# **To Do for Weds 4/18 Meeting**

# **Derek:**

# RMD

# D3 viz (see artists)

# Look into text entry

# **Everyone:**

# Push Code

* Familiarize yourself with the basics of Rshiny:
  + <https://shiny.rstudio.com/gallery/>
  + <https://shiny.rstudio.com/tutorial/>

# **Manksh:**

# Will add descriptions to notebooks

* Potentially help Kathy LOL

# **Kathy:**

# Rshiny Evolution over Time section

# **Louis**

# Top words over time

# Interactive Word Cloud app (see https://shiny.rstudio.com/gallery/word-cloud.html)

# 

# **ARTISTS:**

# **#Unique artists in top 100: diversity of artists**

# **Line Plot by Derek**

# **Artists with most top 100 singles**

# **Bar graph**

# **Artists' relevance / life span (RSHINY)**

# **Horizontal Bar ish Plot of Top 20 longest life span artists on y axis, years on x, nonaligned rectangles in this plot (D3 -> export png/jpeg to RMD)**

# **Tick Marks**

# **Most collaborative artists (RSHINY)**

# **Line plot of artist collaboration**

# **Bar graph of most featured**

# **Bar graph of most collaborated**

# **Most verbose artists (RSHINY)**

# **Bar graph**

# **Dendrogram -**

# **Still missing some explanation of what is a dendro**

# **Scatterplots of Strongest Correlations:**

# **Hexbins if we have time**

# **Energy vs Loudness -- Perhaps we can have an interactive slider to transform the data**

# **Most explicitness**

# 

# **SONGS:**

# **Top 100 ngrams (RSHINY - Slider)**

# **By decade**

# **Most frequent words of all time (word count without stop words)**

# **# of Covers ( RSHINY slider for bin width on bargraph)**

# **Enter random word, plot line graph of usage over**

# 

# 

# **EVOLUTION OVER TIME (RSHINY)**

# **Scattteryplot of correlations**

# **Spotify Features Over time (Avg value, boxplots, parallel coordinate plots, correlation matrix / dendrogram)**

# **Duration**

# **Acousticness**

# **Danceability**

# **Energy**

# **Explicitness**

# **Instrumentalness**

# **Liveness**

# **Loudness**

# **Popularity**

# **Speechiness**

# **Tempo**

# **Valence**

# **Verbosity**

# 

# 

# **Kathy**

# Interactive Plot Plotly Parallel Coordinate

# Basis of Rshiny Dashboard

# **Manksh**

# Artists lifespan /longevity

# Most explicit artists

# Music stuff working together with Derek

# **To Do for Weds 4/11 Meeting**

# **Derek:**

# From correlation matrix, make scatterplots of pairs of highly correlated features

# Translate from Python to R:

# Diversity / Unique Artists in Top 100

# Artists with Most top 100 songs

# Check most featured artists

SONGS:

* Scatterplots / Correlation Matrix

# **Louis**

# Louis will redo word count analysis with R word cloud, + stop words

# Translate Derek's EDA-words.ipynb:

# Word subsets over time

# # Covers

# **Kathy**

# Interactive Plot Plotly Parallel Coordinate

# Basis of Rshiny Dashboard

# **Manksh**

# Artists lifespan /longevity

# Most explicit artists

# Music stuff working together with Derek

# 

# **To Do for Tues 4/3 Meeting**

# **Louis**

* redo word count analysis with R word cloud, + stop words :(
* Add explanation to your python notebook such as:
* There's no explicit way of detecting covers so we string matched the title and checked if they have different artists
* We have duplicates since a song may be popular at the end of one year and continue to be popular at the beginning of another

**Manksh**:

* Add explanation to your python notebook
* Explore when different genres became popular in regards to year
* Start translating your notebook into R

**Kathy**

👍 Add boxplots over time (maybe moving avgs if there’s time?)

👍 Check if the 337 NAs correspond to older songs

👍 Parallel coordinate plot for average variable values over year

# **Derek**

* Add explanation to your notebook
  + Ie: Not much change in artist diversity since we'd like to see more years before we claim there's a sinusoidal pattern
* Check if madonna’s 36/50 times in the top 100 were because of different singles or repeated same single? √
* Confirm if rap counts as speechiness √ https://beta.developer.spotify.com/documentation/web-api/reference/tracks/get-several-audio-features/
* Who is the most featured artist? √
* Tidy the dataframe √

# 

# 

# 

# **Introduction**

* Why did you choose this topic and questions we're interested in studying
* List team members and a description of what everyone did

**Description of Data**

* Describe how data was collected/accessed
  + Who is included in this study
  + When was this query processed
  + Definition of any features
  + How is data collected?
  + Estimated accuracy rate?
* Merge the data / Match
  + Match song name from Billboard - Spotify: lowercase everything -- Derek (over spring break)

Note: Is\_explicit, popularity columns were feature engineered

* 94 % of songs were successfully queried. The rest were NA's
* Billboard top 100 for 50 years (1964-2016) songs were matched
* No genre included in the pull

## **Analysis Questions**

Artists

1. Artists Career Span Over Time
2. Artists' Word Choice over time -- which artists have songs that reflect the development of their life?
3. **# unique artists that have #1 Hits in Billboard Hot 100 -- Diversity of artists - Derek**
   1. We have a lot more diversity in artists / genre but we hypothesize that this diversity decreases in the Top 100.
4. Collaboration between artists - Derek

Songs

1. Top 100 ngrams by decade Loulou
2. Most frequent words of all time Loulou
3. **Song length over time** Kathy
4. **Danceability /Energy / Tempo over time:** Kathy
   1. **What are the most danceable words: Are there words that are indicative of danceability ?**
   2. Parallel coordinate plot of all metrics over time (plotly) - Derek (no plotly, just parcoords)
      1. We expect an inverse relationship between speechability / acousticness
      2. Instrumental vs speechiness
   3. Plotly: Can we create a slider for Year and observe how the parcoord plot changes over time?
5. 5. Is there one genre that dominates the top 100 nowadays? How does genre vary over time? (can we tie that with energy of a song?)
6. Verbosity over time: #words song / length of song = words / sec -- who are the most verbose / least? - Derek (will compare to ‘speechiness’)
   1. Repetitive lyrics in Top 50 ? We'd expect the vocabulary rate to go down debateable if good or bad - Derek
7. # Covers in top 50 over time -- how likely are we to see repeats / covers? We'd express that to increase Loulou
8. When does 4 count 3 count music peak ? Anything that's not 4 could be a lower tempo?
   1. Decision Tree Features Visualization - Manksh (Trial Decision tree)
9. **How does explicit do lyrics get over time** -> CULTURAL DIFFERENCE - Manksh
10. **Race of #1 Hit artist / Year :** Racial inclusiveness or no? we'd hope so - Manksh

**For week of 3/26**, try to investigate the questions and summarize your findings. Then we'll decide on which kind of visualizations would be best suited for each finding.

**Analysis of Data Quality**

* Graphs, textual description
  + Univariate graphs/distributions
  + Outliers
  + Clusters

**EDA**

* Findings
* Data Cleaning Process
* Approaches that worked / didn't
* Data challenges

**Conclusion**

* Limitations,
* Next Steps