, we select the similarities between the representative musicians in each genre in the first Problem to analyze, find the correlation data between the genre representatives in the above correlation matrix, and recreate the correlation matrix, and perform cluster analysis and draw thermodynamic images.

The cluster analysis and the heat map are as follows:

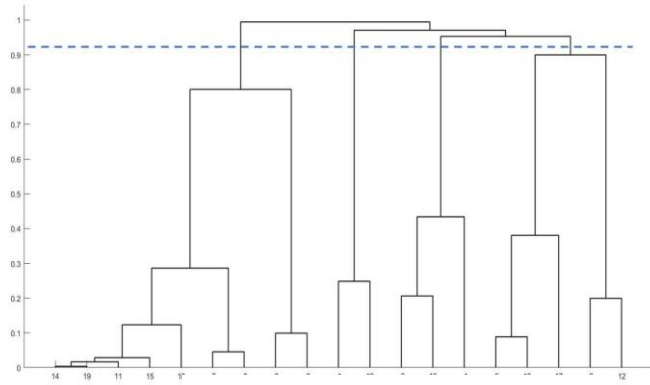


Figure 8: The cluster map

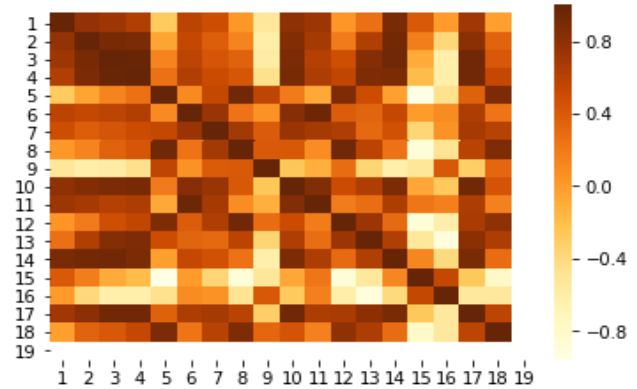


Figure 9: The heat map

From the figure 8, we can see that the clustering results number the musicians as (14, 19, 11, 15, 18, 7, 8, 3, 9), (1, 10), (2, 13, 4), (5, 16, 17, 6, 12) clustered into 4 categories. Musicians in the same category are more similar.

### 3.5 Comparison between two types of similarity data

Calculate the average of the similarity of all music genre and the average of the similarity of different genre for convenience of observation.

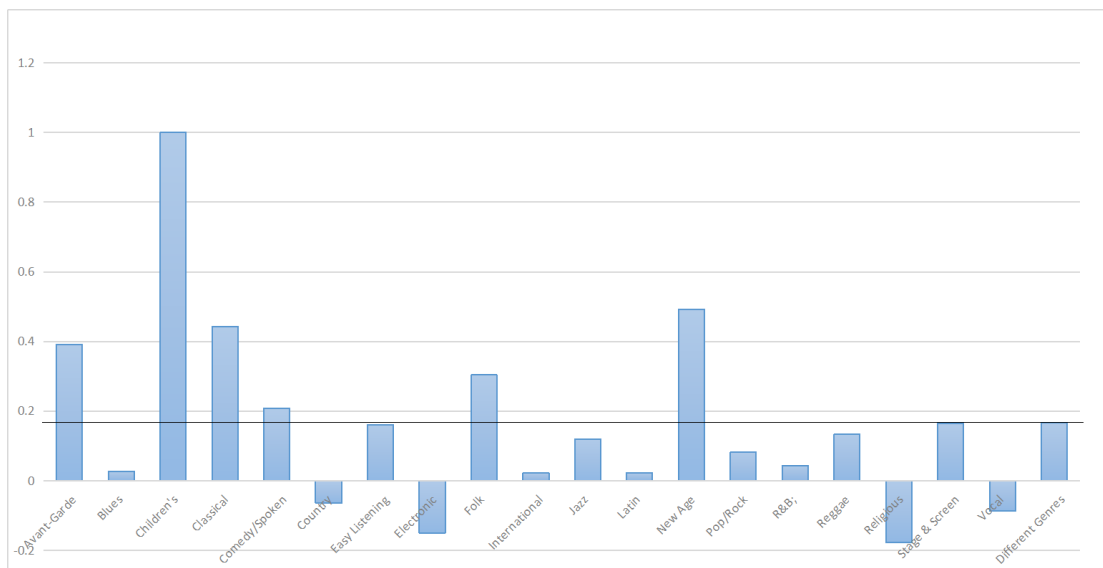


Figure 10: The similarity of different genres

Through the image of the average similarity, it can be clearly observed that the average similarity within the 6 genres is greater than the average similarity between different genres, and the average within the remaining genres is less than the average between different genres. Among them, the average similarity of Children's genre is the highest, reaching 1. This is because there is only one artist in the genre, so it is 1. However, because the sample data is too small, the average similarity within the genre does not have any reference value.

## 4. Problem3

### Problem Analysis

Problem 3 first requires us to compare similarities and influences between and within schools and to explain the relationships between some genres and other genres, and second, to reveal how genres change over time.

First of all, using the data of Problem II and the similarity fever, we try to find out the influence of musicians between genres and within genres, and analyze the relationship between genres and genres by comparing the various information of these musicians.

Secondly, we analyze how the genre changes over time, and our analysis of the problem is divided into the following two parts. The first part, mapping the changes of the prevailing genres over time, reflects the emergence of new genres as the times change, and the second part analyzes the changing trends in the number of people within a particular genre, where we choose to analyze the Pop/Rock kind of data in detail. Organize the data given in the topic and sort out the number of people in Pop/Rock for each time period. In order to reflect the trend change of the number of people, the change of the overall trend is judged by Sen's trend estimation method, and the point with the large slope change is analyzed by making a line chart.

## The establishment of the Model

### 4.1 Relationship between genres and within genres

This relationship has been described in detail in problem number two.

### 4.2 The prevailing genres have evolved over time

We visually show the prevalence of genres over time in a way that combines graphics:

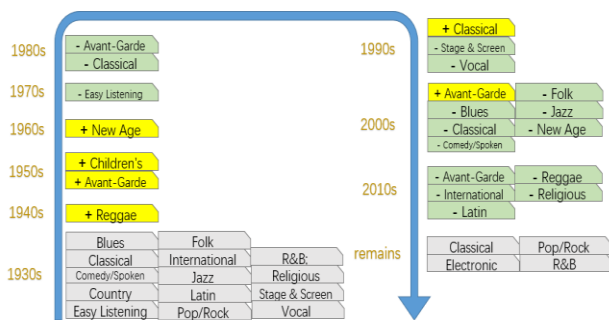


Figure 11: Calendar figure

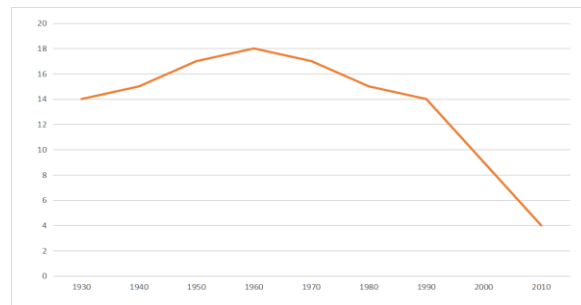


Figure 12: Line figure of genres

You can see from the pictures in 1930, the world has Pop / Rock and other 14 popular music genres, over time, until 1960, the popular music genres showed a clear upward trend, in 1960 the number of popular music genres peaked, there are 18, of which from 1940 to 1950, the largest upward trend. But after 1960 the popular music genres showed a downward trend, with the prevalence of music genres dropping by 10 between 1990 and 2010. It can be seen intuitively from the line chart, in the course of the development of the world, the early and present popular music genres of a relatively single species, the general aesthetic trend is consistent, and in the middle of the popular music genres diverse and rich.

### 4.3 Trends in the number of people in the Pop/Rock genre

For the analysis of the overall trend, to show the change in the trend should be seen in terms of the number of changes, we use sen's trend test. Select an indicator that can reflect the overall rate of change - the inclination rate, the inclination rate is the slope calculated back by each time node, and finally the slope obtained is worth it, so the inclination rate can indicate the total change of the overall trend. If the dumping rate is positive, the overall trend is upward, and if the dumping rate is negative, the overall trend is decreasing. The size of the absolute value of the inclination reflects the extent to which the trend changes.

The inclination is calculated as follows:

$$\alpha = \text{mean}\left(\frac{z_j - z_i}{j - i}\right), \forall j > i \quad (16)$$

Where  $z_i, z_j$  is the number of people per decade, which means that the slope is  $j > i$  calculated for subsequent years.

Calculate the tilt rate of the change in the number of people:

$$\alpha = 1.3746938 \quad (17)$$

The value of the resulting tilt rate is positive, indicating that the overall trend of the number of people in the Pop/Rock genre is on the rise, with a greater degree of upward trend.

The number of people per node is represented by a line chart:

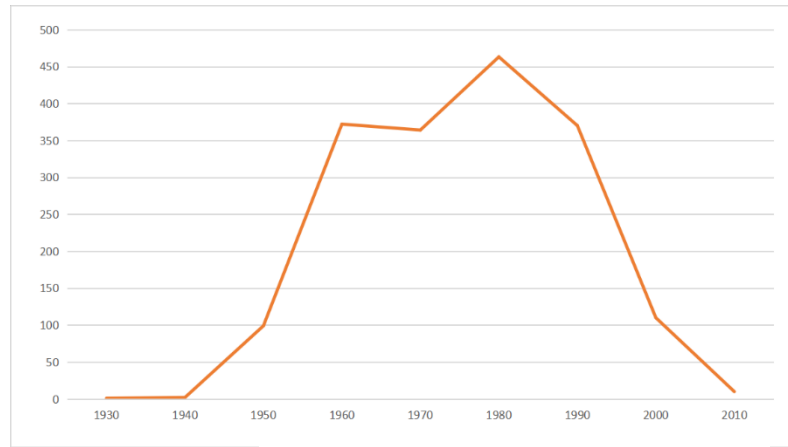


Figure 13: Line chart

The data for each node is shown in the table below:

1930	1940	1950	1960	1970	1980	1990	2000	2010
1	2	99	372	364	463	370	110	10

Table 8

From the figure 13 we find that between 1950 and 1960 the number of people in this genre grew rapidly, with a slope of 27.3, and combined with Problem one we found that the famous Beatles began to prevail during this period, the band was the most direct suitor of all influencers, so the growth rate was faster. Pop/Rock's numbers have been declining since peaking in 1980, and the fastest period of decline was from 1990 to 2000, and looking at the data in influence data reveals that since 1990 the R and B genres have been prevalent, and we judge that the increase in the number of other musical genres has led to a decrease in the numbers of Pop/Rock.

## 5. Problem4

### Problem Analysis

Problem 4 requires the analysis of three sub-Problems. The first is to determine whether the influencer directly influences the suitor based on the data in data influence, and from which aspects the influence should be made. The second is to judge that those characteristics of music are more infectious. The third is that more contagious features play a similar role to a particular musician in becoming a famous pop.

First, we analyze whether influencers directly influence suitors. The degree of influence can be reflected by the similarity between indicators. We choose a genre for analysis, combined with the detailed subnet given in Problem 1, we analyze Hank Williams, a representative of country genre. Perform cosine similarity analysis between the representative and the suitor using the five main ingredients in Problem

2, and make a line chart to observe whether the influencer directly affects the suitor; then calculate the suitor's 14 indicators in data by artist and the influencer's the similarity between the 14 indicators, through similarity data, shows in which aspects the influencer has a greater influence on the pursuer.

Second, we have to determine which characteristics of music are more contagious. Among them, the appeal can be reflected by the indicator popularity, which indicates how much the public likes music. Use Spearman rank correlation to calculate the correlation between the data of all indicators in full music data and the popularity indicator data, and select the indicators with greater correlation with the popularity indicator.

Finally, we have to reflect whether the selected indicators above have an impact on a musician becoming a popular person. We selected country genre as the main research object through the genre network made in Problem 1. Study the infectious power indicators of all musicians in the genre, and do a comprehensive gray-relation analysis of the selected indicators to find out who have outstanding comprehensive indicators. If these characters have a large influence index in Problem 1, it means that the selected index plays an important role in the country genre musician becoming a famous pop. Otherwise, it is not important.

## The establishment and solution of the model

### 5.1 Influencer influences the pursuer relationship model

First, we analyze whether the influencer directly affects the suitor, and the degree of influence can be reflected by the similarity between the indicators. We selected Hank Williams, the representative of country genre, to analyze, using the data in Problem 2 on the analysis of the main ingredients made by the musician, to note that the main ingredient data of  $h(h_1, h_2, h_3, h_4, h_5)$  Hank Williams  $C$  is the main data matrix of  $h$  184 direct suitors, and to calculate the cosine similarity between the main ingredient data of the musician in the  $C$  matrix, as follows:

$$\delta_i = \cos\theta = \frac{h \times C(i)}{\sqrt{h^2} \cdot \sqrt{C(i)^2}} \quad (18)$$

The similarity coefficient between influencer Hank Williams and all musicians in his genre is shown by the broken line chart below:

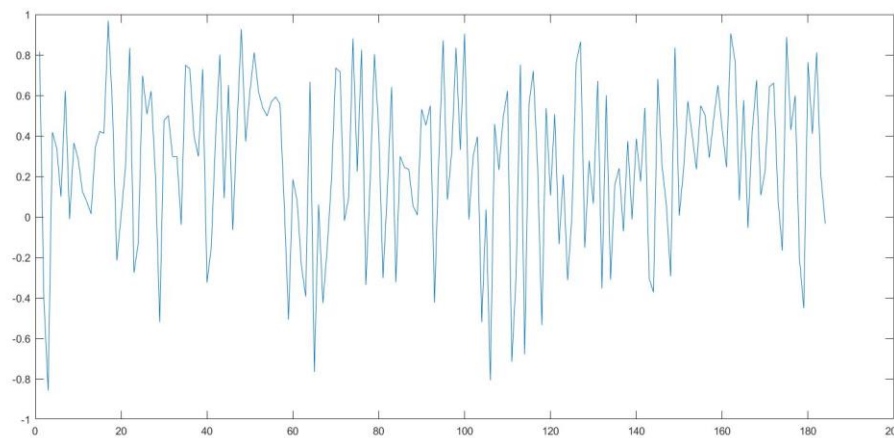


Figure 14: Line chart

From this image, it is clear that the similarity between Hank Williams and his direct followers fluctuates considerably and fluctuates significantly, and for some direct followers, Hank Williams has a greater overall impact on them, while for others, the combined influence is less.

To get a direct result, we analyzed the similarity of each indicator, selecting specific data from Hank Williams in the data by artist and his direct followers to calculate the similarity of the 14 indicators between direct followers and Hank Williams, respectively. Note that Hank Williams' indicator data is,

and that the indicator data matrix of  $g(g_1, g_2, g_3, \dots, g_{14})$  184 direct suitors is, where the data  $D$  for  $D$  each indicator is a column vector of,  $d_1, d_2, \dots, d_{14}$  calculated using cosine similarity, and the result is shown in the figure below:

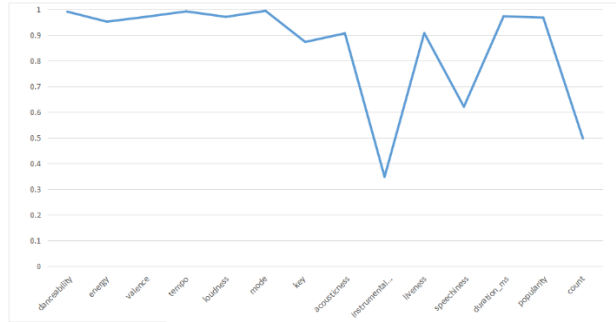


Figure 15: Line chart

A line chart is the similarity between Hank Williams's musical characteristics indicators and lyrics of his followers. As we can see from the graph, indicators with similarity above 0.9 have danceability, energy, valence, tempo, loudness, mode, acousticness, liveness, duration ms, popularity, similarity indicators below 0.9 have key, instrumentalness, speechiness, count. We believe that indicators with a similarity above 0.9 have a greater impact on followers. It's worth noting that Hank Williams has less influence on followers in the instrumentalness, speechiness, count indicators, which also has a lot to do with Hank Williams' time environment and personal views. Because he was an early singer active in the 1930s in the United States, the times are far from the present, the public aesthetic has changed greatly, people of different eras like different styles of artists, so artists will follow the times to choose their recognized instrumentalness, speech form to match their performances, and Hank Williams has no greater relationship, so he has less influence in this area. In the count indicator, the reason for his small influence is mainly influenced by his own views, regardless of others, his followers do not change their own count because of the number of counts he sends.

## 5.2 Spearman Rank Correlation - Select an infectious indicator

Infectiousness can be reflected by the indicator popularity, which indicates how much the public loves music. The full music data all indicators in the data used in the data are related to the Popularity indicator data using The Speedman rank correlation, and the correlation coefficient is as follows:

danceability	energy	valence	tempo	loudness	mode	key
0.0754	0.1881	-0.2231	0.2275	0.2182	-0.1401	0.1195
acousticness	instrumentalness	liveness	speechiness	duration_ms	popularity	count
-0.3996	-0.0811	-0.0917	-0.1524	0.4809	1	-0.1366

Table 9

Select the indicators that are more relevant to the popularity indicator: According to the literature, the more similar the two variables are when the Spielman correlation coefficient increases in absolute terms. So we select the indicators that are more infectious:

valence	tempo	loudness
acousticness	duration_ms	popularity

Table 10

## 5.3 Comprehensive evaluation of the infectious index model

We chose country genre as the primary study object. This paper studies the infectious index of all musicians in this genre, and does a comprehensive analysis of the gray correlation of selected indicators to find out the outstanding figures of the comprehensive index.



### Step 1: The performance of a gray correlation analysis of the infectious forces indicator

Using the problem of a gray association solution method, the weight of each indicator is obtained and the comprehensive index is obtained:  $Z$

$$Z = r'_1 x_{k1} + r'_2 x_{k2} + \cdots + r'_6 x_{k6} \quad (19)$$

The weights are shown in the table below:

Index	valence	tempo	loudness	acousticness	duration_ms	popularity
Weight	0.523797	0.523944	0.523783	0.523797	1	0.523838

Table 11

Select the musicians with larger composite indicators as shown in the table below:

artist_name	Comprehensive score	artist_name	Comprehensive score
Jeff Foxworthy	406957.6	The The	322790.5
James McMurtry	355484.3	The Marshall Tucker Band	313573.3
Grateful Dead	337784.3	Jeff Buckley	307734.7
Jerry Garcia	335233.9	Greg Brown	283397.2
The Waterboys	329112.8	Don Henley	281026.7

Table 12

### Step 2: The Analysis of the impact of musicians with large composite metrics

The direct suitors of musicians with larger comprehensive indicators above, the time range of influence, the number of influence genres, and the degree of impact indicators in Problem I are shown in the table below:

name	followers	genre	con	Influence
Jeff Foxworthy	0	0	0	0
James McMurtry	2	1	20	0
Grateful Dead	60	5	50	21.06033
Jerry Garcia	10	3	40	107.681
The Waterboys	11	4	20	48.72083
The The	7	1	20	32.19163
The Marshall Tucker Band	12	2	30	25.87655
Jeff Buckley	24	2	20	40.67005
Greg Brown	8	3	20	43.07241
Don Henley	12	2	20	28.4812

Table 13

Looking at the table above, we found that the musicians in the table had little influence, so it shows that infectious indicators have little effect on musicians becoming a well-known pop star.

## 6 Problem5

### Analysis of the Problem

The fifth Problem asks us to find out from the data given in the topic whether there is a potential for a revolutionary major leap in the evolution of music. and find out which musicians are the influencers of major changes.

First of all, we want to identify major historical changes, so we take into account the changes in music over time, using the d-ata\_by\_yearthe topic. Because there are more indicator types in this data, and the

similarity between many indicators is higher, so we choose the main ingredient analysis method to reduce the index to the dimensional, with fewer indicators to reflect the information. The information contribution rate and cumulative contribution rate of the indicator are calculated by the main ingredient analysis method, and the indicators are sorted, the main ingredients of the index with the contribution rate of more than 99.5% are selected, and the values of each main ingredient are calculated.

The second is to identify the years from 1921 to 2021 when musical characteristics mutated. Using the Mann-Kendall trend test, the time node where the positive correlation sequence of time intersects the negative correlation represents the year in which the mutation occurred.

Finally, the time node is analyzed near the year in which the mutation occurred, and the representative figures of the sudden year in Problem one is found and analyzed.

## 6.1 The establishment and solution of the model

### 6.1 Analysis of the main ingredients

Select the main ingredient analysis method to turn 13 indicators into a few more comprehensive indicators, calculate the information contribution rate and cumulative contribution rate of each indicator, with the main ingredient contribution rate weight, establish a comprehensive evaluation model of the main ingredient. The calculation formula method is consistent with the problem.

The sequence consisting of 13 indicator data is  $Y = Y_1, Y_2, \dots, Y_{13}$ .

The information and cumulative contribution rates for the 13 indicators are as follows:

Index	Information contribution rate	Cumulative contribution rate	Index	Information contribution rate	Cumulative contribution rate
1	54.97443276	54.97443276	8	0.827175288	98.53133182
2	13.93677281	68.91120557	9	0.692320971	99.22365279
3	11.75770216	80.66890773	10	0.601372029	99.82502482
4	7.27659598	87.94550371	11	0.142655703	99.96768053
5	5.271383484	93.2168872	12	0.032319474	100
6	2.587363089	95.80425028	13	3.6538E-16	100
7	1.89990625	97.70415654			

Table 14

Select 4 main ingredients from the table above, resulting in a cumulative contribution of more than 90% of the feature vector. The resulting data is large and is not presented in the text

Using the weights as a  $E$  matrix, the 13 indicators are reduced to 4 main ingredients and calculated using the following formula:

$$z_i = \sum_{j=1}^{14} e_j \tilde{Y}_j \quad (20)$$

Due to the large amount of data after calculation, it is not shown in the body.

### 6.2 Mann-Kendall Trend Test

There are mutations in musical characteristics, so we wanted to find out the year in which the musical characteristics mutated from 1921 to 2021. Using the principles of statistics, using the Mann-Kendall trend test, the positive correlation sequence with time and negative correlation are graphed, and the time node at which the two lines intersect is the year in which the mutation occurred. This Problem conducts Mann-Kendall trend detection of the four main ingredients derived from the analysis of the main ingredients.

The steps of the Mann-Kendall trend are follows:

1. For the four main ingredients that affect the characteristics of extreme precipitation in summer, the order column is constructed separately:

$$s_k = \sum_{i=1}^{n-1} r_i, k = 2, 3, \dots, n, (n = 100) \quad (21)$$

The  $r_i$  values are as follows:

$$r_i = \begin{cases} 1, & z_i > z_j \\ 0, & \text{else} \end{cases} \quad (22)$$

2. The average and variance of the sequence are calculated

$s_1, s_2, \dots, s_n$  is a separate sequence, and the average and variance are calculated as follows:

$$\begin{cases} E(s_k) = \frac{k(k-1)}{4} \\ var(s_k) = \frac{k(k-1)(2k+5)}{72} \end{cases}, k = 2, 3, \dots, n, (k = 100) \quad (23)$$

3. The Calculation of  $UF_k$

Based  $s_k$  on what we can see is the tiredness of all the numbers of values greater than the time value. The defined statistic  $UF_k$  is:

$$UF_k = \frac{(s_k - E(s_k))}{\sqrt{var(s_k)}}, k = 1, 2, \dots, n, (n = 100) \quad (24)$$

4. The Calculation of  $UB_k$ :

$$UB_k = -UF_k, k = n, n-1, \dots, 1 \quad (25)$$

The four main ingredients will be calculated and  $UF_k$   $UB_k$  on the year will be as follows:

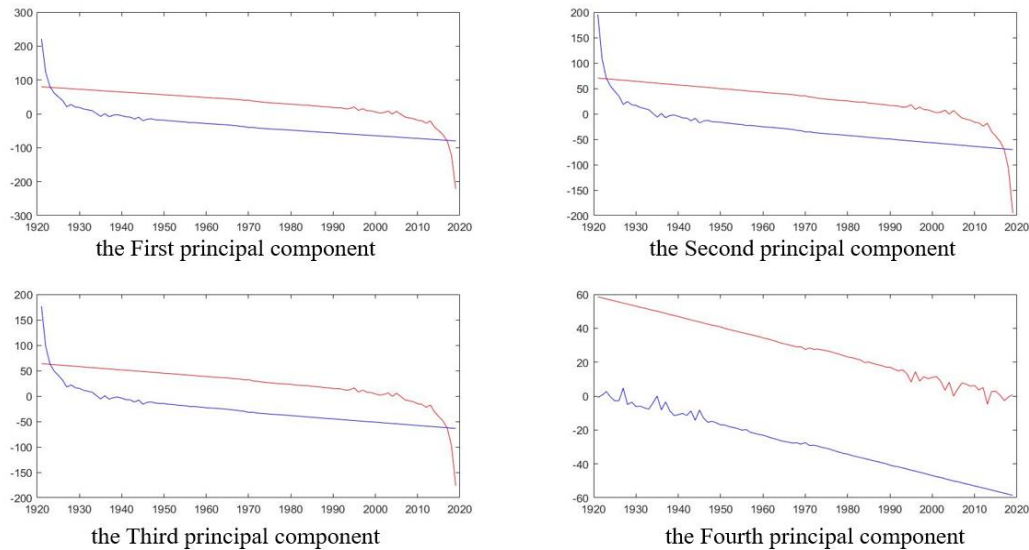


Figure 16: Time series analysis

The year in which the first three main ingredients mutated was 1923, 2017.

### 6.3 Influencers of major changes

Near 1923 we identified more influencers of suitors as follows:

**Hank Williams**

Country

184



Hiram "**Hank**" Williams (September 17, 1923 – January 1, 1953) was an American singer-songwriter and musician. Regarded as one of the most significant and influential American singers and songwriters of the 20th century, Williams recorded 35 singles (five released posthumously) that reached the Top 10 of the Billboard Country & Western Best Sellers chart, including 11 that ranked number one (three posthumously).

**Billie Holiday**

Vocal

106



Eleanora Fagan (April 7, 1915 – July 17, 1959), known professionally as **Billie Holiday**, was an American jazz and swing music singer with a career spanning 26 years. Nicknamed "Lady Day" by her friend and music partner Lester Young, Holiday had an innovative influence on jazz music and pop singing. Her vocal style, strongly inspired by jazz instrumentalists, pioneered a new way of manipulating phrasing and tempo. She was known for her vocal delivery and improvisational skills.

The number in the upper right corner is the number of direct suitors.

The information we've identified around 2017 for those with more influencers is as follows:

**Frank Ocean**

R&amp;B

3

**Frank Ocean** (born October 28, 1987)[3] is an American singer, songwriter, record producer, photographer, and visual artist. Recognized for his idiosyncratic musical style, introspective and elliptical songwriting, unconventional production techniques, and wide vocal range, Ocean is considered to be among the most acclaimed artists of his generation. Music critics have credited him with revitalizing jazz and funk influenced R&B, as well as advancing the genre through his experimental approach.[4] He is considered a representative artist of alternative R&B.

**Flume**

Pop/Rock

2



Harley Edward Streten[2] (born 5 November 1991), known professionally as **Flume**, is an Australian musician, DJ and record producer.[1] His self-titled debut studio album, *Flume*, was released on 9 November 2012 to positive reviews, topping the ARIA Albums Chart and reaching double-platinum accreditation in Australia. Flume is regarded as the pioneer of future bass who helped popularize the genre.

## 7. Problem 6

### Problem Analysis

The sixth Problem asks us to analyze the influence process of music evolution over time in a genre, to identify indicators that reveal dynamic influencers, and to explain how genres or artists change over time.

First of all, we choose the Blues genre as the research object, and draw the line chart of the indicators in the genre about the age to observe the dynamic changes of the indicators. In order to find out the indicator that reveals the dynamic influencer, we will analyze the trend of each indicator over time and select an indicator that reflects the overall rate of change -- the inclination rate. The size of the absolute value of the inclination

can reflect the degree of trend change, and the greater the degree of change, the more obvious the dynamic change of the indicator. Select the indicator with a large absolute yield as the dynamic index of the influencer.

Secondly, for the school over time changes we have made a detailed description in the third Problem, in this Problem do not repeat.

Finally, we're going to show how musicians have changed with the times, and we've chosen to analyze the Beatles band's data. Where this data is obtained by using online data crawling.

## 7.1 The establishment of the model

### 7.1 Dynamic changes in the Blues genre

First, the various types of information for the Blues genre are summarized, some of which are shown in the table below:

Artist name	Year	danceability	energy	...	count
Elmore James	1930	0.45725	0.70285	...	40
Big Joe Williams	1930	0.634	0.2205	...	4
Lightnin' Hopkins	1930	0.562917	0.379125	...	48
Professor Longhair	1930	0.5885	0.621167	...	12
Joe Bonamassa	1980	0.393857	0.578714	...	7
Chris Thomas King	1980	0.52	0.0204	...	2
Susan Tedeschi	1990	0.508111	0.337222	...	9

Table 15

The chart of the indicators in the table above on the changes in age is as follows:

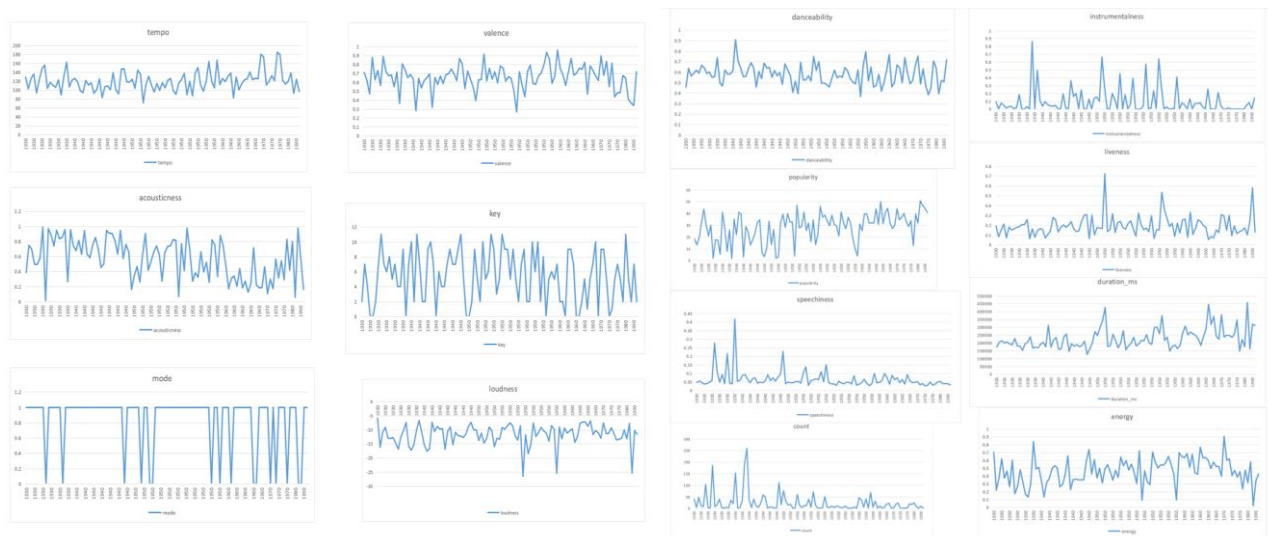


Figure 17: Line chart

Looking at the image above, we find that mode does not indicate dynamic changes. For other indicators, we first average the indicator data on each time node, especially for key, count two types of indicator data we do average and then round down. The index data of each node are calculated by tilt rate, and the calculation formula is referenced above. The yield value is shown in the table below:

danceability	energy	valence	tempo	loudness	key
-8.2E-05	-0.00051	-0.00234	-0.03409	0.002667	-0.01564
acousticness	instrumentalness	liveness	speechiness	popularity	count
-0.0053405	-0.001194935	0.002144	-0.000951	0.395886	-0.52075

Table 16

Select an indicator with a large absolute inclination as a dynamic indicator of the influencer, which is: popularity, count, tempo.

## 7.2 Dynamic changes in the Beatles band

The captured data is presented as an image:

Changes in band personnel:

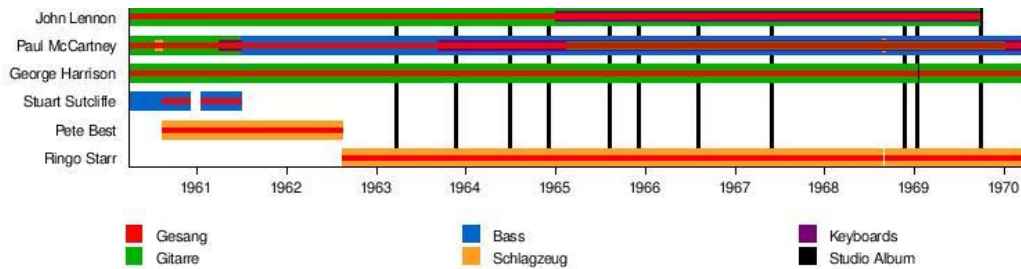


Figure 18: Team member

Album updates and awards:

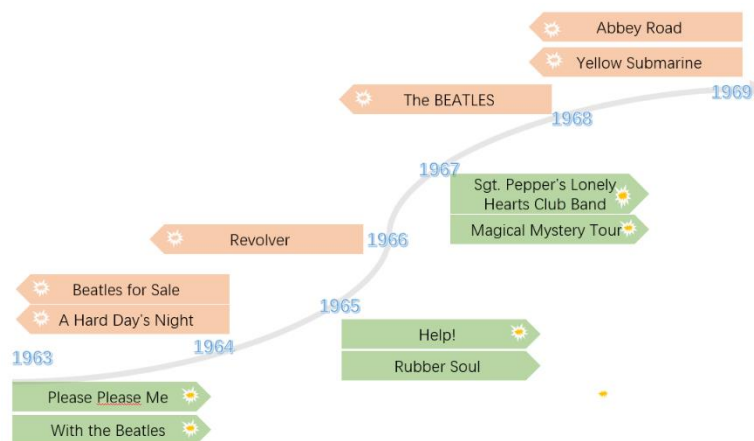


Figure 18: Time line

## 8. Problem7

### Problem analysis

Problem 7 requires us to express the impact of music on the environment, society, and culture. We consulted relevant literature and chose to illustrate the mutual influence between culture, society and music genre by giving a case of music genre.

We choose Rap as an example to illustrate the mutual influence between culture, society and music.

Rap genre originated in 1970. It is an important part of black music and a keynote of street culture. It started on the street, but it developed rapidly in 1980 and 1990. Rap genre moved from an unknown environment to a big stage of world attention. Today, Rap genre is very popular even in China across the ocean. This, in turn, makes people pay more attention to discrimination in society.

Another famous case is The Woodstock Festival. In 1969, the country was deep into the controversial Vietnam War, a conflict that many young people vehemently opposed. It was also the era of the civil rights movement, a period of great unrest and protest. Woodstock was an opportunity for people to escape into music and spread a message of unity and peace. Anti-success itself, anti-consumerism, anti-commercial society, just like the documentary at Woodstock Music Festival said. Then why follow the rules of the game that the adults said? Why do we have to fight, why should we make money for promotion and buy a house? What to do about the American dream that didn't belong to me?



The era has made music, and music has marked the era.



Figure 19

### **The manifestation of the evolution of social existence in music**

With the development of technology, people can more easily hear the performances of people on the other side of the world. On the one hand, this promotes the exchange and development of different genres, and on the other hand, it also improves the similarity between different genres. This is also reflected in the data.

## **9. Strengths and Weaknesses**

### **Advantages:**

1. We use the **sen's** trend test method to build a model, reasonably use the characteristics of the uniform distribution of time data nodes, convenient and efficient to obtain the trend of musical characteristics change;
2. In this Problem, we use the Mann-Kendal model, which can find the year in which the musical feature mutated and make a reasonable analysis of the conceptual year during the change of musical characteristics.

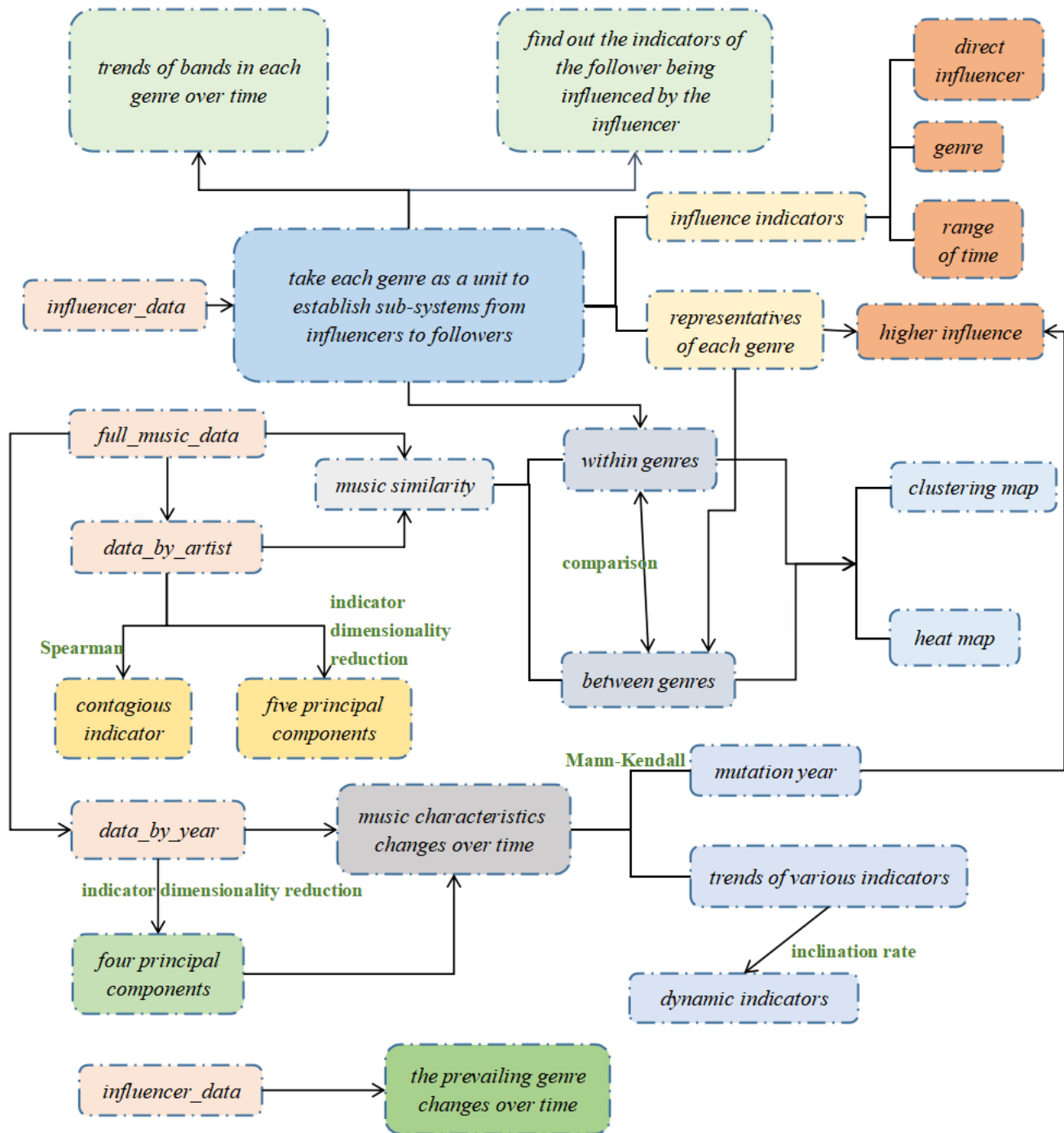
### **Disadvantages:**

1. In the description of the model, because the amount of data is too large, we can't show it all, we can only select a portion of the special amount of data to present.

## **Reference**

- [1] Robeyns I . SEN'S CAPABILITY APPROACH AND GENDER INEQUALITY: SELECTING RELEVANT CAPABILITIES[J]. Feminist Economics, 2003.
- [2] Gocic M , Trajkovic S . Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia[J]. Global & Planetary Change, 2013, 100(JAN.):172-182.
- [3] Spearman C . The proof and measurement of association between two things. By C. Spearman, 1904.[J]. American Journal of Psychology, 1987, 100(3-4):441-471.
- [4] Perkins, Jean M . The principal component analysis of genotype-environmental interactions and physical measures of the environment[J]. Heredity, 1972, 29(1):51-70.
- [5] [ Sue J. Lin, Lu I J , Lewis] C . Grey relation performance correlations among economics, energy use and carbon dioxide emission in Taiwan[J]. Energy Policy, 2007.

## One-page Document





## Appendix

Code:

```
#gray
clc,clear
a=xlsread('grey_data.xlsx','Sheet1','A1:C19');

for(i=1:3)
    a(i,:)=(max(a(i,:))-a(i,:))/(max(a(i,:))-min(a(i,:)));
end
[m,n]=size(a);
cankao=max(a)';
t=repmat(cankao,[1,n])-a;
mmin=min(min(t));
mmax=max(max(t));
rho=0.5;
xishu=(mmin+rho*mmax)./(t+rho*mmax);
guanliandu=mean(xishu);
%[gsort,ind]=sort(guanliandu,'descend');
xlswrite('grey_data.xlsx',cankao,'cankao');
xlswrite('grey_data.xlsx',guanliandu,'guanliandu');
xlswrite('grey_data.xlsx',xishu,'xishu');
xlswrite('grey_data.xlsx',t,'t');

#principal main
clc,clear
X=xlsread('data_by_artist.csv','data_by_artist','C2:P5855');
z=zscore(X)
M=cov(z)
[V,D]=eig(M);
d=diag(D);
eig1=sort(d,'descend')
v=fliplr(V)
S=0;
i=0;
while S/sum(eig1)<0.85
    i=i+1;
    S=S+eig1(i);
end
res1=eig1./S;
NEW=z*v(:,1:i)
W=100*eig1/sum(eig1)
figure(1)
pareto(W);
xlswrite('T2zhu.xlsx',D,'D');
xlswrite('T2zhu.xlsx',M,'M');
xlswrite('T2zhu.xlsx',v,'v');
xlswrite('T2zhu.xlsx',V,'V1');
```