1 Method's Current Usage

1.1 To Which World Regions Does the Valence–Dominance Model of Social Perception Apply?

Authors: Jones et al.
Published: Nature Human Behaviour, 2021
Summary: This study evaluated the cross-cultural applicability of the valence–dominance model of social perception. PCA was used to analyze social evaluations of faces across 11 regions and 41 countries, identifying valence and dominance as primary dimensions. Results indicated that while the model generalized well, regional variations emerged, reflecting cultural differences. Congruence coefficients and exploratory factor analysis validated the PCA findings.
Takeaway: The study's use of PCA for dimensionality reduction and validation provides a framework for analyzing global patterns. This inspires my project to explore music track features, identifying universal structures in global music distributions.

1.2 The Structure and Development of Explore–Exploit Decision Making

Authors: Madeline B. Harms et al.
Published: Cognitive Psychology, 2024
Summary: This study explored developmental differences in explore-exploit decision-making through tasks like bandit games and foraging simulations. PCA identified two components: random exploration (non-goal-directed) and directed exploration (purposeful uncertainty reduction). Results showed that adolescents rely more on random exploration, while adults demonstrate strategic directed exploration, highlighting developmental changes in decision strategies.
Takeaway: The use of PCA to uncover latent components and analyze task-related variability is highly relevant for my project. It demonstrates that I can first use PCA to identify core components of music features, then examine music difference and evolution among components.

1.3 Public Concerns and Attitudes Towards Autism on Chinese Social Media Based on K-Means Algorithm

Authors: Qi Zhou et al.
Published: Scientific Reports, 2023
Summary: This study analyzed autism-related discussions on Zhihu using TF-IDF for feature

extraction, PCA for dimensionality reduction, and K-Means for clustering. The elbow method determined the optimal cluster number as four, grouping topics into self-experiences, societal views, caregiver stresses, and informational content. Results highlighted growing public awareness, stereotypes, and challenges faced by individuals with autism and their families.
Takeaway: The integration of PCA and K-Means clustering demonstrates an effective methodology for uncovering similarity patterns in large datasets. This approach aligns with my project to analyze global music distribution, enabling feature reduction and meaningful clustering of tracks.

2 My Experimentation

2.1 Goal

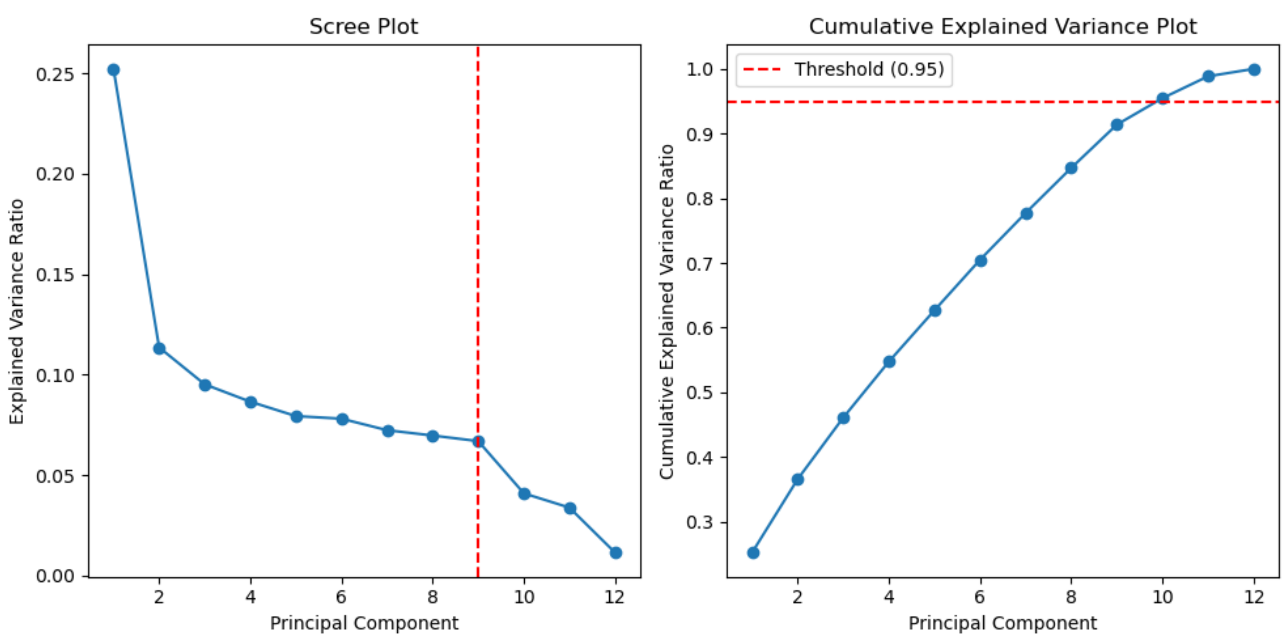
I performed **dimensionality reduction** and **clustering** among global tracks over the past century (1920-2020).
The PCA space can be used to visualize the global music distribution, and the KMeans clustering can be used to briefly identify the similarity patterns among tracks without extensive pairwise calculation.

2.2 Challenges

1. **Parameterization**: PCA and KMeans both require selecting the number of components or clusters. Some data-driven criteria is needed to determine the optimal number.
2. **Interpretability**: To make sense of the PCA results, I need to "decode" the principal components into the original features; similarly, to understand the clustering results, I need to examine the top songs in each cluster to get a sense of the cluster's characteristics.

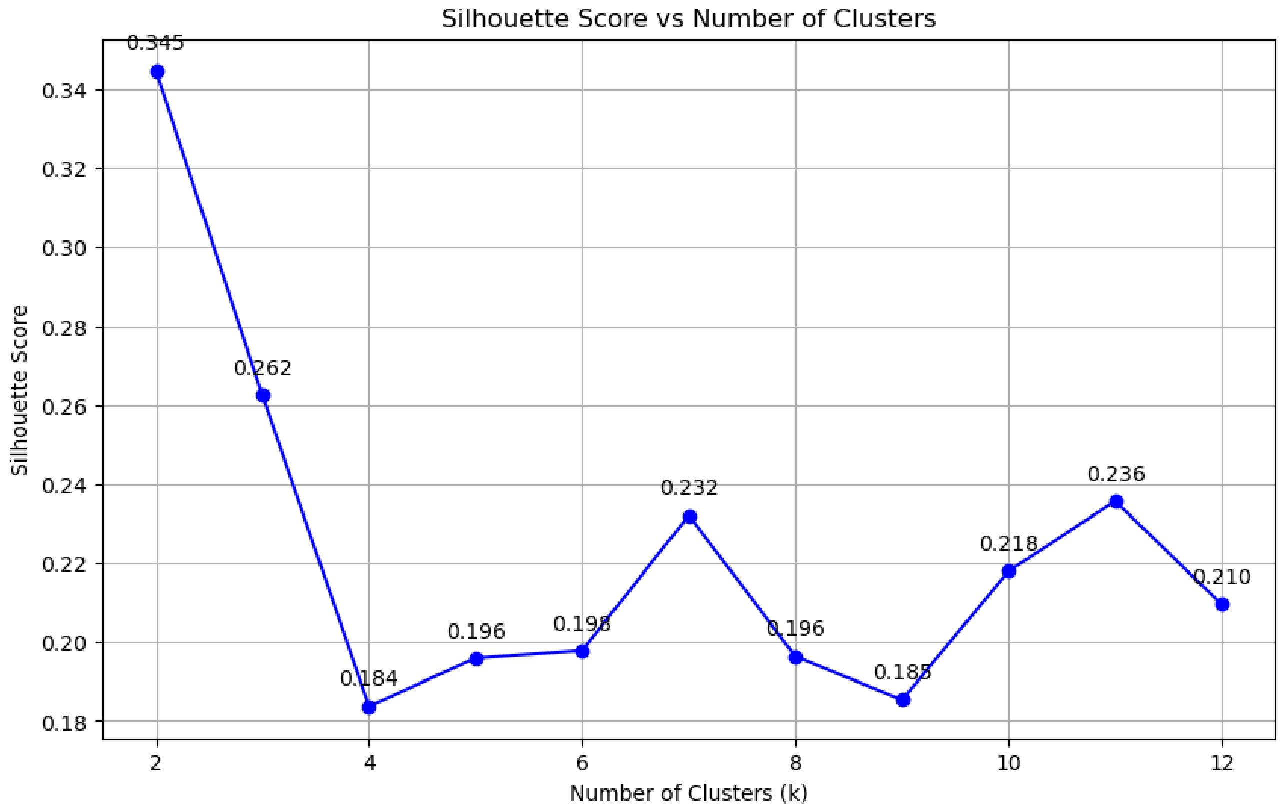
2.3 Process

1. I first scaled the features, then performed PCA to reduce the dimensionality of the data.
2. According to the scree plot and cumulative explained variance, I found that the first 9 components can explain 95% of the total variance. Thus I chose the first 9 components as the input of KMeans clustering.



3. Then I performed KMeans clustering on the PCA-transformed data. Through the Silhouette score plot, I found that the optimal number of clusters is 2. Thus I chose the number of

clusters as 2.



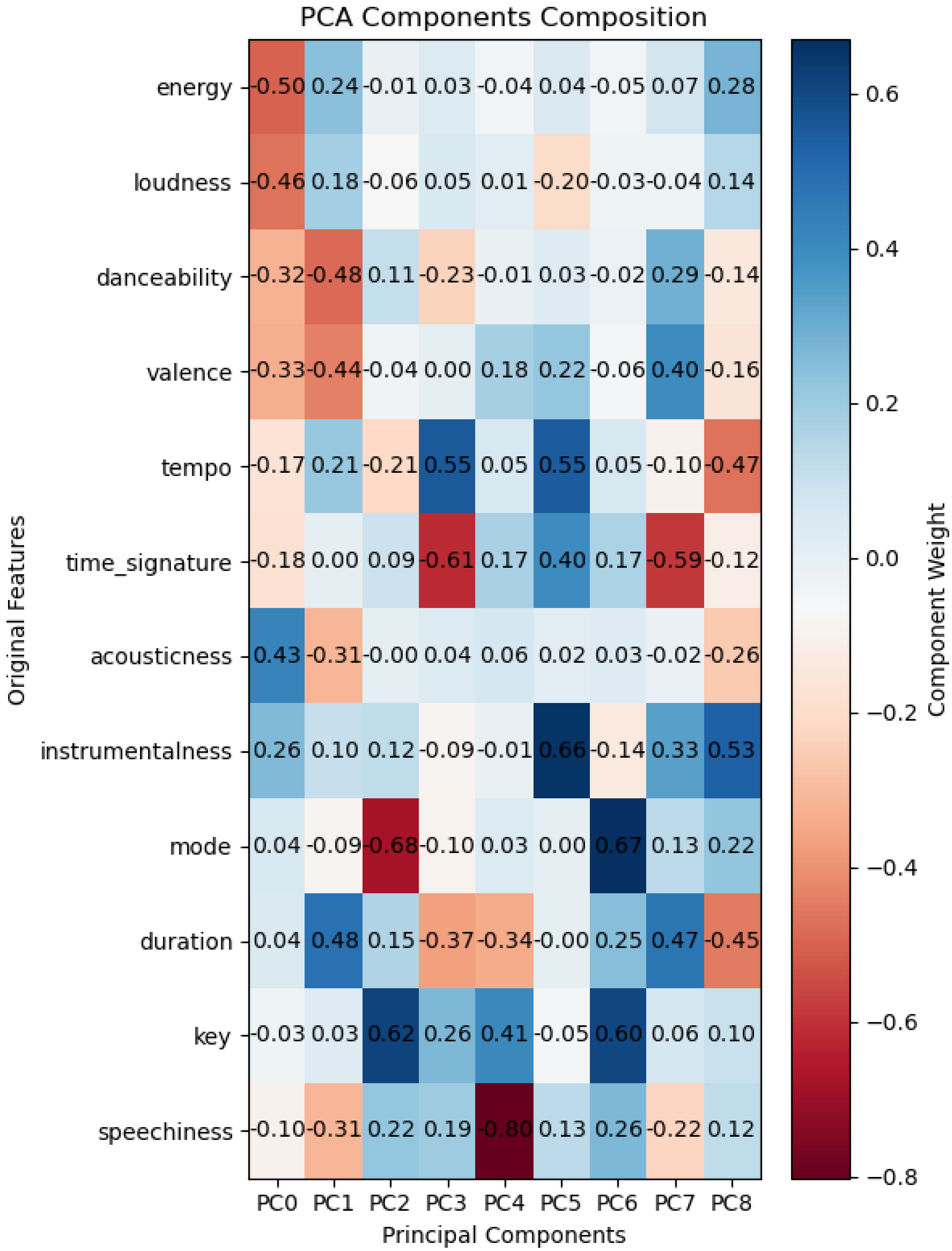
4. Visualization:
    - PCA: For interpretability, I plot the correlation between the principal components and the original features. Blue means positive correlation, red means negative correlation.
    - KMeans: I plot the distribution of the tracks in the PCA space (three 2D space of combinations of the first three principal components). Clusters are indicated by different colors. I also examnied the top 10 songs in each cluster.

3 Findings and Implications

3.1 PCA Findings

PCA reduced the dataset to 9 principal components (PCs). Through the scree plot, PC 1 is the most important component. And its pattern is also the most interpretable.

The heatmap shows how original features contribute to each PC:



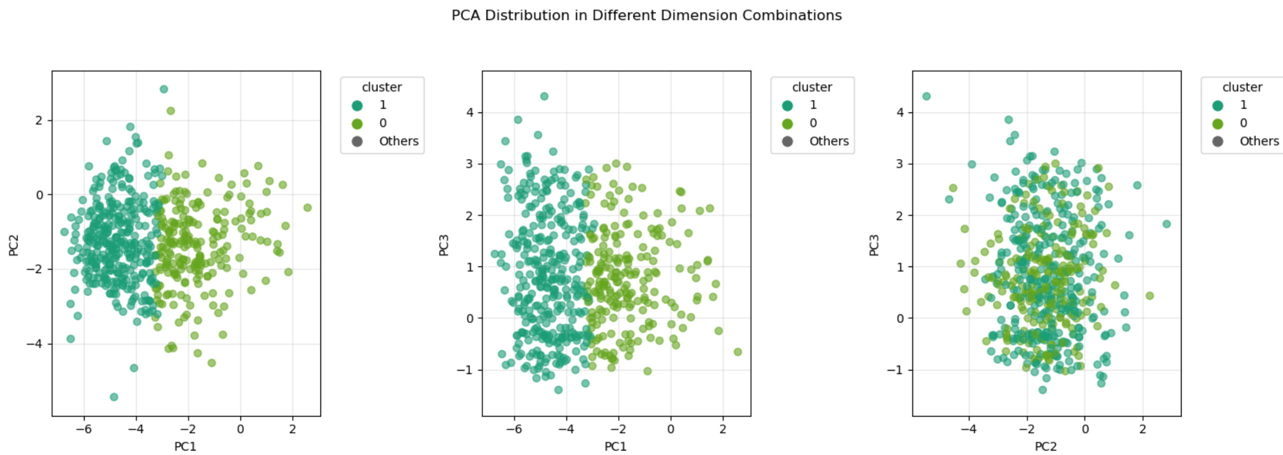
- PC1: Strongly influenced by negative weights on energy, loudness, danceability and valence, suggesting it captures the emotional intensity and energy level of tracks. It likely distinguishes between high-energy, emotionally upbeat music (e.g., pop/electronic) and softer, acoustic or melancholic styles.

- PC2: Dominated by positive contributions from duration and negative weights on speechiness, danceability and valence, potentially capturing rap music characteristics.
- PC3: Highlights mode and key, relating to the melodic structure and tonality.

3.2 KMeans Findings

Optimal k=2 clusters were identified based on PCA-transformed data:

- Cluster 0 and Cluster 1 are primarily separated along PC1, suggesting emotional energy is the key differentiating factor for tracks. Spatially, Cluster 0 leans toward higher PC1 values, while Cluster 1 leans lower.



- Cluster 0 tracks are more acoustic, slower-paced, and emotionally reflective. For example, “Heartbreak Anniversary” by Giveon, “Someone You Loved” by Lewis Capaldi, “All of Me” by John Legend.

```
+-------+---------------------+----------------------------+----------+
|cluster|name                 |artists                     |popularity|
+-------+---------------------+----------------------------+----------+
|0      |Heartbreak Anniversary|['Giveon']                 |94.0      |
|0      |you broke me first   |['Tate McRae']              |91.0      |
|0      |Someone You Loved    |['Lewis Capaldi']           |90.0      |
|0      |lovely (with Khalid) |['Billie Eilish', 'Khalid']|89.0      |
|0      |Arcade               |['Duncan Laurence']         |89.0      |
|0      |Heather              |['Conan Gray']              |89.0      |
|0      |Afterglow            |['Ed Sheeran']              |88.0      |
|0      |All of Me            |['John Legend']             |87.0      |
|0      |everything i wanted  |['Billie Eilish']           |87.0      |
|0      |Train Wreck          |['James Arthur']            |87.0      |
+-------+---------------------+----------------------------+----------+
```

- Cluster 1 tracks exhibit high energy, upbeat rhythms, and electronic influences. For example, “Save Your Tears” by The Weeknd, “Blinding Lights” by The Weeknd, “The Business” by Tiësto.

```
Cluster 1:
+-------+------------------+-----------------------+----------+-
|cluster|name              |artists                |popularity|
+-------+------------------+-----------------------+----------+-
|1      |Save Your Tears   |['The Weeknd']         |97.0      |
|1      |telepatía         |['Kali Uchis']         |97.0      |
|1      |Blinding Lights   |['The Weeknd']         |96.0      |
|1      |The Business      |['Tiësto']             |95.0      |
|1      |Bandido           |['Myke Towers', 'Juhn']|94.0      |
...
|1      |Good Days         |['SZA']                |93.0      |
|1      |Watermelon Sugar  |['Harry Styles']       |92.0      |
+-------+------------------+-----------------------+----------+-
```

The examine of the top 10 songs in each cluster confirms our interpretation of the PCA results, further suggesting the importance of emotional energy in distinguishing between music styles.

3.3 Summary

The analysis reveals that PCA effectively highlights **emotional energy** (PC1) as the key factor distinguishing music styles, supported by KMeans clustering into two groups: reflective acoustic tracks (Cluster 0) and high-energy electronic tracks (Cluster 1).
These findings demonstrates the significance of energy and mood in shaping musical categorizations and listener preferences.