RiSK: Analysis of the Changes in Popular Music in the World

Keena Gao, Rhea Tejwani, Sue Zhang

STA199 Fall 2020

Topic

- Our project is an **analysis of the popular music of the world over time.**
- As students who frequently listen to music, we want to analyze the trends of popular music over time

Our research questions:

- Does our data provide sufficient evidence that the proportion of songs in non-English languages has increased from the time period of 1956-1989 to the time period of 1990 to present?
- Does our data provide sufficient evidence that the standard deviation in Acousticness has increased over time?
- **The goal of our report is to observe how popular music has shifted to be more diverse in the characteristics of non-English genres or Acousticness over time.**

Data

- Curated by Sumat Singh (@iamsumat) on Kaggle
- Contains variables that measure various characteristics of the 1,994 most popular songs in the world over the years 1956 to 2019 on Spotify
- Scraped from the playlist "Top 2000s" by the user PlaylistMachinery (@plamere) using Selenium with Python.

The variables that we explored in the raw data set:

- **Top Genre**
- **Acousticness**
- **Year**

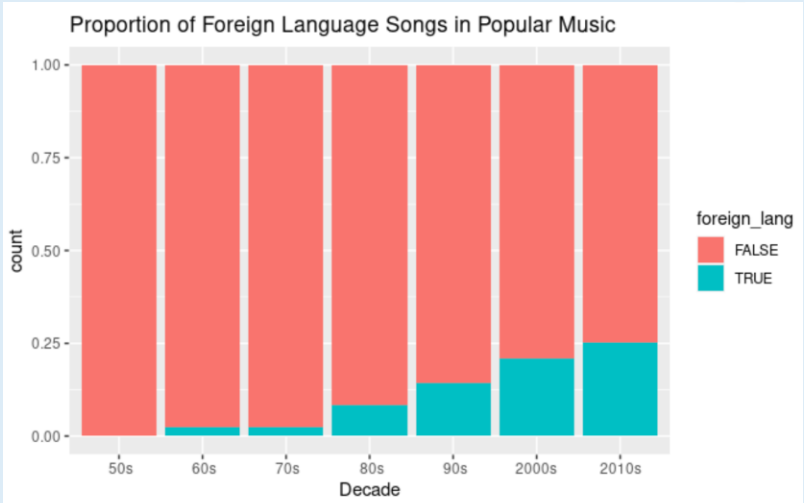
Mutated variables:

- **decade**
- **period**
- **foreign_lang**
- **pop**
- **sd**
- **yearSince1956**

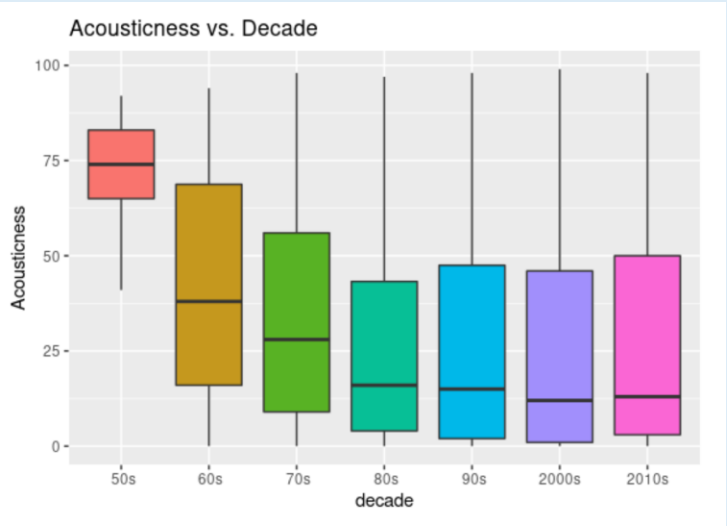
Exploratory Data Analysis

```
## # A tibble: 6 x 3
## # Groups:   decade [6]
##   decade  pop       n
##   <chr>   <chr> <int>
## 1 " 60s"  pop      12
## 2 " 70s"  pop      37
## 3 " 80s"  pop      47
## 4 " 90s"  pop      39
## 5 "2000s" pop      74
## 6 "2010s" pop     138
```

The count of songs in the "pop" genre have generally increased by decade.



The proportion of popular songs in non-English languages have increased by decade.



The median value for Acousticness is high in the 50s, 60s, and 70s compared to the remaining decades.

Simulation-Based Testing: Non-English Songs

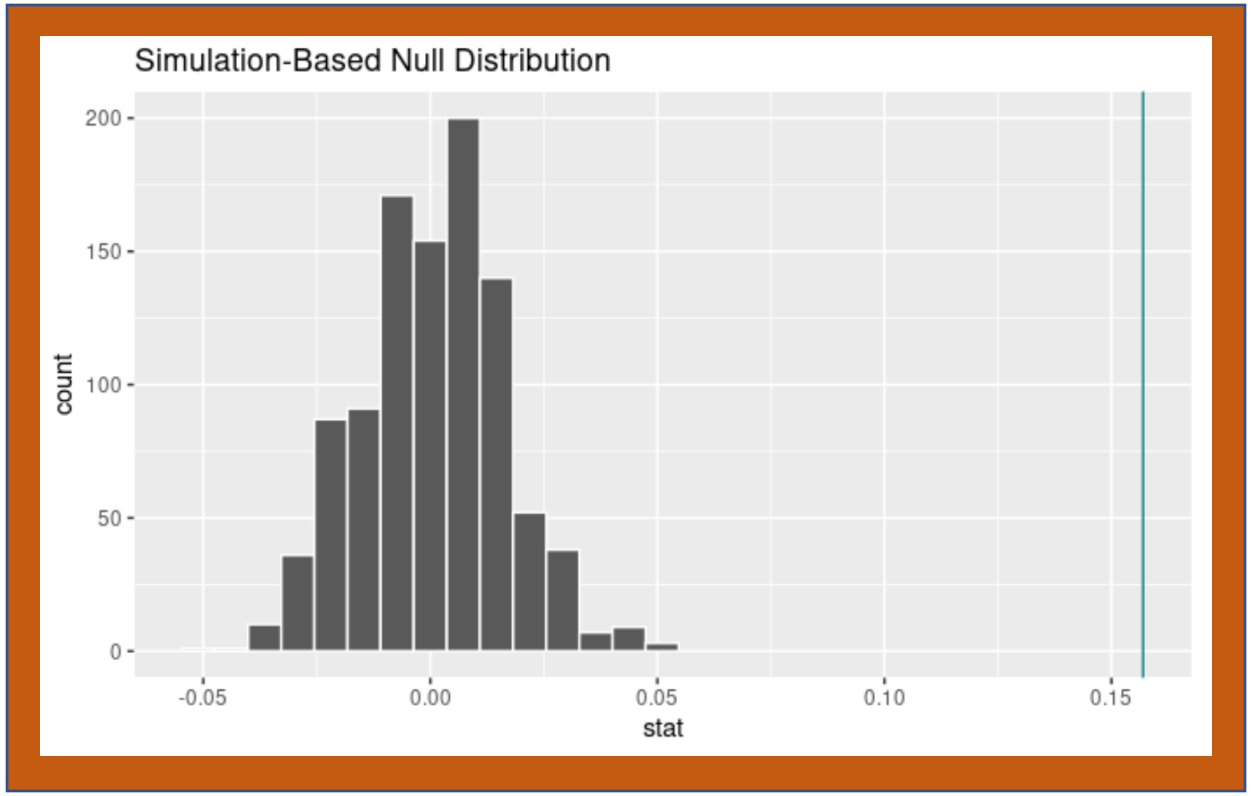
H0: The proportion of popular songs in non-English languages in the "1990s and After" time period is equal to the proportion of songs in non-English languages in the "Pre 90s" time period

Ha: The proportion of popular songs in non-English languages in the "1990s and After" time period is greater than the proportion of popular songs in in non-English languages in the "Pre 90s" time period

$\alpha = 0.05$

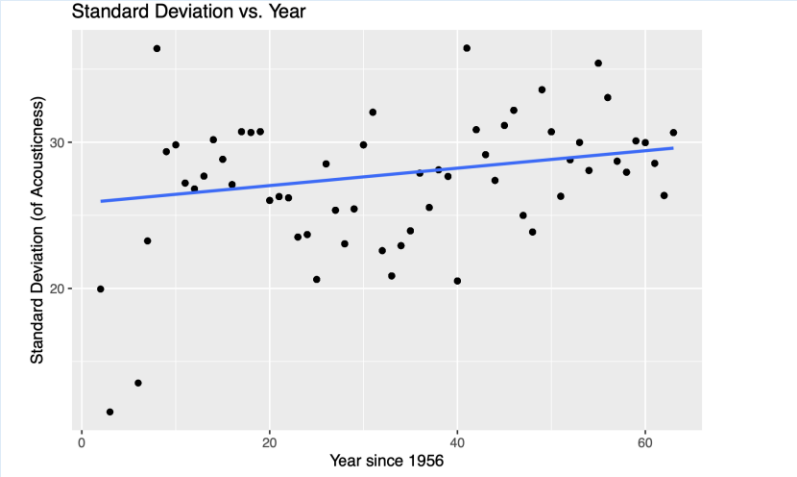
```
## # A tibble: 1 x 1
##   p_val
##   <dbl>
## 1     0
```

There is sufficient evidence to conclude that the proportion of popular foreign language songs after 1989 is greater than the proportion of popular foreign language songs in before 1990.

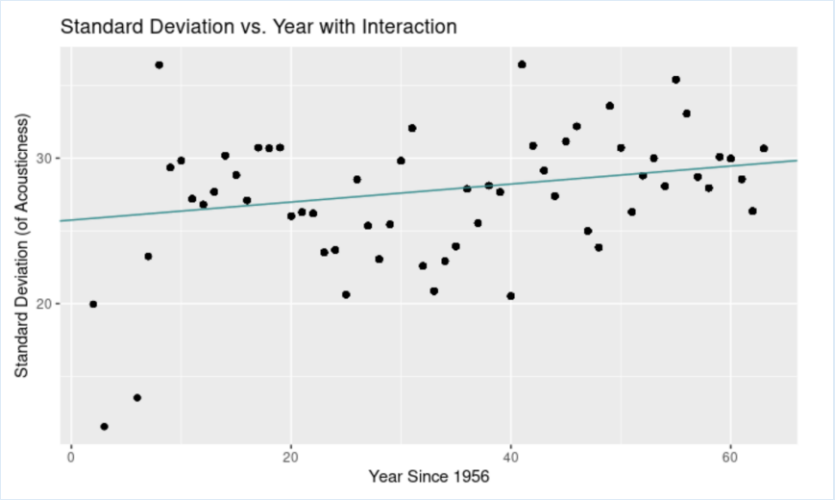


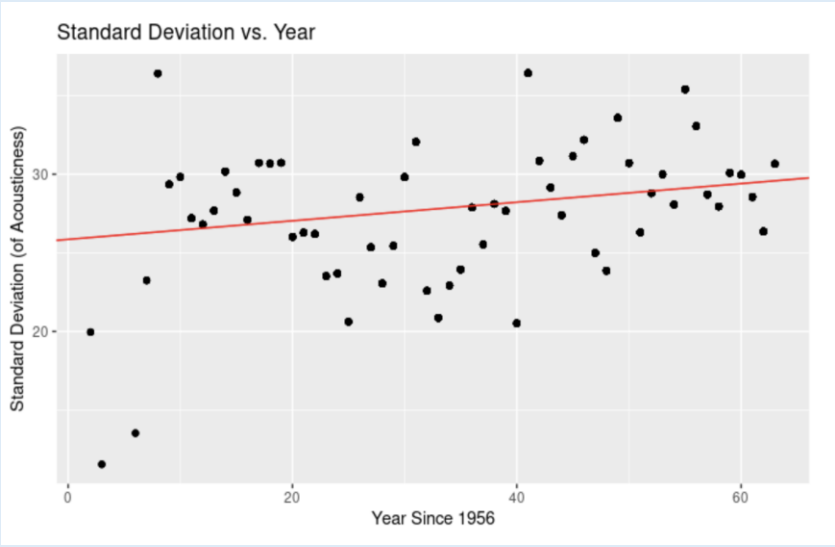
Linear Modeling: Standard Deviation of Acousticness

Standard Linear Model



Linear Model with Interaction Effects with the Genre "Pop"





Both models show that the standard deviation for Acousticness has generally increased over time.

Conclusion

The proportion of non-English songs in popular music and the standard deviation of Acousticness in popular music has increased over time.

Future Changes

- Find a data set with popular songs from more years
- Explore different regression models that will better fit the data
- Use hypothesis tests to compare individual years or smaller time frames