

Final Project: Swiftie

Abigail Ponsetto

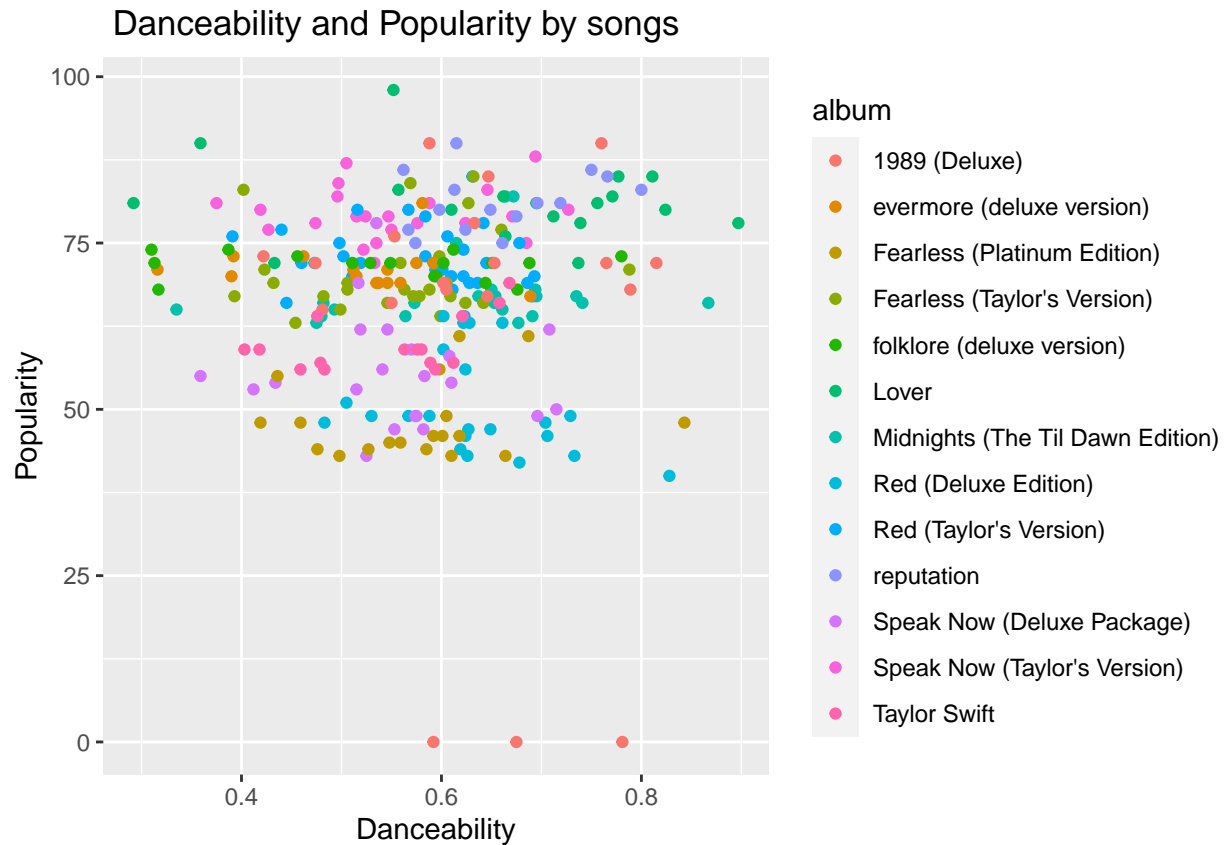
12/2023

My data 115 project focuses on Taylor Swift's music catalog. No matter who you are in 2023 you know who Taylor Swift is, Times 2023 Person of the Year and new member of the billionaire club. The Eras Tour is an international phenomenon. Taylor Swift has spent the last few years rerecording her first 6 albums after her masters were sold off. These re-records of the albums and for the most part the same songs have become more popular than the originals, as shown in my project. We will look at all of her released music including the stolen versions and Taylor's Version up to Speak Now (Taylor's Version). We will look at the dance ability and popularity of each album. We will look at whether dance ability affects popularity or not. My big question is do fans shows more support for Taylor's Versions songs because of the albums being better or despite them being basically the same as the originals (also known as Stolen Version to the Swifties)? Along with that are the re-records really that similar to the originals. I processed through the original data to slim it down to what I wanted to take a closer look at. I also created another file with the calculated averages of this data. Below I supplied my source for the original data set. With the rerecords there has sparked a lot of debate on if there is any difference between the originals and Taylor's Versions of her albums (aside for the from the vault tracks). I am curious to analyze any differences or similarites to the basis of the songs.

source:<https://www.kaggle.com/datasets/jarredpriester/taylor-swift-spotify-dataset>

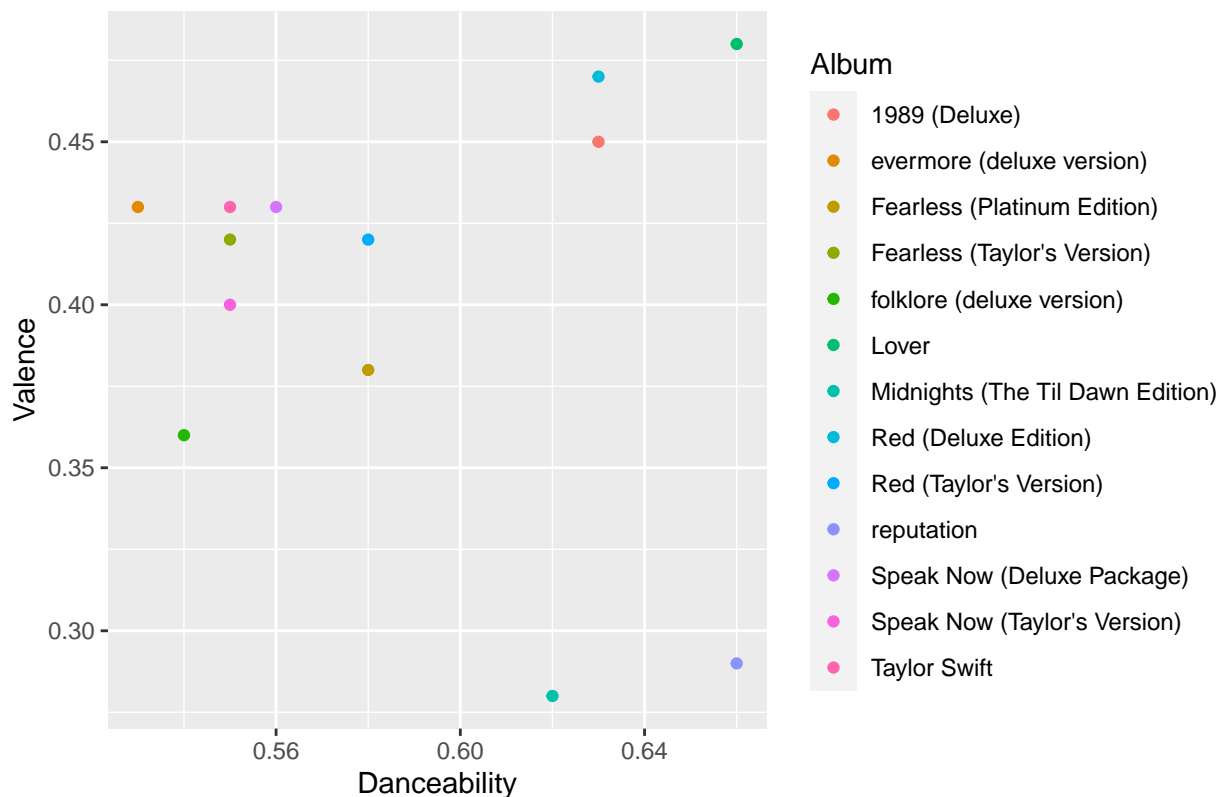
```
TS <- read.csv("TaylorSwiftFull.csv")
TS2 <- read.csv ("TaylorS.csv")
REDSV <- TS[TS$album == 'Red (Deluxe Edition)', ]
REDTV <- TS[TS$album == "Red (Taylor's Version)", ]
FEARLESSSV <- TS[TS$album == "Fearless (Platinum Edition)", ]
FEARLESSTV <- TS[TS$album == "Fearless (Taylor's Version)", ]
SPEAKNOWTV <- TS[TS$album == "Speak Now (Taylor's Version)", ]
SPEAKNOWSV <- TS[TS$album == "Speak Now (Deluxe Package)", ]
library(ggplot2)
library(gridExtra)

# Scatter plot of Danceability and Popularity by songs
ggplot(TS, aes(x = danceability, y = popularity, color = album, group = album)) +
  geom_point() +
  labs(title = "Danceability and Popularity by songs",
       x = "Danceability",
       y = "Popularity")
```



```
##valence vs danceablity
ggplot(TS2, aes(x = Avg.Dance, y = Avg.Valence, color = Album)) +
  geom_point() +
  labs(title = "Danceability and Valence by Album",
        x = "Danceability",
        y = "Valence")
```

Danceability and Valence by Album



```
# Bar plot of Popularity by Album
bar_plot_popularity <- ggplot(TS, aes(x = popularity, y = album)) +
  geom_bar(stat = "identity", fill = "blue", alpha = 0.7) +
  labs(title = "Popularity by Album",
       x = "Popularity",
       y = "Album")

# Bar plot of Danceability by Album
bar_plot_danceability <- ggplot(TS, aes(x = danceability, y = album)) +
  geom_bar(stat = "identity", fill = "blue", alpha = 0.7) +
  labs(title = "Danceability by Album",
       x = "Danceability",
       y = "Album")

REDSVBAR <- ggplot(REDSV, aes(x = popularity, y = name)) +
  geom_bar(stat = "identity", fill = "red", alpha = 0.7) +
  labs(title = "Red SV Popularity",
       x = "Popularity",
       y = "Album")

REDTVBAR <- ggplot(REDTV, aes(x = popularity, y = name)) +
  geom_bar(stat = "identity", fill = "red", alpha = 0.7) +
  labs(title = "Red TV Popularity",
       x = "Popularity",
       y = "Album")

FEARLESSSVBAR <- ggplot(FEARLESSSV, aes(x = popularity, y = name)) +
```

```

geom_bar(stat = "identity", fill = "orange", alpha = 0.7) +
labs(title = "FEARLESS SV Popularity",
     x = "Popularity",
     y = "Album")

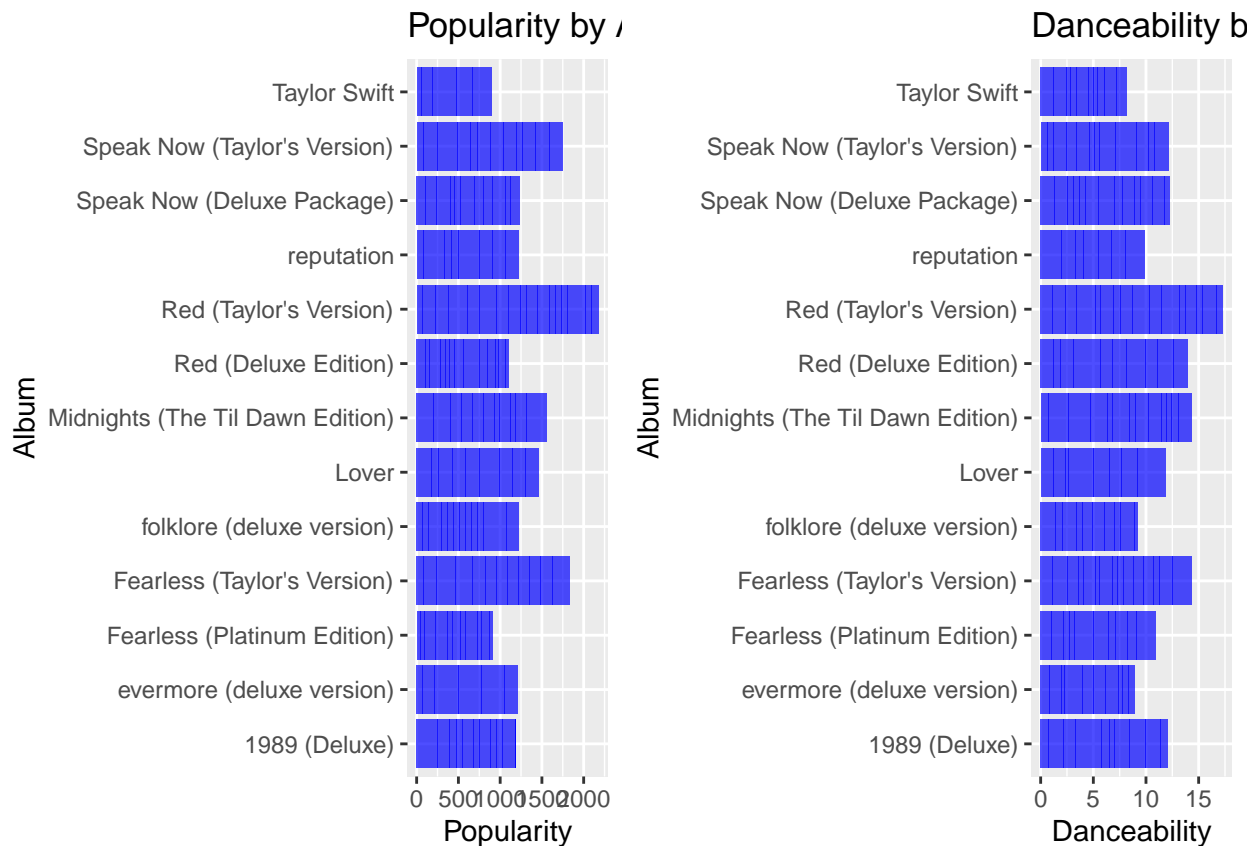
FEARLESSTVBAR <- ggplot(FEARLESSTV, aes(x = popularity, y = name)) +
  geom_bar(stat = "identity", fill = "yellow", alpha = 0.7) +
  labs(title = "FEARLESS TV Popularity",
       x = "Popularity",
       y = "Album")

SPEAKNOWSVBAR <- ggplot(SPEAKNOWSV, aes(x = popularity, y = name)) +
  geom_bar(stat = "identity", fill = "pink", alpha = 0.7) +
  labs(title = "SPEAK NOW SV Popularity",
       x = "Popularity",
       y = "Album")

SPEAKNOWTVBAR <- ggplot(SPEAKNOWTV, aes(x = popularity, y = name)) +
  geom_bar(stat = "identity", fill = "purple", alpha = 0.7) +
  labs(title = "SPEAK NOW TV Popularity",
       x = "Popularity",
       y = "Album")

# Arrange plots in a 1x2 grid
grid.arrange(bar_plot_popularity, bar_plot_danceability, ncol = 2)

```



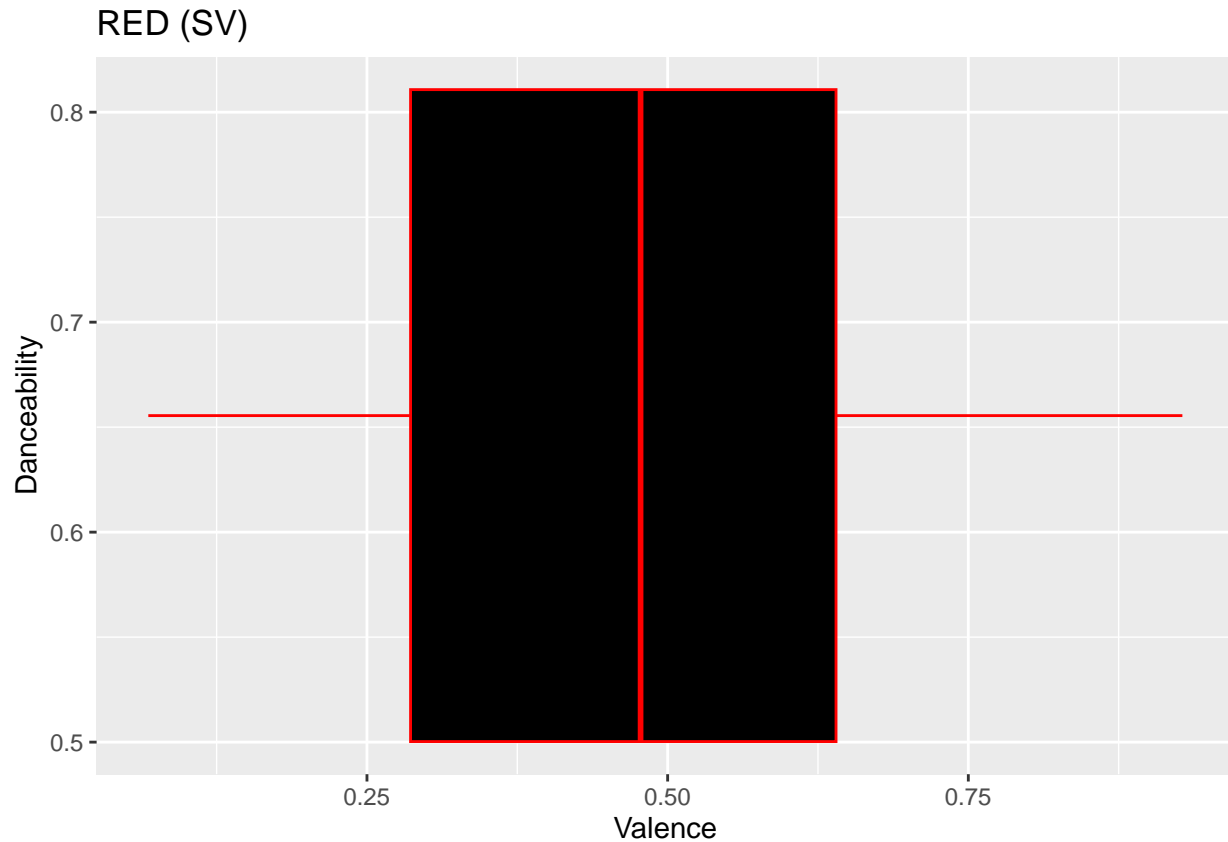
```

##boxplot RED
ggplot(REDSV, aes(x = danceability, y = valence )) +
  geom_boxplot(fill = "black", color = "red") +

```

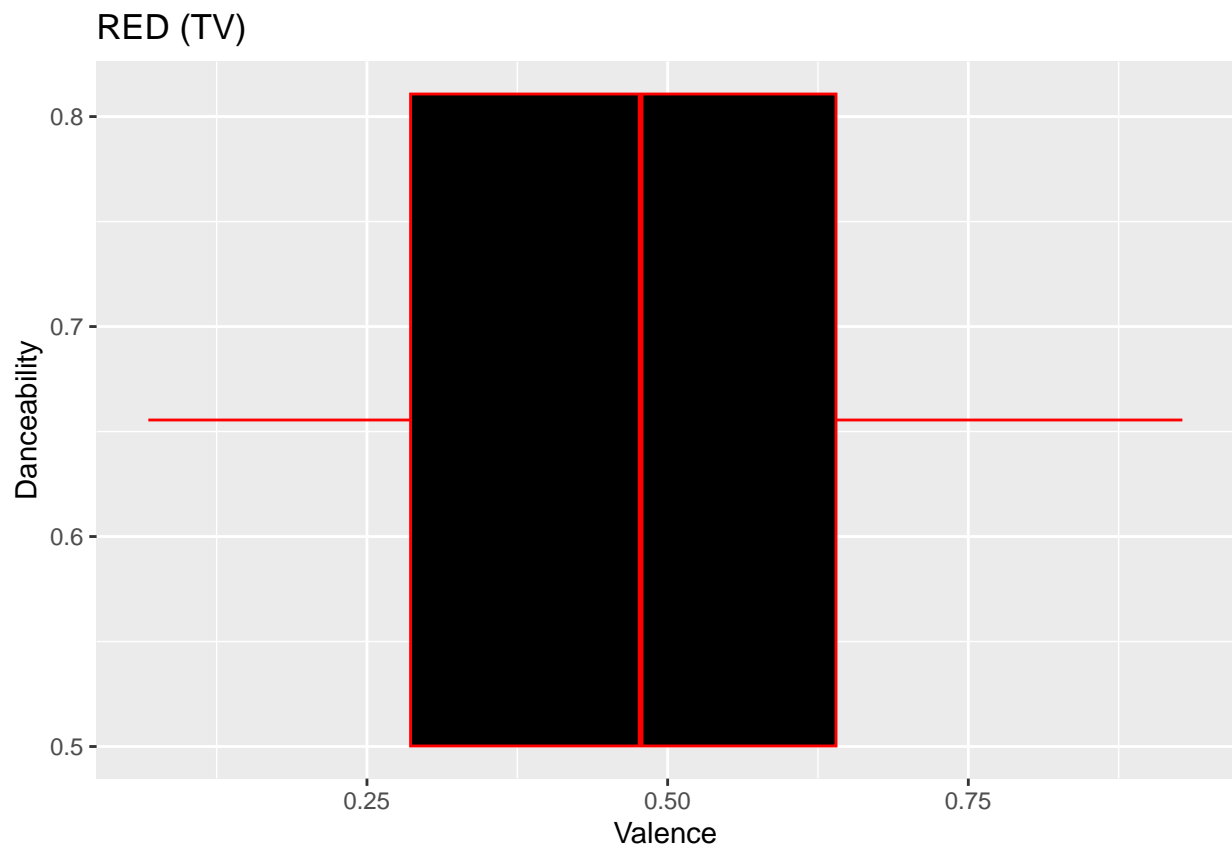
```
labs(title = "RED (SV)",
      x = "Danceability",
      y = "Valence") +
coord_flip()
```

```
## Warning: Continuous x aesthetic
## i did you forget `aes(group = ...)`?
```

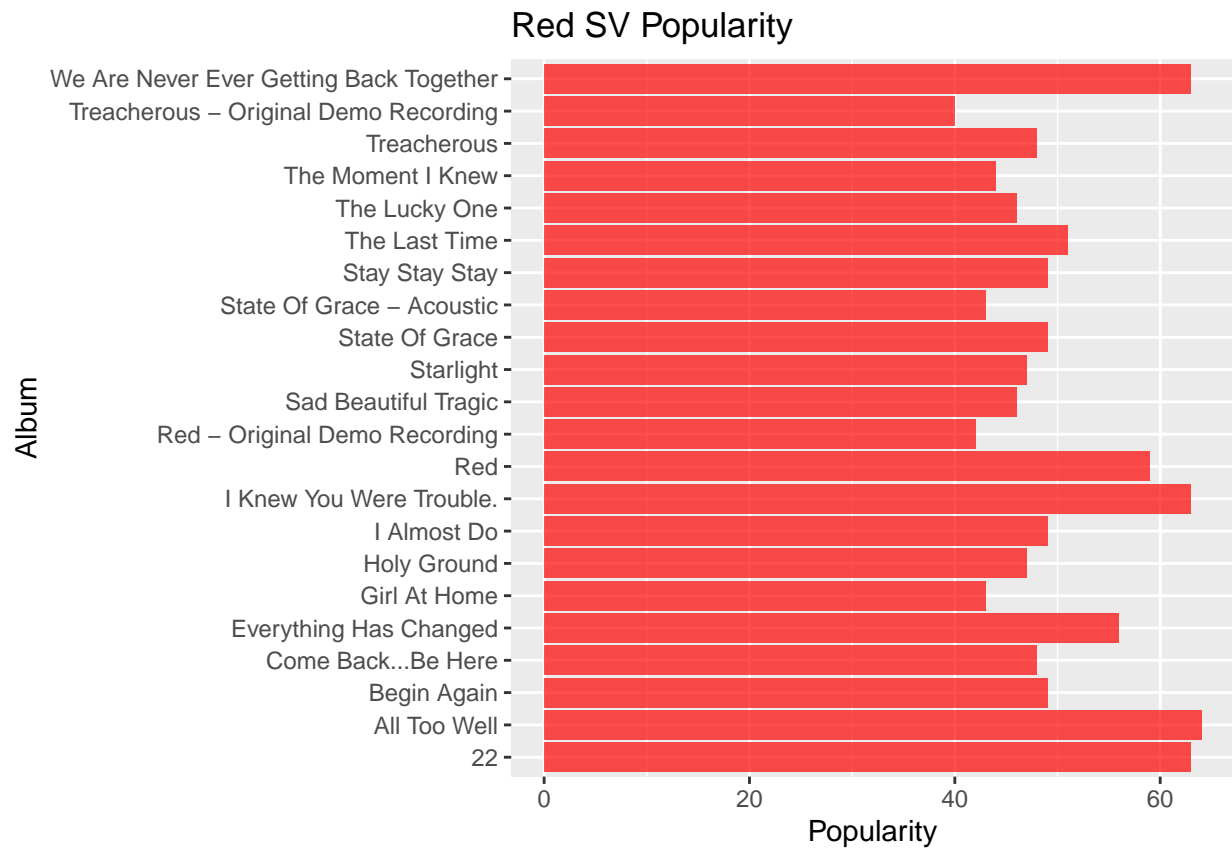


```
##boxplot RED
ggplot(REDSV, aes(x = danceability, y =valence )) +
  geom_boxplot(fill = "black", color = "red") +
  labs(title = "RED (TV)",
        x = "Danceability",
        y = "Valence") +
  coord_flip()
```

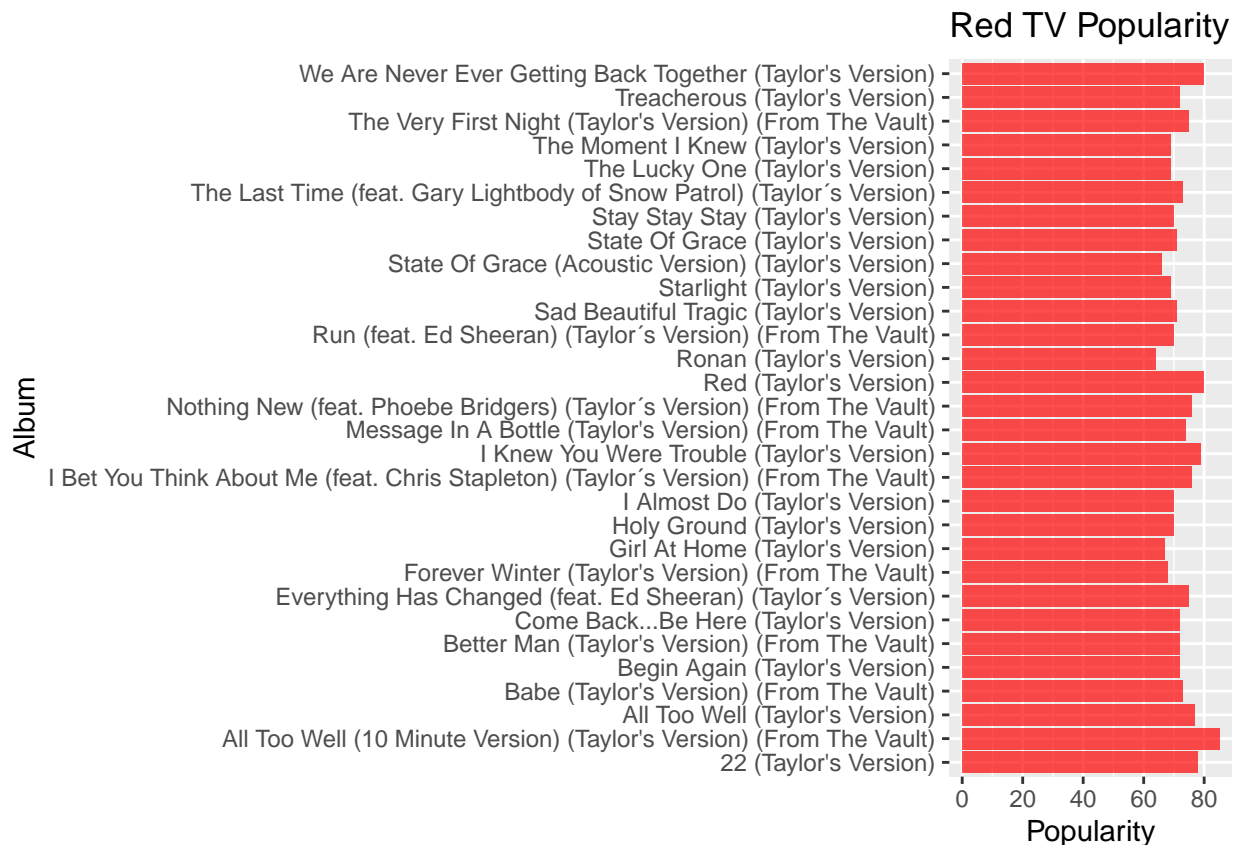
```
## Warning: Continuous x aesthetic
## i did you forget `aes(group = ...)`?
```



```
# Print the plot  
print(REDSVBAR)
```

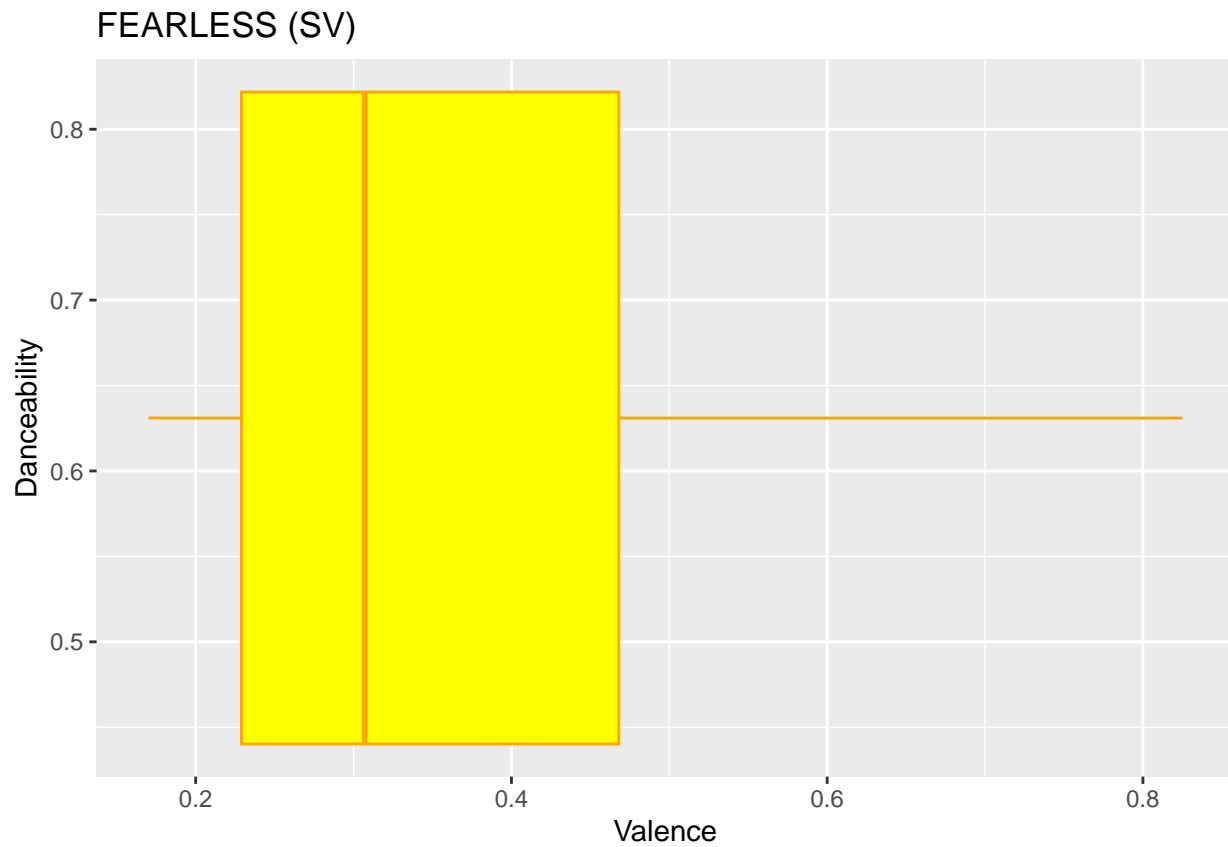


```
# Print the plot  
print(REDTVBAR)
```



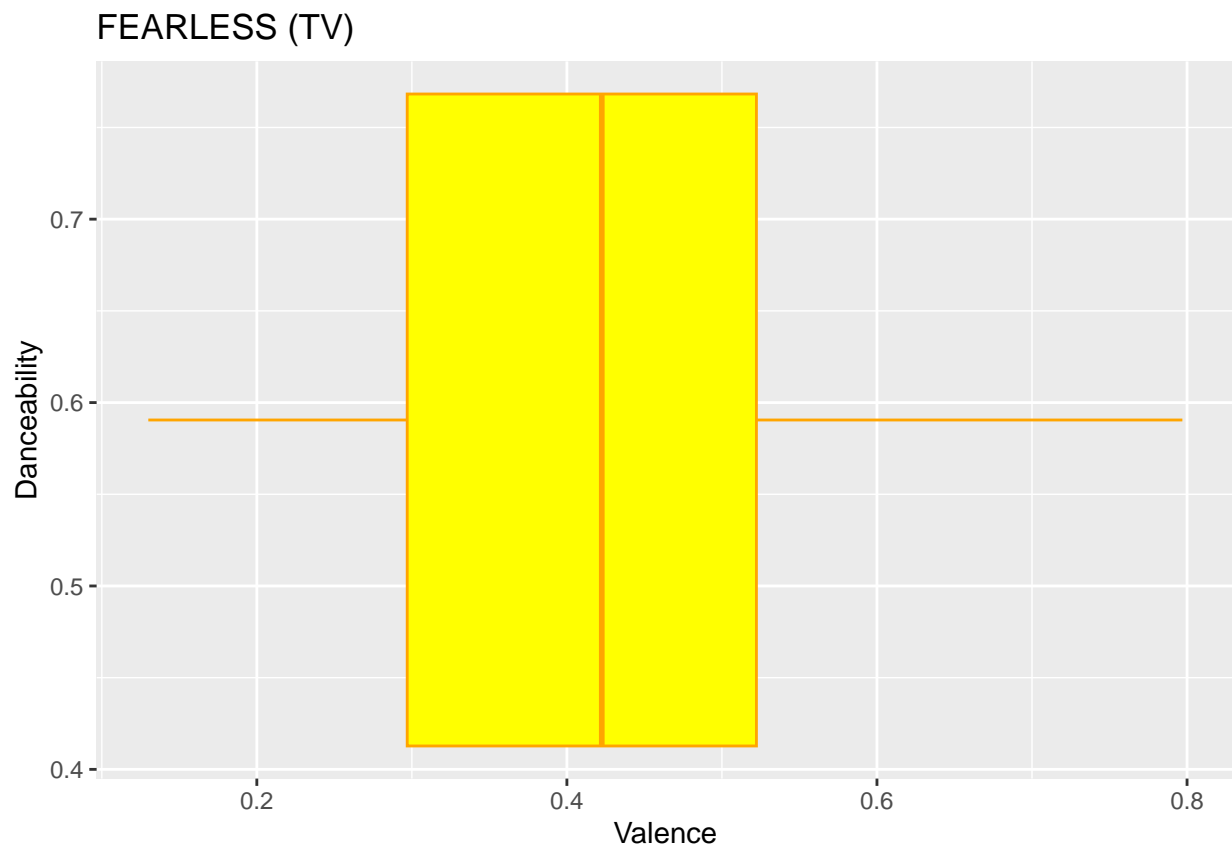
```
##boxplot FEARLESS
ggplot(FEARLESSSV, aes(x = danceability, y =valence )) +
  geom_boxplot(fill = "yellow", color = "orange") +
  labs(title = "FEARLESS (SV)",
        x = "Danceability",
        y = "Valence") +
  coord_flip()
```

```
## Warning: Continuous x aesthetic
## i did you forget `aes(group = ...)`?
```

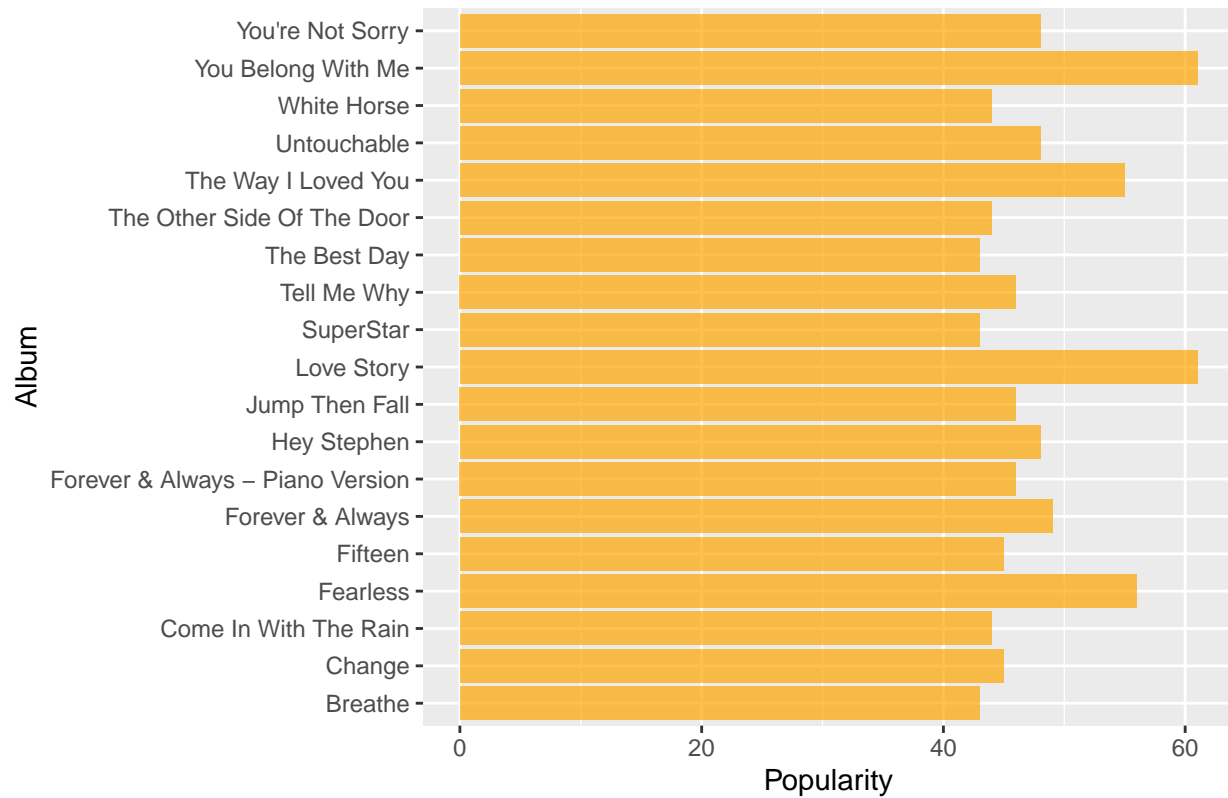
```
##boxplot
ggplot(FEARLESSTV, aes(x = danceability, y =valence )) +
  geom_boxplot(fill = "yellow", color = "orange") +
  labs(title = "FEARLESS (TV)",
        x = "Danceability",
        y = "Valence") +
  coord_flip()
```

```
## Warning: Continuous x aesthetic
## i did you forget `aes(group = ...)`?
```



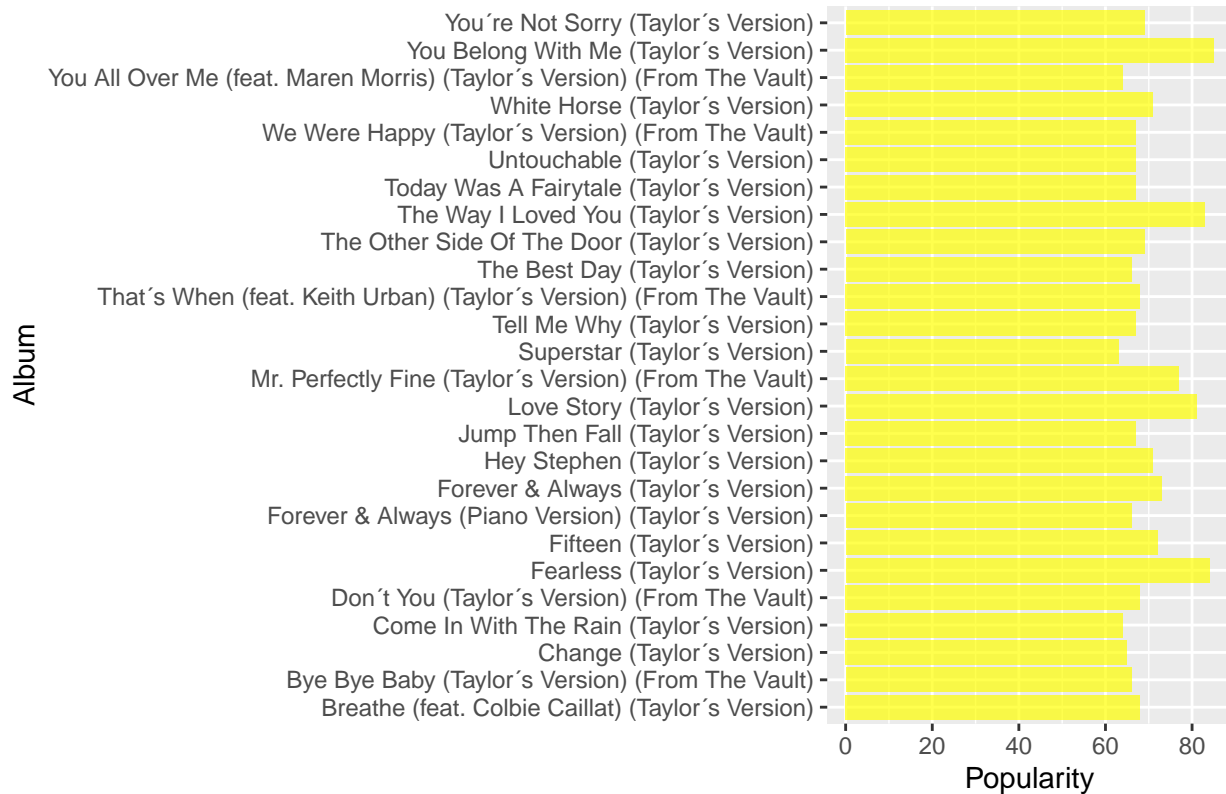
```
# Print the plot  
print(FEARLESSSVBAR)
```

FEARLESS SV Popularity



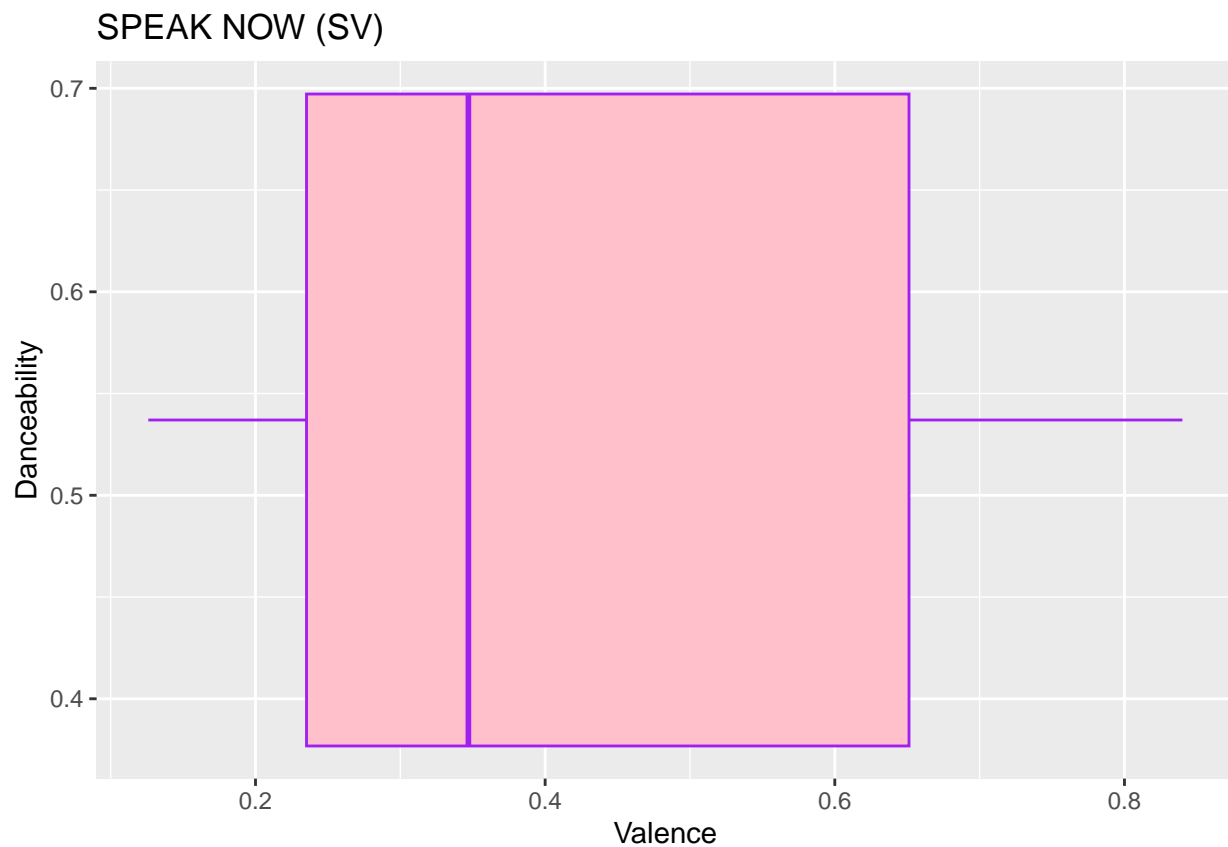
```
# Print the plot  
print(FEARLESSTVBAR)
```

FEARLESS TV Popularity



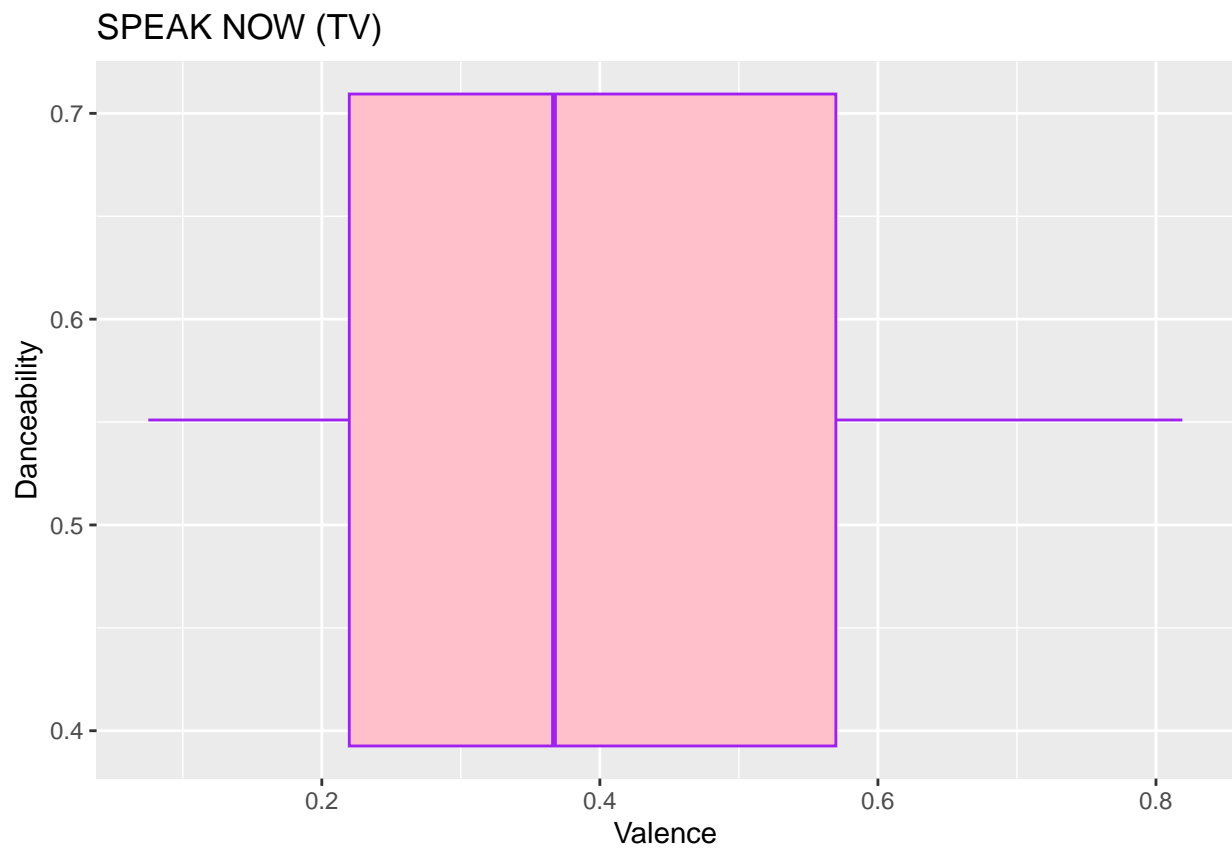
```
##boxplot
ggplot(SPEAKNOWSV, aes(x = danceability, y =valence )) +
  geom_boxplot(fill = "pink", color = "purple") +
  labs(title = "SPEAK NOW (SV)",
        x = "Danceability",
        y = "Valence") +
  coord_flip()
```

```
## Warning: Continuous x aesthetic
## i did you forget `aes(group = ...)`?
```



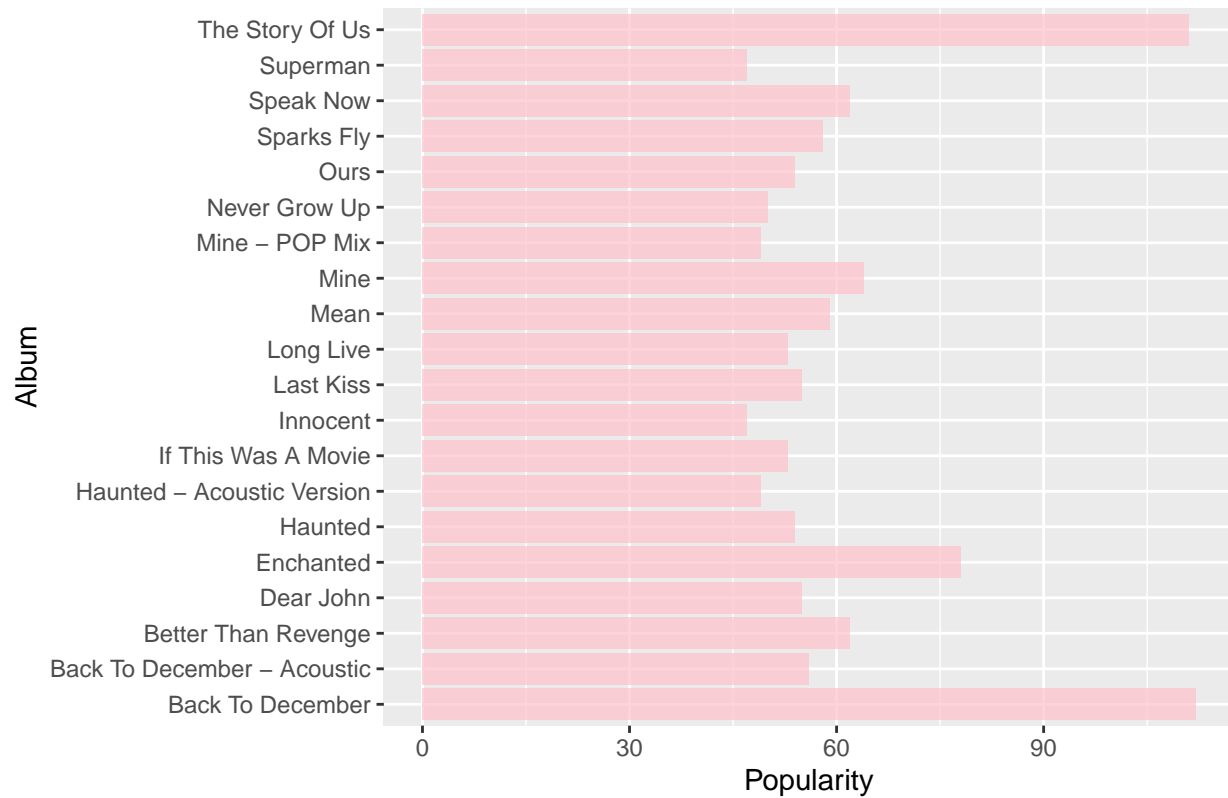
```
##boxplot
ggplot(SPEAKNOWTV, aes(x = danceability, y =valence )) +
  geom_boxplot(fill = "pink", color = "purple") +
  labs(title = "SPEAK NOW (TV)",
        x = "Danceability",
        y = "Valence") +
  coord_flip()
```

```
## Warning: Continuous x aesthetic
## i did you forget `aes(group = ...)`?
```

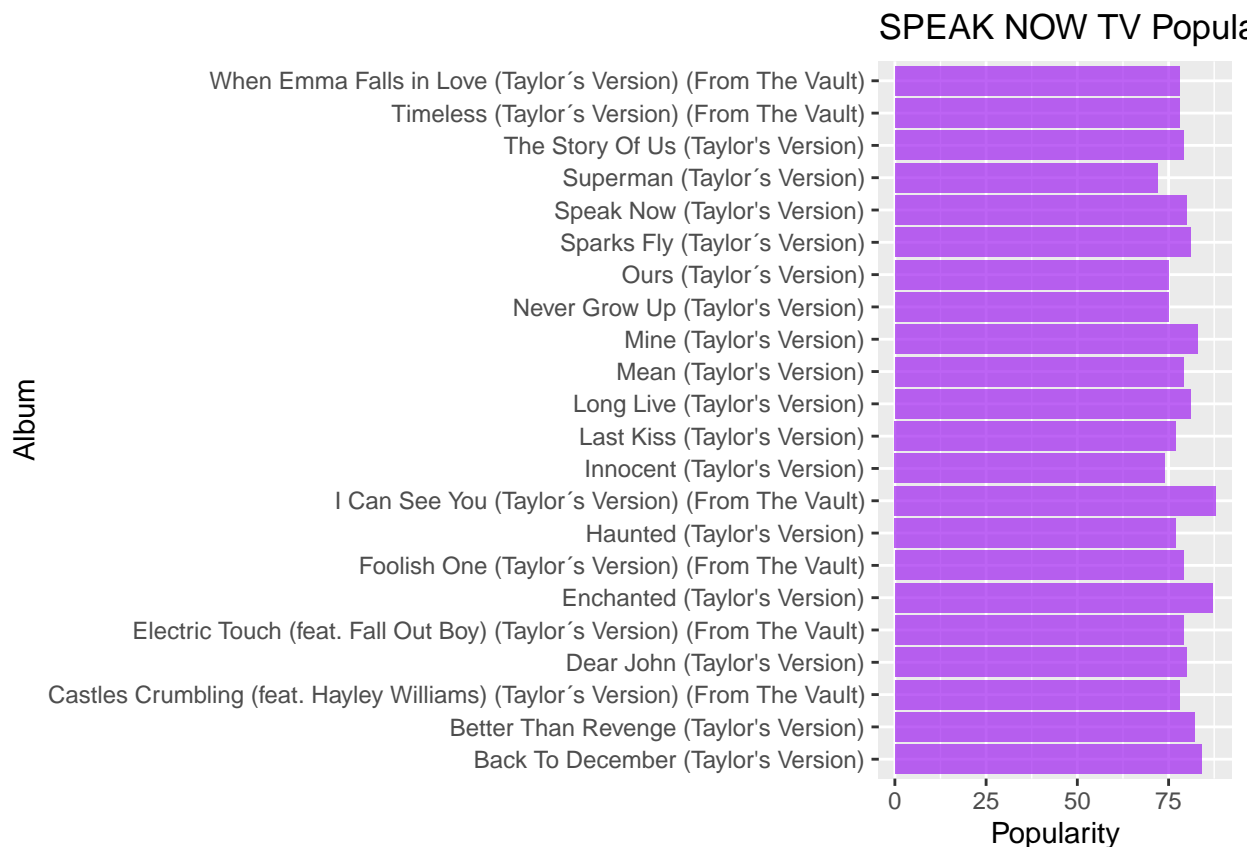


```
# Print the plot  
print(SPEAKNOWSVBAR)
```

SPEAK NOW SV Popularity



```
# Print the plot
print(SPEAKNOWTVBAR)
```



Conclusion In my analysis and shown above the RED, Fearless, and Speak Now albums have the similar Danceability and Valence, yet Taylor's Versions are more popular. However you can also see that Speak Now's top songs The Story of Us and Back TO December were more popular in their original version, but the rest of the album was not nearly as popular as they are for the re-records. You can deduce that the Taylor's Versions albums are more popular not because the albums are something new and different but because Taylor's fan base would rather support her owning her own music.

terms explained:

danceability - Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is the least danceable and 1.0 is most danceable.

popularity - the popularity of the song from 0 to 100

valence - A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).

energy - Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.

tempo - The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.