

# Final Project Progress

By: Brandon Connell & Noah Vanscoyoc

# What do we have so far?

- So far, we have an entire introductory section on our github as well as a lot of work into a bunch of very rudimentary graphs that need to be fleshed out and some beginning data points that we can use to find some interesting things about the data.
- One of our first initial interests was an overall “grade” or score for each individual song based on important key factors to figure out which song is the best song (statistically)

## What is the best Summer Billboard Hit?

This question can seem to be a loaded question, and truthfully very opinionated. But we can take some key factors into consideration to decide factually what is the best song? \*Taking the key factors into consideration, giving a weight to each factor then gives us the chance to multiply it by its respective data point and figure out its overall score once added \*Once the score has been calculated, all that must be done is put everything into the original table and select the data points we only want to see \*Finally, once the scores are calculated and given, we then will give each score a “Grade” based on this scale: #1 song = The Overall Best Song 70+ = “S Tier” 60-69 = “A Tier” 50-59 = “B Tier” 40-49 = “C Tier” 39 and below = “D Tier”

```
songs_data <- all_billboard_summer_hits

weights <- c(danceability = .3,
            liveness = .3,
            energy = .3,
            valence = .02,
            acousticness = .02,
            speechiness = .04,
            instrumentalness = .02)

all_billboard_summer_hits %>%
  mutate(score = rowSums(songs_data[, names(weights)] * weights)) %>%
  select(track_name, artist_name, score, year, danceability, liveness, energy, valence, acousticness, speechiness, instrumentalness)
  arrange(desc(score))
```

This is our function!

```
grade <- function(score){
  if(score>=78){
    return("The Overall Best Song")
  }
  else if (score >= 70 & score <= 77){
    return("S Tier")
  }
  else if(score >= 60 & score <= 69){
    return("A Tier")
  }
}
```

And after a quick look, we can see that most songs have used C major in the billboard summer hits. Maybe this means C major is the best key mode for making the top billboard summer hits?

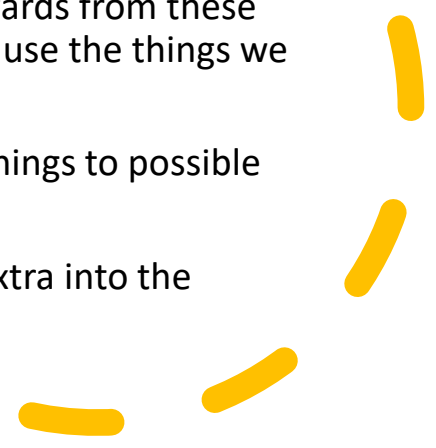
The table below is just to show some songs that are in the C major key mode:

```
all_billboard_summer_hits %>%  
  filter(key_mode == "C major") %>%  
  select(key_mode, track_name, everything())
```



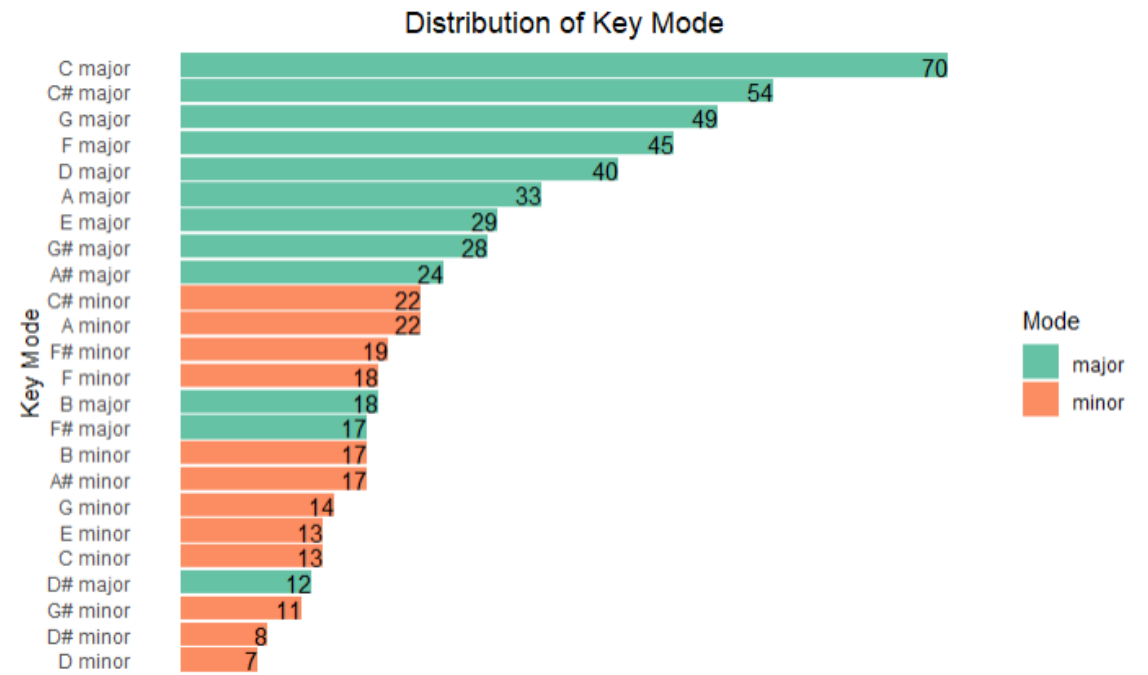
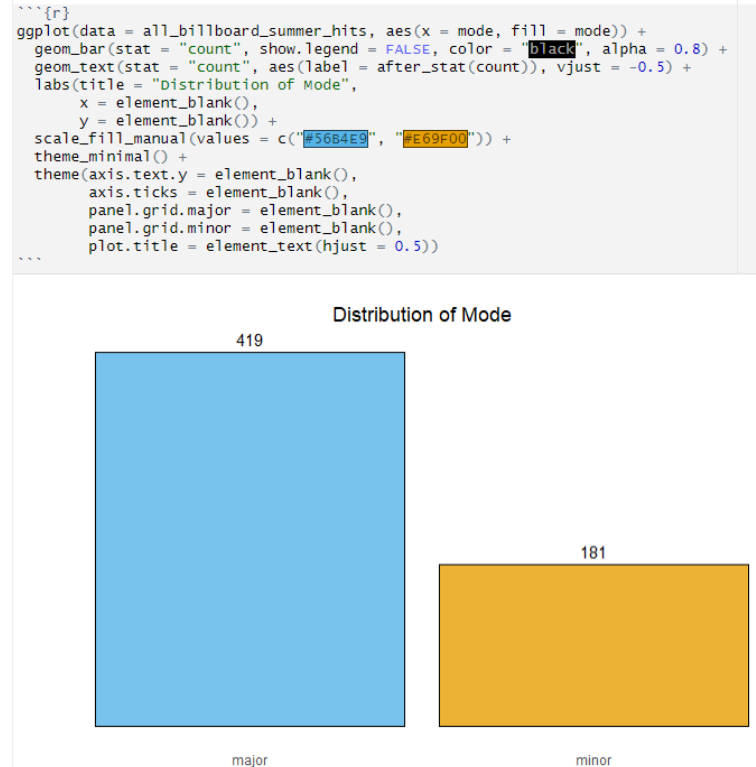
## What else?

- Well, currently we are pretty much complete for graphs and visualizations for the topics we were interested. Next plan is to work backwards from these data visualizations and find interesting tables and ways to use the things we learned from the lessons in the project.
- This means we will be going through and finding certain things to possible mutate, or summarize and filter out what we want to see.
- For instance, what is above, just something small to see extra into the dataset to see the songs that were only made in C major.



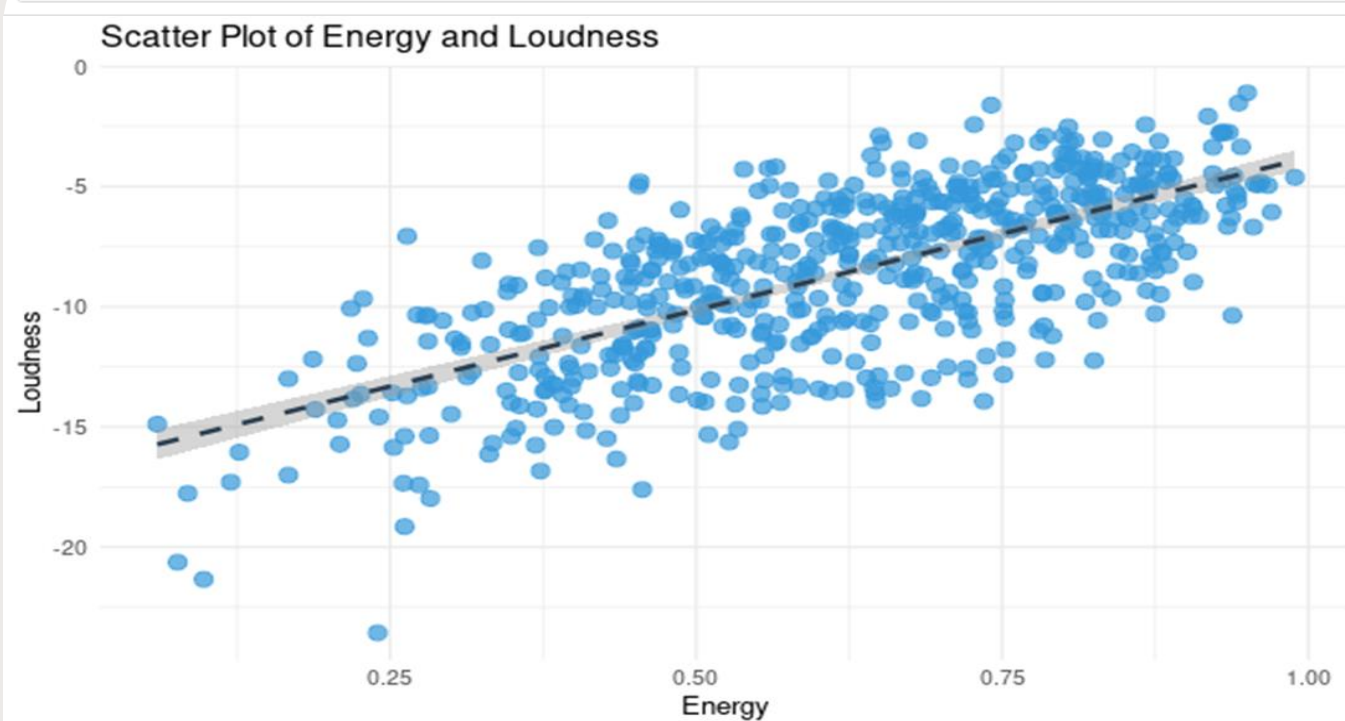
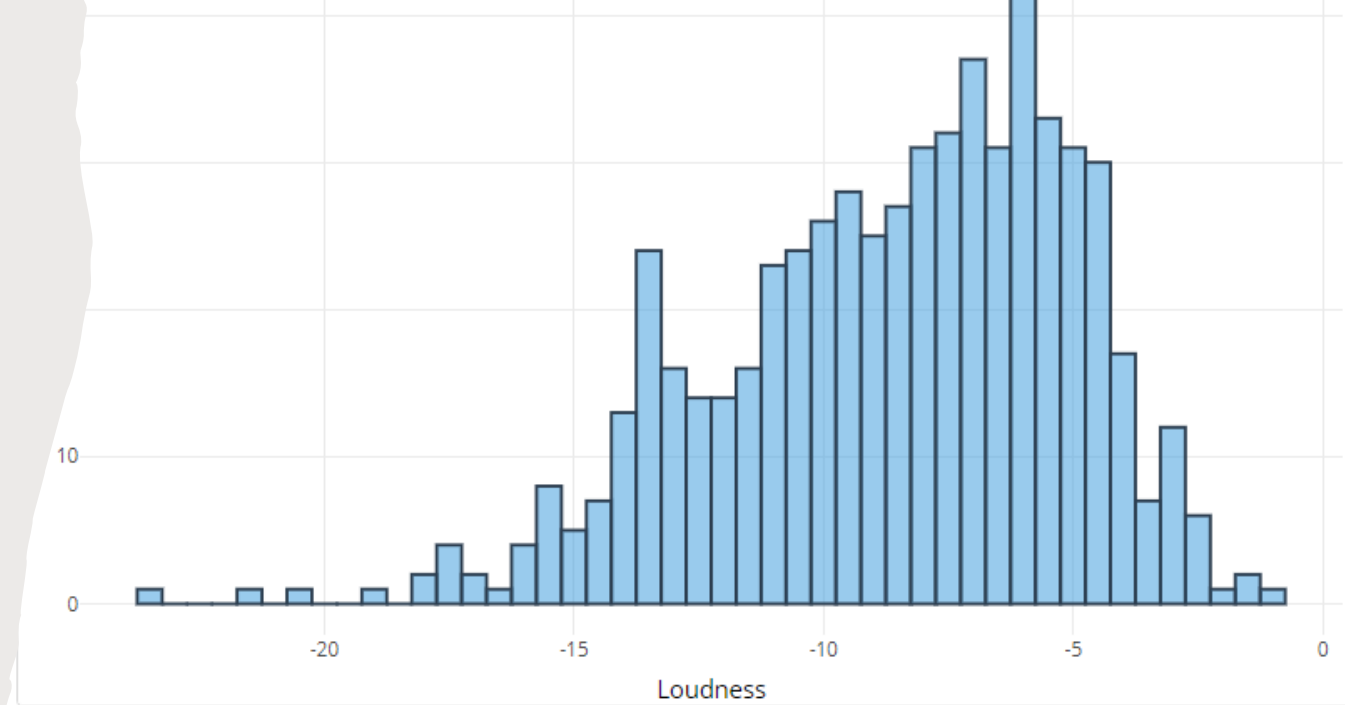
# Some Graphs

- The first data type we looked at was if a song was in major or minor
- Secondly was the number of different key modes in every top hit



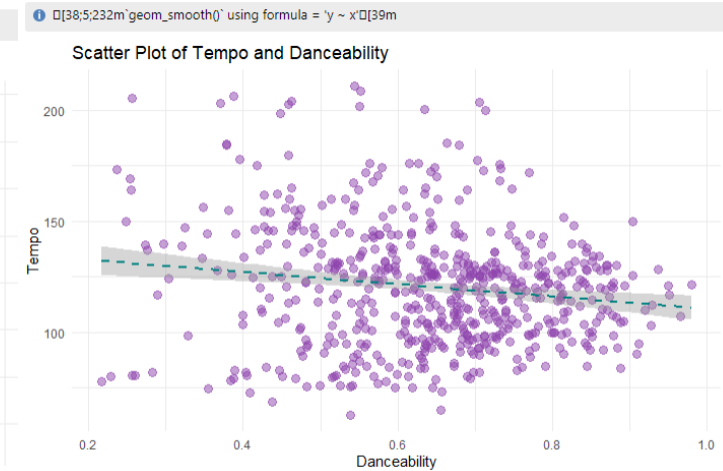
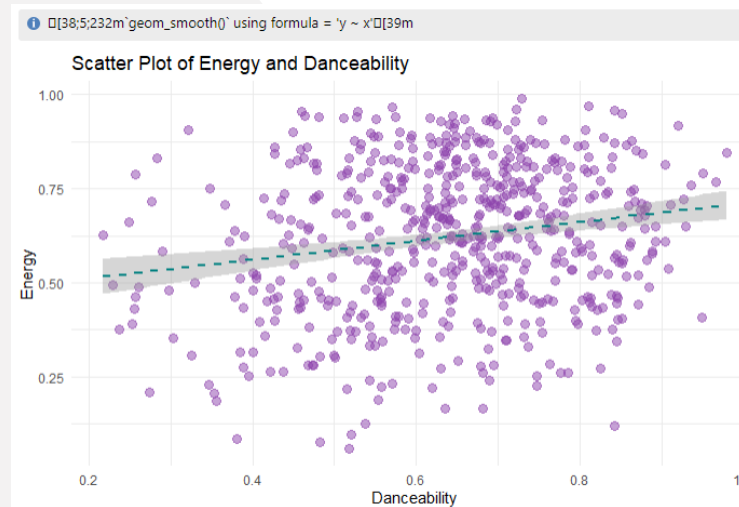
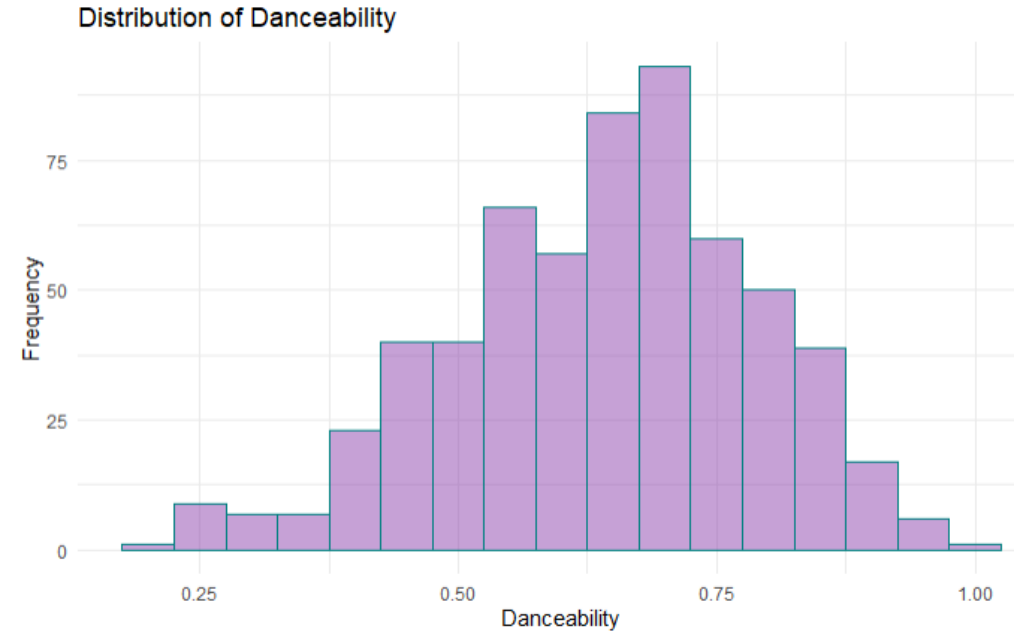
# Loudness Graphs

- Our third graph was to demonstrate loudness in the songs and count them out with a interactive graph with plotly
- Alongside our initial loudness plot, we then decided to compare the songs to determine whether or not energy and loudness had any correlation over the years



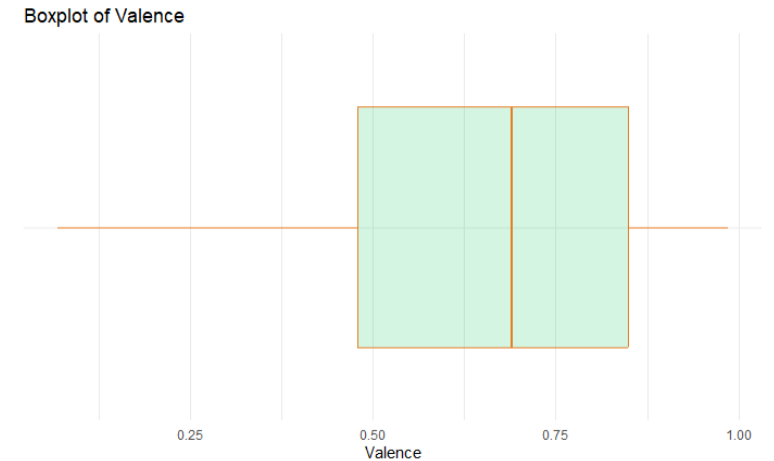
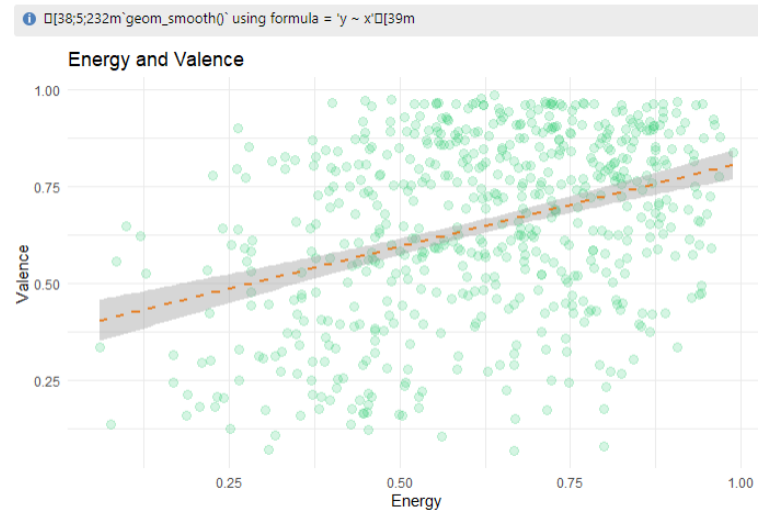
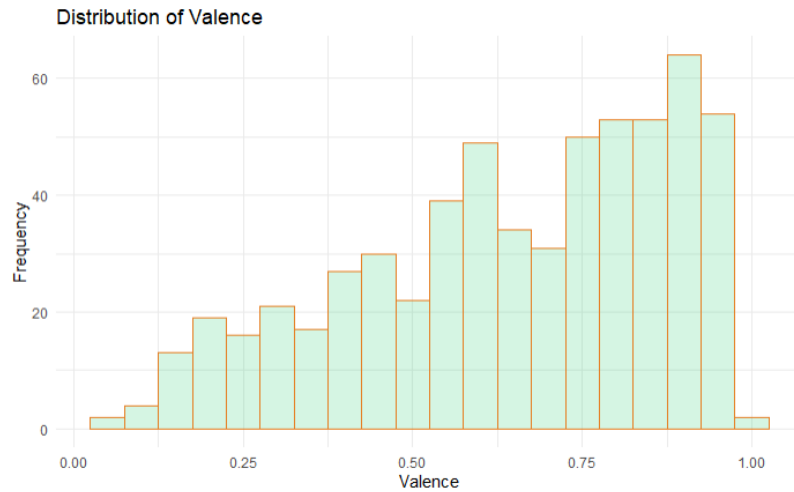
# Danceability Graphs

- For these couple graphs, we also wanted to find any correlation between different variables and danceability



# Valence Graphs

These graphs demonstrate valence and its numbers for the dataset. Alongside it is a comparison between Valence and Energy.



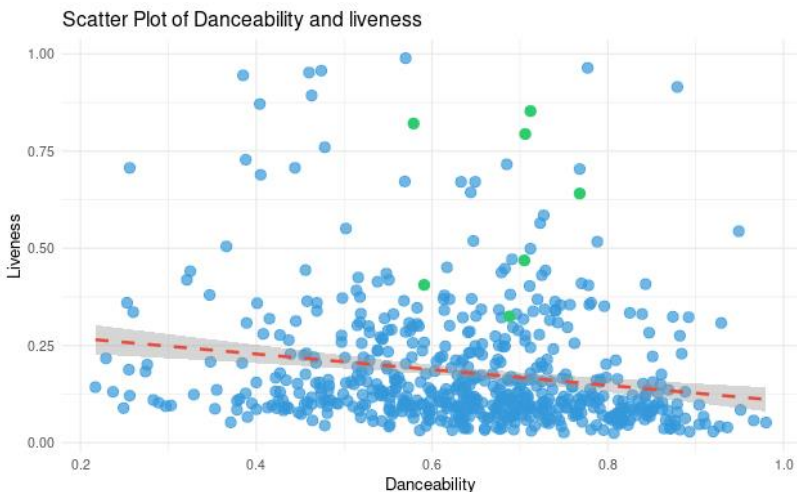
A tibble: 7 × 22

danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness
0.768	0.641	G	-8.320	major	0.2010	0.000228	0.000571	0.0944
0.591	0.406	C	-8.466	minor	0.0331	0.839000	0.000000	0.2580
0.579	0.821	C#	-4.616	major	0.1520	0.011700	0.000000	0.0631
0.705	0.469	A	-7.460	major	0.0771	0.242000	0.000000	0.1040
0.706	0.794	B	-4.321	minor	0.0568	0.110000	0.000000	0.2760
0.712	0.853	G#	-6.337	major	0.0894	0.003060	0.014900	0.2450
0.688	0.325	F	-8.083	minor	0.2510	0.107000	0.000000	0.0837

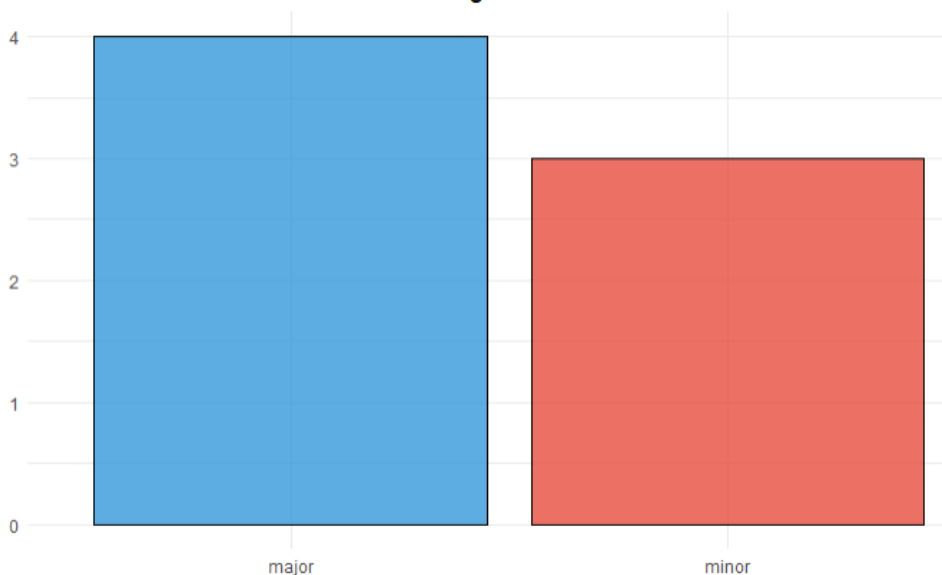
7 rows | 1-9 of 22 columns

# Rihanna

- Lastly, we found the artist with the most summer billboard hits, This being Rihanna we then dug a bit deeper into some of her difference key statistics.



Distribution of Modes in Rihanna's Songs



And what were her songs?

```

```{r}
rihanna_songs <- all_billboard_summer_hits %>%
  filter(artist_name == "Rihanna")

rihanna_songs %>%
  select(track_name, year)
```

```

A tibble: 7 × 2

| track_name          | year |
|---------------------|------|
| Pon de Replay       | 2005 |
| Unfaithful          | 2006 |
| Umbrella            | 2007 |
| Take A Bow          | 2008 |
| Disturbia           | 2008 |
| Where Have You Been | 2012 |
| Needed Me           | 2016 |