





# What Makes a Song Popular on Spotify?

# The Data



- 1 From Kaggle, sourced from Spotify Web API
- 2 169,000 songs from 1921 - 2020
- 3 Each row represents 1 song
- 4 No issues regarding Data Ethics

# The Variables - General Info

---

- ❖ Artists
- ❖ Track Name
- ❖ Track id
- ❖ Release Year
- ❖ Duration
- ❖ Genre
- ❖ Decade
- ❖ Key
- ❖ Mode
- ❖ Time Signature
- ❖ Tempo
- ❖ Explicit
- ❖ Followers
- ❖ Number of Artists

# The Variables - Audio Features

All numeric, on a scale of 0 - 100

- ❖ Acousticness
- ❖ Danceability
- ❖ Energy
- ❖ Instrumentalness
- ❖ Liveness
- ❖ Loudness
- ❖ Speechiness
- ❖ Valence

# The Variables - Audio Features

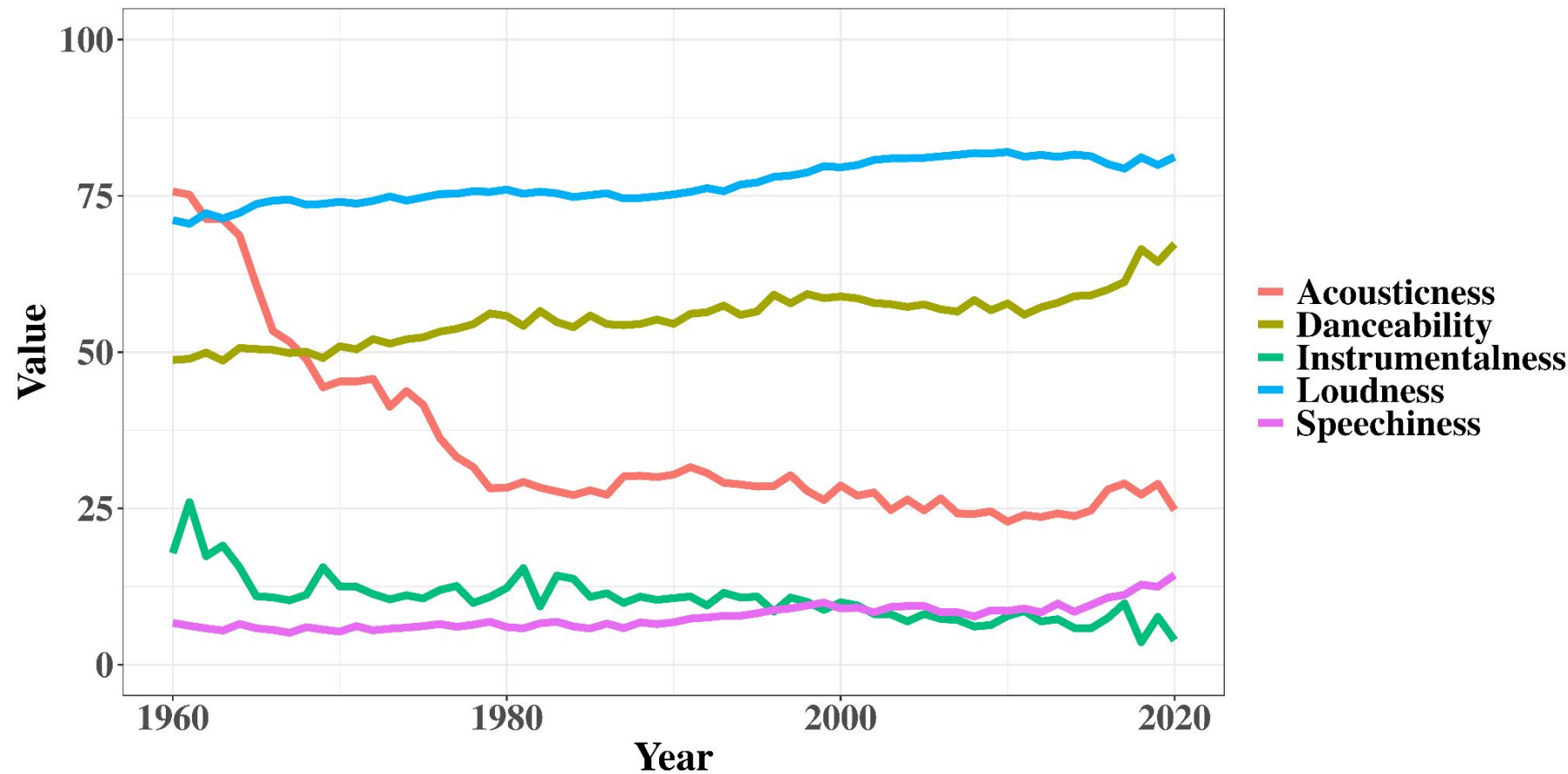


Numeric, on a scale of 0 - 100

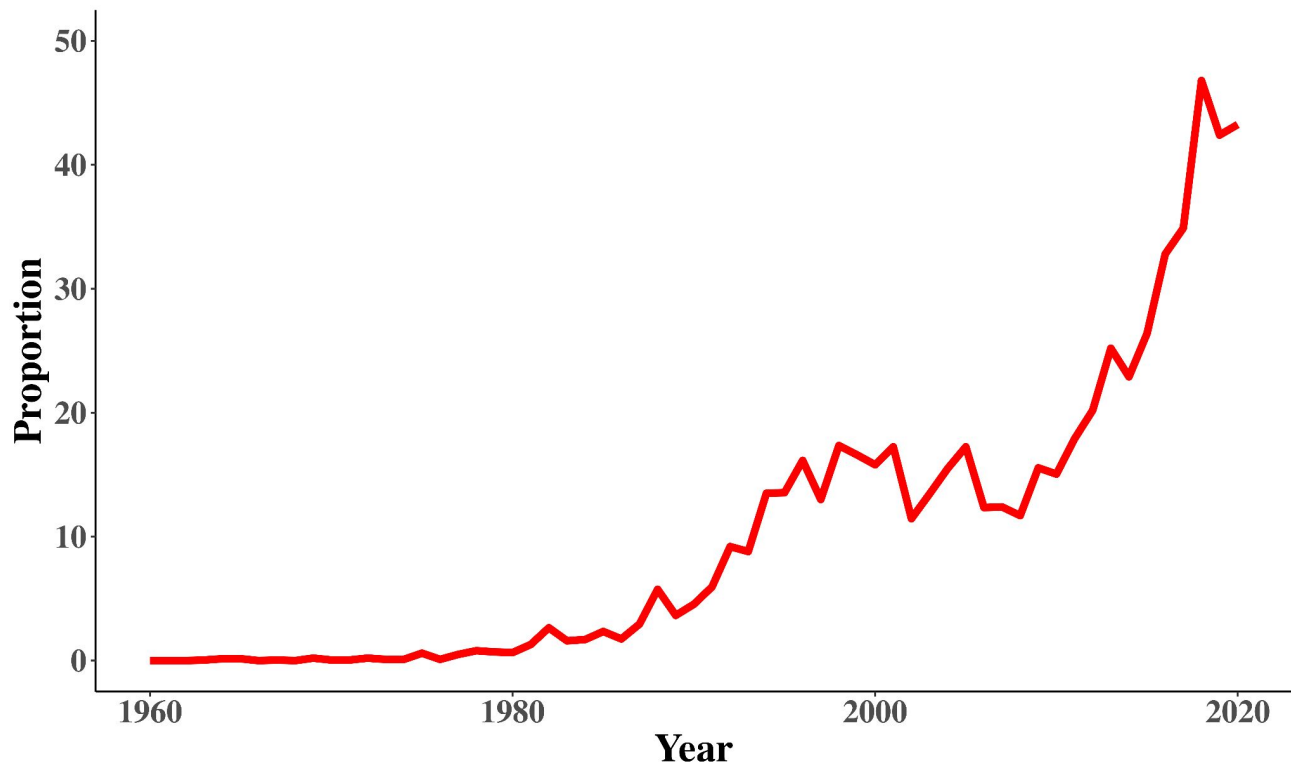
And most importantly...

- ❖ **Popularity** - *calculated by algorithm and is based, in the most part, on the total number of plays the track has had and how recent those plays are.*

# Avg Audio Features by Year

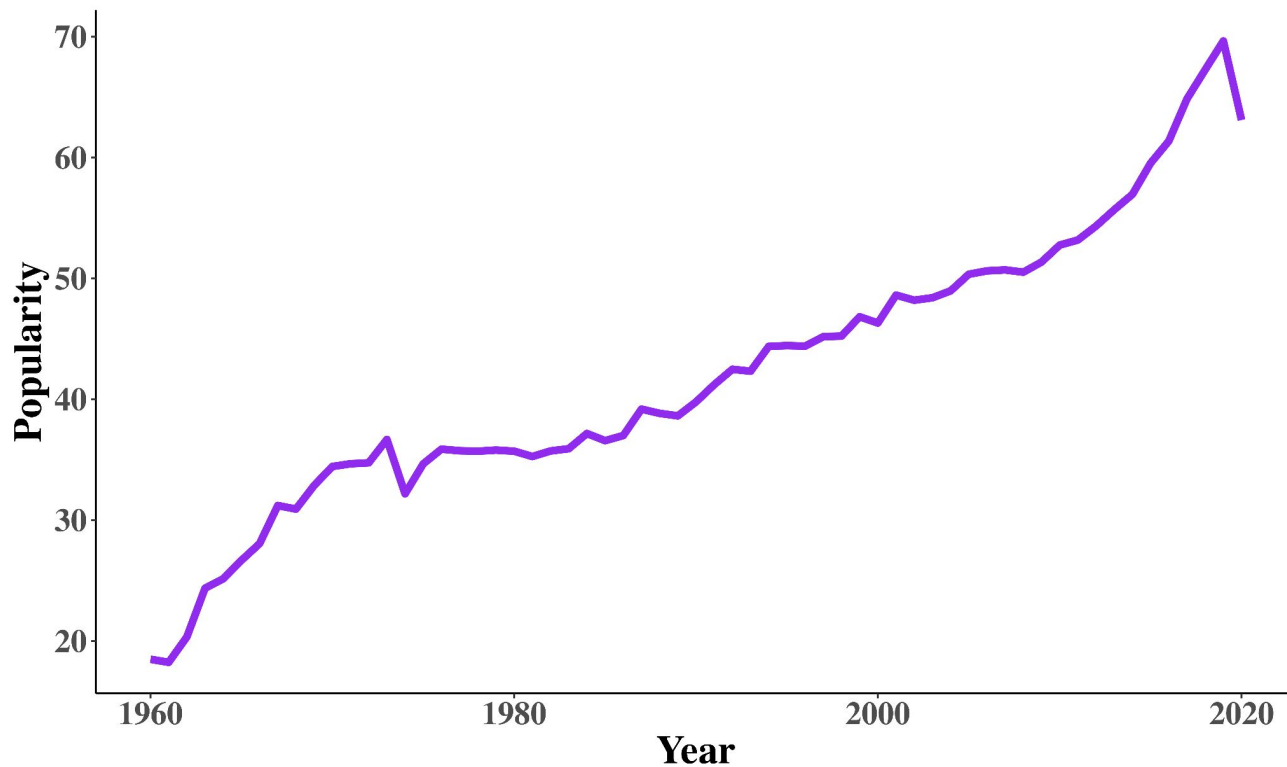


# Proportion of Explicit Songs per Year





# Avg Popularity by Year



# Linear Regression



$$Popularity = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$$



Determine the strength of the relationship between a response variable and multiple explanatory variables

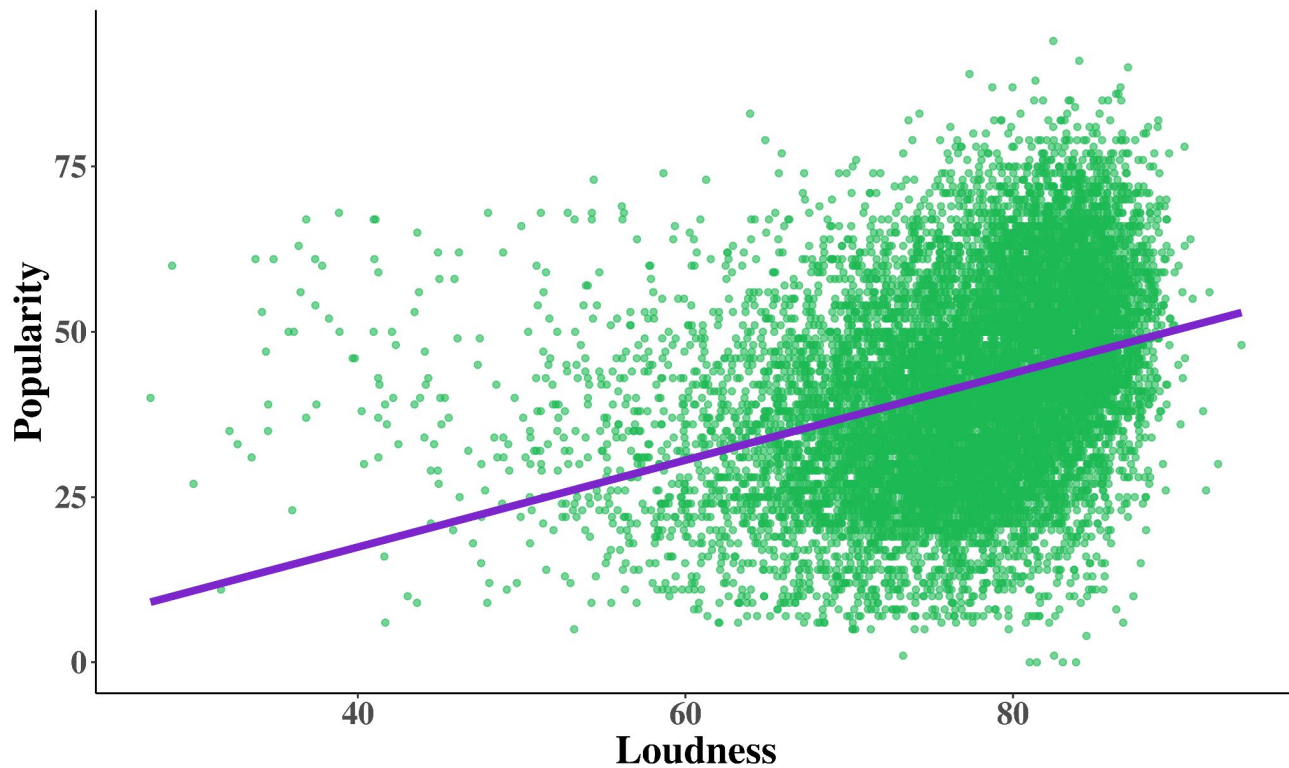


Explanatory vs Predictive Model?



Look for correlations, find 'most significant' variable, add to model - Repeat!

# Popularity vs Loudness



# Linear Regression - Final Model



***Popularity ~ Year + Danceability + Loudness + (-Liveness) + Explicit***



55% variance of Popularity explained



All explanatory variables have a P-value < 0.001



Problems with this model?

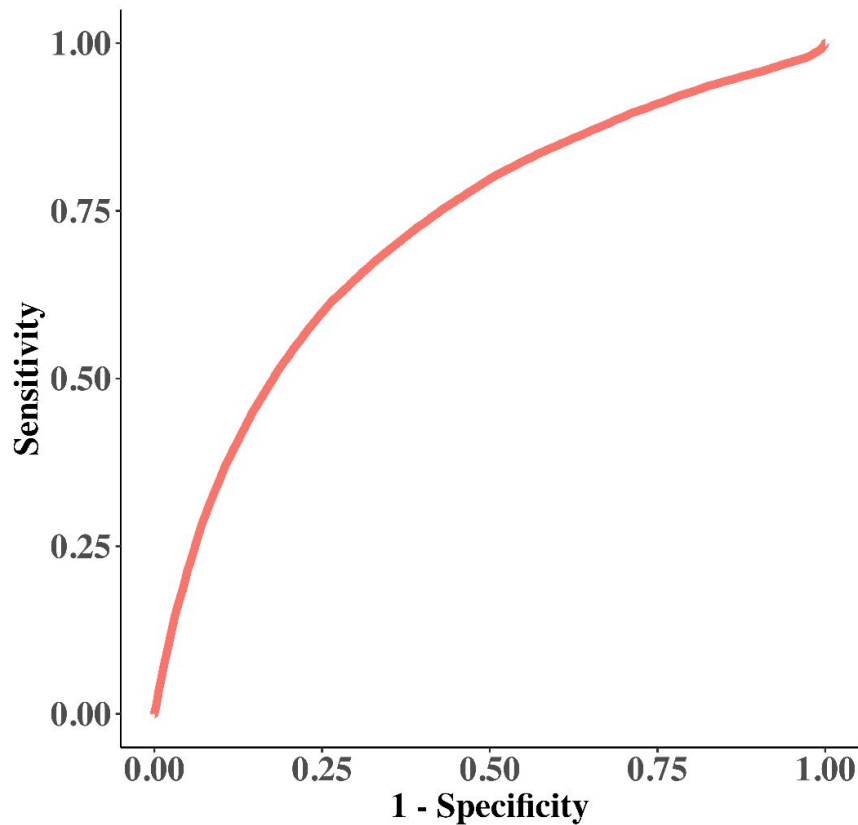
# 50 is the magic number!



## Logistic Regression

- Similar to Linear Regression, used to predict, or explain a binary outcome.
- The outcome is now **is\_popular**
- Looking for a high AUC score, variables must still be significant
- Remove any dates from the model!

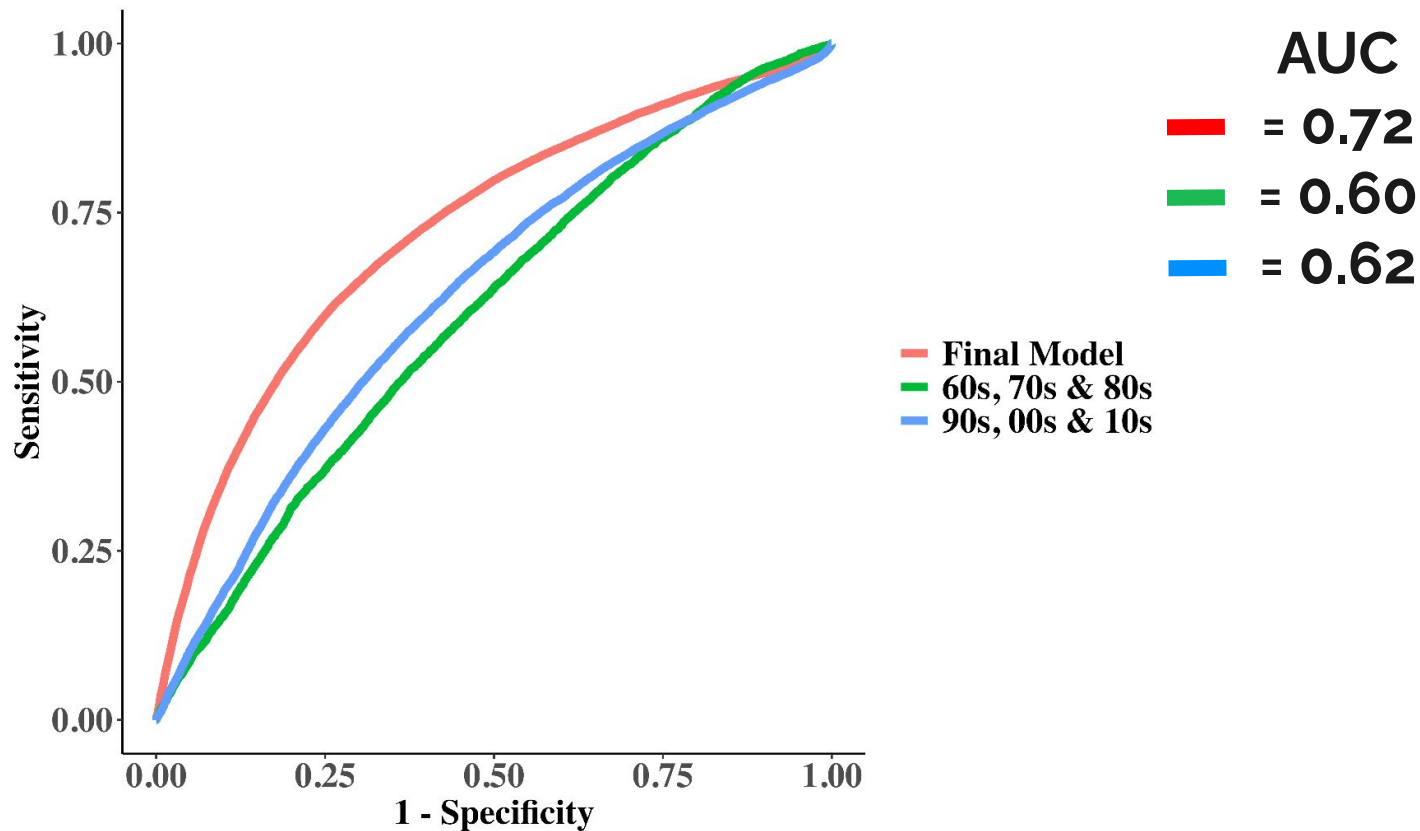
***is\_popular ~ loudness + explicit + danceability + number of artists***



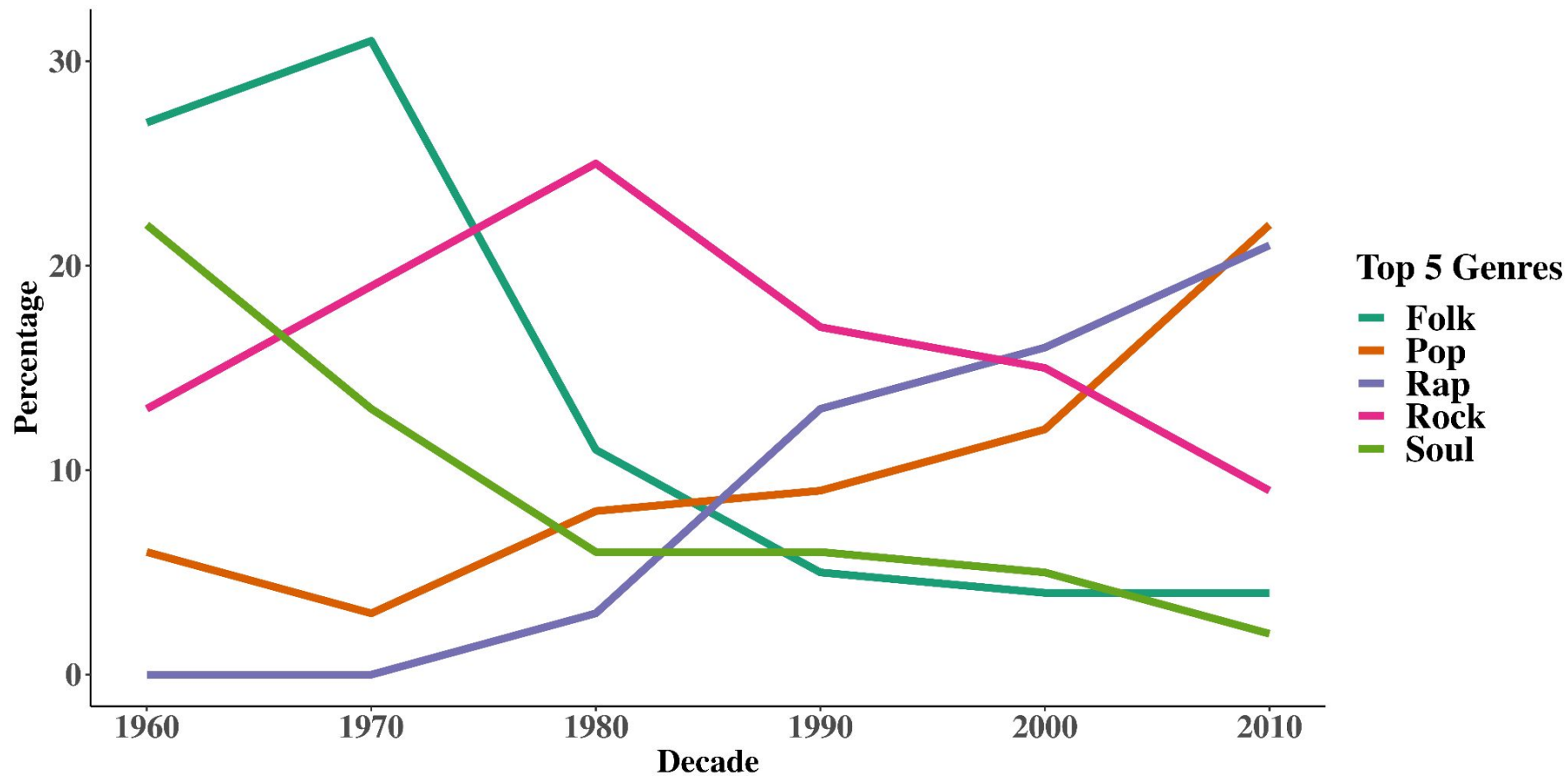
**AUC = 0.72**

— Final Model

*is\_popular* ~ *loudness* + *explicit* + *danceability* + *number of artists*



# Proportion of Top 5 Genres Over Time





# In Conclusion



## What Makes a Song Popular on Spotify?

*Is\_popular (50+) ~ loudness + explicit + danceability + no of artists*



AUC = 0.72



Top Recent Genres - Rap & Pop