Gender Representation in Music

Section 2 Group 2: Jarid Haslinger, Ryan Kilpadi, Sriram Soundarapandian

Fall 2022

Introduction

Our initial motivation for choosing this topic stems from our interest in learning more about how the wealth of a country influences gender representation in the music industry. We wanted to learn more about whether a certain gender is over or under-represented in specific countries. There are many industries where gender misrepresentation is prevalent and we are hoping to uncover more information regarding the music industry and therefore more about how the creative space is a form of gender representation or misrepresentation. In addition, spitting up our sampled countries into sections of "higher" and "lower" GDP per capita allows us to analyze how the development of a country connects to gender representation in the music industry.

Furthermore, musical choices are strong representations of cultural preferences, therefore, understanding gender representation will allow us to learn more about how GDP might influence these choices. In many ways, top musical artists are role models and represent certain societal preferences. Therefore, countries with lower GDP per capita may have some differences in how easy it is for women or men to experience success in the music industry.

We chose a cutoff point at 22,000 because one widely used measure of whether a country is deemed "developed" is the threshold of GDP per capita of at least 22,000 which has been stated by the World Bank as one of the measures. Therefore, we choose to split the data between countries above and below the per capita GDP of 22,000. Additionally, we analyzed 72 total countries because those were listed on Spotify's top 50 playlists.

We use the following hypothesis tests:

HO: Gender proportions in musical artists and GDP segments are independent, (gender identity does not vary by nationality)

HA: Gender proportions in musical artists and GDP segments are dependent, (gender identity varies by nationality)

This is a two-sided hypothesis test because we are measuring both sides since it is not a greater than/less than hypothesis test. In addition, we will use a significance level of 0.05

Data

For our data collection, we used Spotify API to compile a set of the most popular artists in over 70 countries. This involved writing a Python script to extract the unique artists in each country's "Top 50" playlist, which is curated by Spotify to display the most streamed tracks in a specific country in the past week. The data collection was conducted on November 6th. An interesting extension to this project might explore whether gender ratios of top artists change over the course of a year. We then added GDP per capita information to our dataset, pulling data from the IMF via Wikipedia. The genders of artists were not available through Spotify's API, so we created a script that performed Google searches on all the unique artists, allowing us to quickly enter the genders of artists manually. We recognize that there may be a small error rate due to this method.

Methods

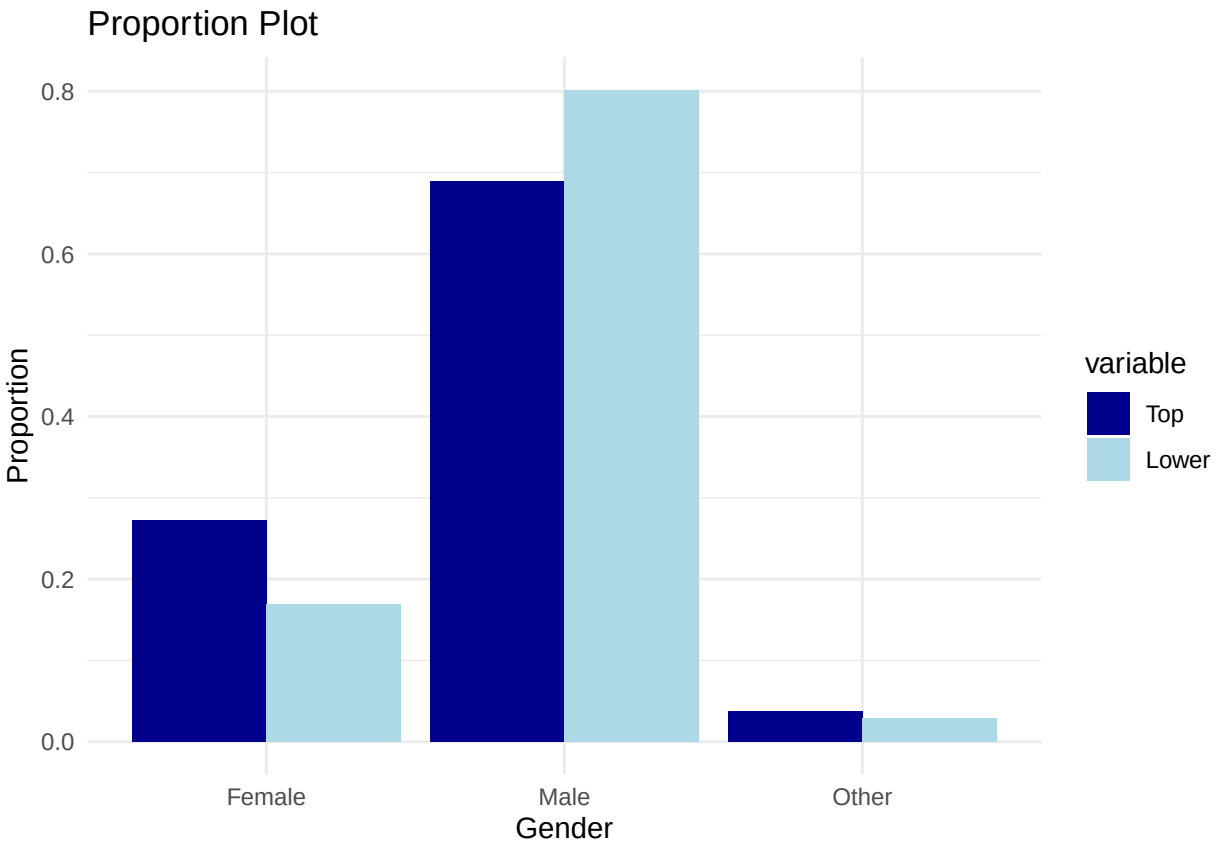
The variables that we're studying are the gender identity of the musical artist which consists of Male, Female, and Other. In addition, we are considering the variable of whether the country falls into the top segment of GDP or the lower segment. These variables are categorical.

We will first create a graph for the proportion of Male, Female, and Other artists in each segment of GDP. This graph will visually show any differences in these proportions. In addition, after finding our chi-squared value and degrees of freedom, we will graph the degrees of freedom graph and where the chi-squared statistic lies on the graph.

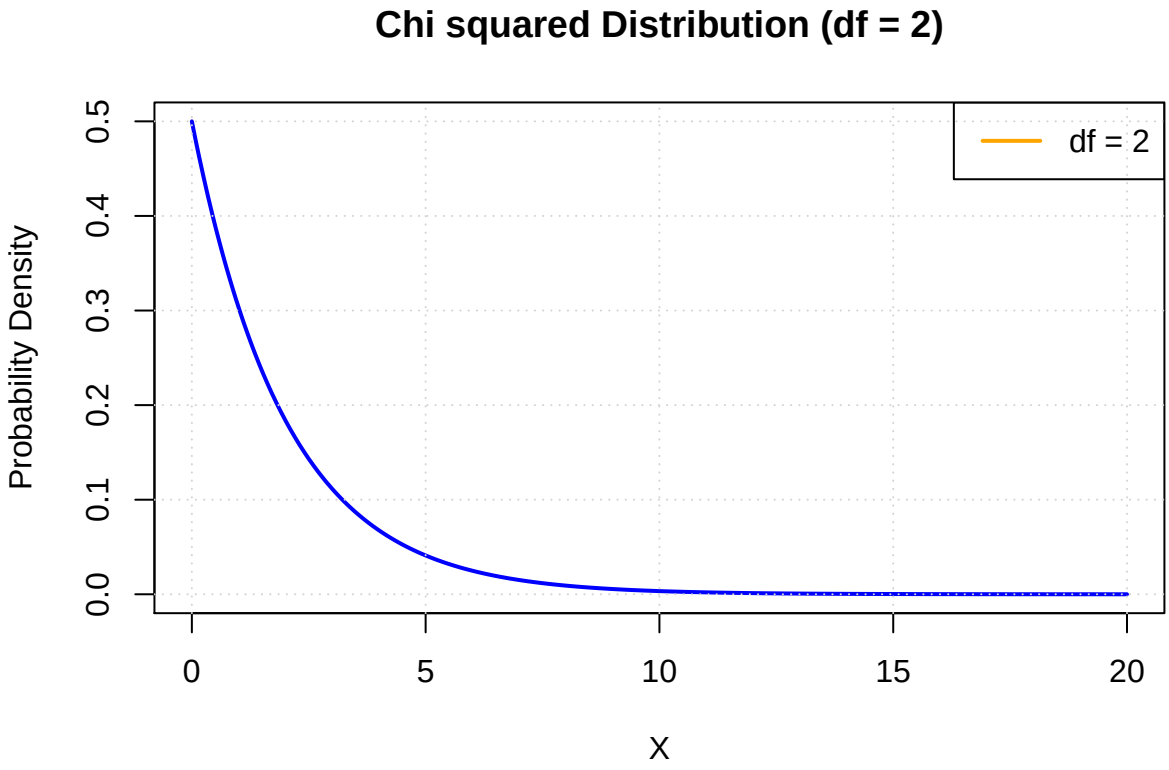
The inferential analysis that we will conduct is the chi-squared test for testing the independence of two categorical variables. This is a two-sided hypothesis test because we are measuring both sides since it is not a greater than/less than hypothesis test. The first step will be to sort the data we have collected into a table and then find the expected value of every combination. Using this, we will calculate the Chi-Squared statistic using a significance level of 0.05 and determine whether the p-value falls below or above our confidence level.

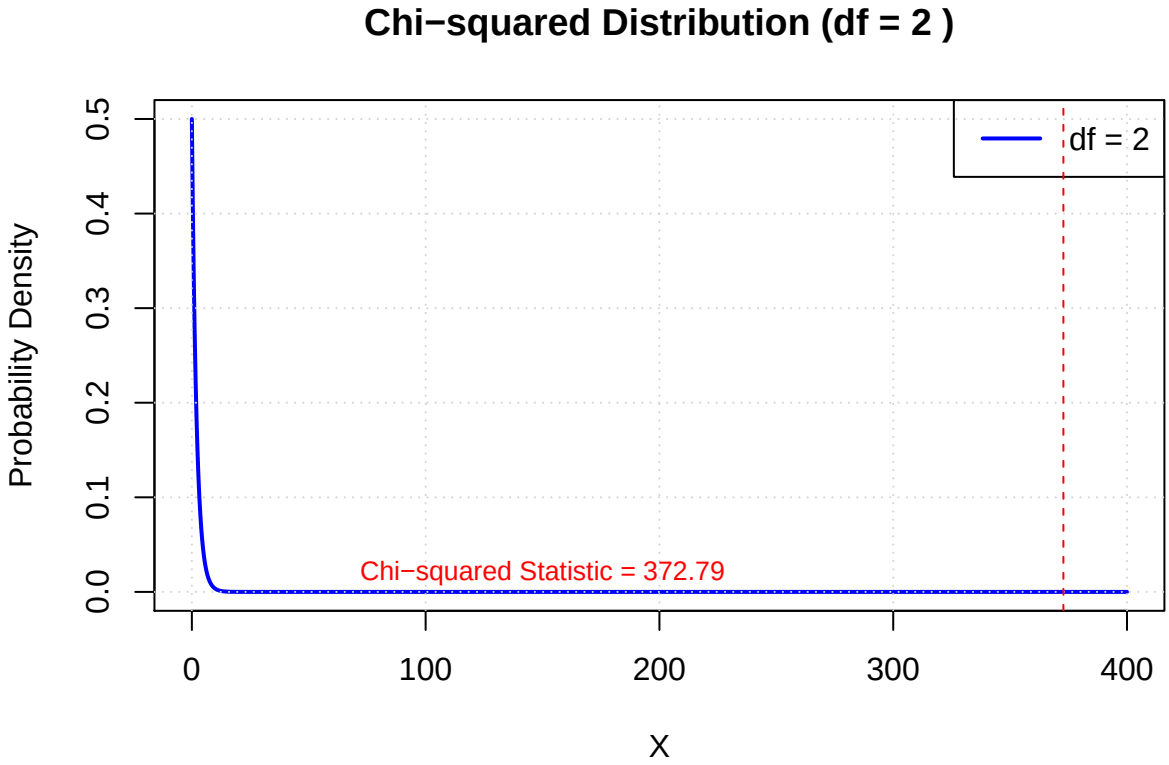
Results

Using the data we collected, we found the following proportions. **The tables for our data which show total counts and expected values will be at the bottom of our report.**



Based on the proportion graph, it initially looks like there is a relatively large difference in proportion between the proportions of male and female musical artists in higher and lower GDP countries. There appears to be a larger proportion of female musical artists in higher GDP countries and a higher proportion of male musical artists in lower GDP countries when compared to each other. Overall, when we look at higher and lower GDP countries individually, we see that both have a much higher proportion of male musical artists compared to female proportions.





Using our data, we calculated our Chi-squared statistic to be 372.79. In addition, we found that our hypothesis test uses 2 degrees of freedom because we have two columns and three rows. In addition, our p-value is 2.2e-16 which is significantly less than 0.05 which is the alpha level that we chose. Therefore, our conclusion is to reject our null hypothesis which means that there is a relationship between gender representation in music and the GDP of a country.

Discussion

Given that our p-value from the Chi-squared test (with a Chi-Squared statistic of 372.79 and degrees of freedom 2) is less than 0.00001, we reject the null hypothesis in favor of the alternate hypothesis, at the 0.05 significance level. This suggests a statistically significant relationship between gender proportions in musical artists and the GDP segments. In practicality, a lower GDP country's correlation with higher male representation in music may indicate deeper gender inequalities in these countries. Less female representation in music may serve as a sign that less female perspectives and voice is present in general.

A potential limitation in our data is that we only surveyed one streaming service. More female artists may be present on different platforms (e.g. SoundCloud) around the world, biasing the results to seem more significant than they are. Another limitation is that this was a cross-sectional study, which means that we cannot apply these results to any other time period other than the single day we collected the data. The time of data collection is a potential confounder. Thus, we cannot make any general claims about the data over any period of time. If we could do this project again, we would use a longitudinal study design, meaning that we would collect data over a long period of time. This would rule out time as a potential confounder.

In conclusion, this data shows an association between gender proportions in musical artists and the GDP segments. This implies that there may be a correlation between gender proportions in musical artists and the GDP segments. However, we cannot draw a causal conclusion because we have not ruled out any sources of bias or confounders. If this association is truly causal, it could possibly indicate deeper gender inequalities associated with GDP. More tests would have to be conducted to provide evidence for this claim.

Tables:

Total Counts	High GDP	Low GDP	Total
Male	1758	842	2600
Female	696	177	873
Other	96	31	127
Total	2550	1050	3600

Expected Values	High GDP	Low GDP	Total
Male	1448.148	1149.852	2598
Female	486.0606	385.9394	872
Other	70.79094	56.20906	127
Total	2005	1592	3597