

**Influence of Individuals' Demographics and Engagement with Music on Music  
Familiarity Scores**

Niva Manchanda

Department of Biostatistics, University of Kansas

PSYC 840: Linear Regression

Dr. Isuzu Ratnayake

## **Abstract**

This study investigates the influence of demographic variables and engagement with music on participants' music familiarity scores, aligning with prior research on cultural influences and music preferences. The results of a multiple linear regression revealed that maximum age and engagement with music significantly predicted participants' music familiarity scores, demonstrating a positive association between higher age, greater engagement with music, and increased music familiarity. Gender, minimum age, and country of origin showed no substantial predictive effect. These findings underline the pivotal role of age and active engagement with music in determining familiarity scores, offering insights into the nuanced relationship between demographic factors, music engagement, and familiarity in cultural contexts.

**Keywords:** demographics, music familiarity, diverse musical tastes

## **Introduction**

Various research studies have revealed that demographic factors shape music preferences and familiarity with artists and tracks. For instance, a large-scale, cross-country study examined the relationship between cross-country differences in artist, album, and genre listening frequencies and the cross-country distance in geographical, socio-economic, linguistic, cultural, and friendship connections. The study found that music listening is an inherently cultural behavior, which may be shaped by users' backgrounds and contextual characteristics (Liu, Hu, & Sched, 2018). Another study on understanding musical diversity via online social media found that age and gender positively affect diversity, indicating that older individuals and males tend to have more diverse musical tastes (Park et al., 2015). Kowald et al. (2021) investigated the correlation between characteristics of beyond-mainstream music, openness, diversity patterns, and recommendation quality among listeners who enjoy exploring music beyond the mainstream. Their study further highlighted how demographics such as country of origin, age and gender impacted these characteristics.

These studies suggest that demographic variables such as gender, age range, and country of origin can predict music preferences and familiarity with different artists and tracks. Comprehending these demographic-based preferences can be crucial for targeted marketing and promotional strategies. This understanding can help tailor campaigns that resonate with specific demographics, potentially leading to increased engagement and sales. Therefore, the hypothesis aims to explore whether gender, age, country of origin, and engagement with music predict participants' familiarity scores with provided classic and modern popular artists and tracks. This comprehensive analysis will assess the significance of these demographic variables and their relationship with participants' music familiarity levels.

## Methods

The dataset was obtained from Kaggle and contains information collected as part of a research project focused on exploring music listening experiences and preferences among diverse individuals. The data was gathered through focus group interviews and includes details about participants' familiarity with various popular artists and music genres. The dataset allows for investigating questions regarding the diversity of music listening experiences across factors like gender, skin color, or nationality. It provides insights into the level of familiarity with classic and modern artists, and the popularity of different tracks. The dataset contains demographic variables and responses to questions about the participants' familiarity with specific artists and tracks, as well as which artists/tracks they chose for each task. The data also includes scores indicating how familiar the participants were and how much they engaged with specific genres of music overall.

## Results

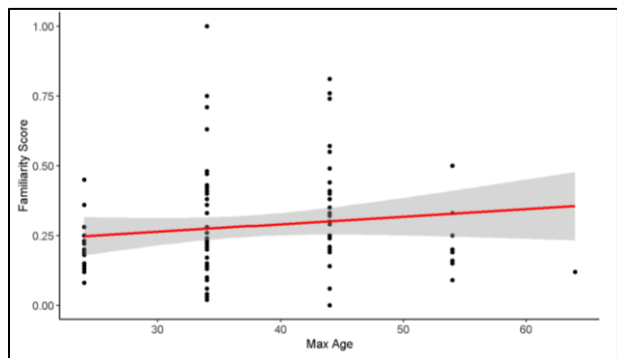
A multiple linear regression analysis was conducted to assess the influence of demographic variables (age, gender, country of origin) alongside engagement with music on participants' music familiarity scores (with classic and modern popular artists and tracks). The predictors included in the model were age (mintage, manager), gender, country of origin, music training, and engagement with music.

The results revealed that participant's maximum age ( $\beta = 0.04$ ,  $t(87) = 1.98$ ,  $p = 0.05$ ) and engagement with music ( $\beta = 0.03$ ,  $t(87) = 2.879$ ,  $p = 0.00$ ), significantly predicted their music familiarity scores. Maximum age demonstrated a positive association, indicating that an increase in age predicted higher familiarity scores (Figure 1). Similarly, engagement with music

positively predicted participants' music familiarity scores (Figure 2). However, other demographic variables, including participants' minimum age, gender, and country of origin did not demonstrate a substantial predictive effect on music familiarity scores with classic and modern popular artists ( $p > 0.05$ ).

**Figure 1.**

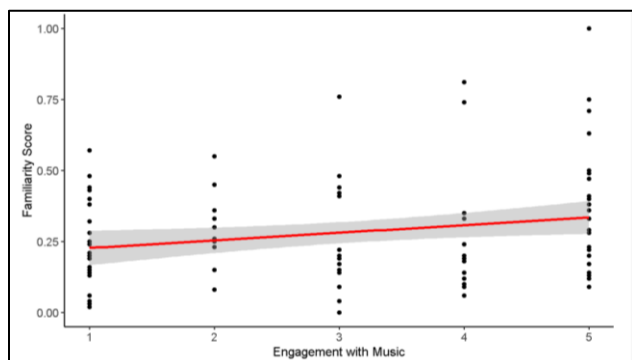
*Participant's maximum age as a predictor of music familiarity scores*



*Note.* Scatterplot of music familiarity score and maximum age including the least squares line (solid) and 95% confidence band (grey).

**Figure 2.**

*Participant's engagement with music as a predictor of music familiarity scores*



*Note.* Scatterplot of music familiarity score and maximum age including the least squares line (solid) and 95% confidence band (grey).

Additionally, the best model fit analysis using the stepwise regression and Bayesian Information Criterion (BIC) revealed that a more parsimonious model with two key predictors: maximum age and engagement with music. The linear regression model comprising 'Managed' and 'Engagement with Music' as predictors indicated that both variables positively predicted music familiarity scores (MaxAge:  $\beta = 0.00$ ,  $t(95) = 1.64$ ,  $p = 0.10$ ; Eng\_with\_music:  $\beta = 0.03$ ,  $t(95) = 2.55$ ,  $p = 0.01$ ). Engagement with Music revealed a significant positive influence on familiarity scores. The overall model explained approximately 7.87% of the variance in music familiarity scores (*adjusted R*<sup>2</sup> = 0.05). The F-statistic for this model was significant ( $F(2, 95) = 4.05$ ,  $p = 0.02$ ), affirming the influence of at least one predictor on the dependent variable.

## Discussion

Contrary to the initial hypothesis, demographic variables such as age, gender, and country of origin did not demonstrate a substantial predictive effect on participants' familiarity scores with classic and modern popular artists and tracks. However, the significant positive association between familiarity scores and engagement with music aligns with prior studies highlighting the influence of active music engagement on familiarity levels with diverse musical content (Sangnark et al., 2021; Madison & Schiölde, 2017). This suggests that while demographic factors may not strongly predict familiarity scores, active engagement with music does play a significant role in enhancing familiarity levels with artists and tracks. These findings have implications for tailored music experiences and marketing strategies, emphasizing the need to prioritize and encourage active music engagement to foster greater familiarity with diverse musical content in individuals across various demographics and cultural backgrounds.

## References

- Kowald, D., Muellner, P., Zangerle, E., Bauer, C., Schedl, M., & Lex, E. (2021). Support the underground: characteristics of beyond-mainstream music listeners. *EPJ Data Science*, 10(1), 14.
- Liu, M., Hu, X., & Schedl, M. (2018). The relation of culture, socioeconomics, and friendship to music preferences: A large-scale, cross-country study. *PloS one*, 13(12), e0208186.
- Madison, G., & Schiölde, G. (2017). Repeated Listening Increases the Liking for Music Regardless of Its Complexity: Implications for the Appreciation and Aesthetics of Music. *Frontiers in Neuroscience*, 11, 147. <https://doi.org/10.3389/fnins.2017.00147>
- Park, M., Weber, I., Naaman, M., & Vieweg, S. (2015). Understanding musical diversity via online social media. *Proceedings of the International AAAI Conference on Web and Social Media*, 9 (1), 308-317.
- Sangnark, S., Autthasan, P., Ponglertnapakorn, P., Chalekarn, P., Sudhawiyangkul, T., Trakulruangroj, M., Songsermsawad, S., Assabumrungrat, R., Amplod, S., Ounjai, K., & Wilaiprasitporn, T., (2021). Revealing preference in popular music through familiarity and brain response. *IEEE Sensors Journal*, 21(13), 14931-14940.

## Appendix (R code)

```
``{r}

library(readr)
library(tidyr)
library(dplyr)

data <- read_csv("C:/Users/User/Downloads/Musicdiv.csv")
data <- data %>%

  filter(Gender != "I prefer not to say", Country != "I prefer not to say")
data <- data %>%

  mutate(across(Age, ~ as.character(.))) %>%
  separate(Age, into = c("minAge", "maxAge"), sep = "\\s*-*\\s*", convert = TRUE) %>%
  mutate(maxAge = sub("years old", "", maxAge))
print(data)


data$maxAge <- as.numeric(data$maxAge)

model1 <- lm(Familiarity_Score ~ minAge + maxAge + Gender + Country + FI_music_training +
Eng_with_music, data = data)

summary(model1)


regsubsets_output <- regsubsets(
  Familiarity_Score ~ minAge + maxAge + Country + FI_music_training + Eng_with_music,
  data = data,
  nvmax = 10)
summary(regsubsets_output)
summary(regsubsets_output)$bic
...

``{r}

library(ggplot2)

data$maxAge <- as.numeric(data$maxAge)

model2 <- lm(Familiarity_Score ~ maxAge + Eng_with_music, data = data)
```



```
summary(model2)
```

```
# Residuals vs Fitted Values Plot
```

```
plot(model2, which = 1)
```

```
# Normal Q-Q Plot
```

```
plot(model2, which = 2)
```

```
# Residuals vs Leverage Plot
```

```
plot(model2, which = 5)
```

```
# Create a scatter plot of the data points and add the regression line
```

```
ggplot(data, aes(x = maxAge, y = Familiarity_Score)) +
```

```
  geom_point() +
```

```
  geom_smooth(method = "lm", se = TRUE, linetype = "solid", color = "red") +
```

```
  labs(x = "Max Age", y = "Familiarity Score", title = "Max. Age as a Predictor of Music Familiarity Score") +
```

```
  theme_classic() +
```

```
  theme(
```

```
    plot.title = element_text(face = "bold", hjust = 0.5, size = 11),
```

```
    panel.grid.minor = element_blank(),
```

```
    axis.text = element_text(size = 10))
```

```
ggplot(data, aes(x = Eng_with_music, y = Familiarity_Score)) +
```

```
  geom_point() +
```

```
  geom_smooth(method = "lm", se = TRUE, linetype = "solid", color = "red") +
```

```
  labs(x = "Engagement with Music", y = "Familiarity Score", title = "Engagement with music as Predictor of Music Familiarity Score") +
```

```
  theme_classic() +
```

```
  theme(
```

```
    plot.title = element_text(face = "bold", hjust = 0.5, size = 11),
```

```
    panel.grid.minor = element_blank(),
```

```
    axis.text = element_text(size = 10))````
```