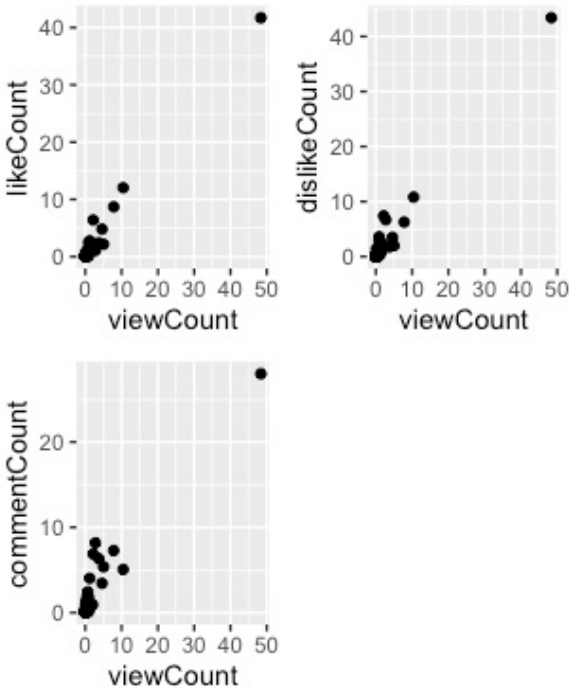
Stephanie Sears

Capstone Project: Applying Statistics

In my sentiment analysis project, I observe data for three official YouTube channels of popular R&B singers—Patti LaBelle, Whitney Houston, and Mariah Carey. As part of an overall objective to understand how social media users understand popularity, the dataset (Videostats) I created for this project consists of view counts, like counts, dislike counts, and comment counts. These are all metrics readily available on the YouTube social media platform. However, given this project's focus on popularity, a popularity metric needed to be defined and created. The popularity metric, named pop_score, was created by taking the sum of view, like and comment counts and then subtracting dislike counts. In order to analyze the data, it was necessary to normalize the data to have a standard basis for comparison. The resulting numerical data points to some interesting insights.

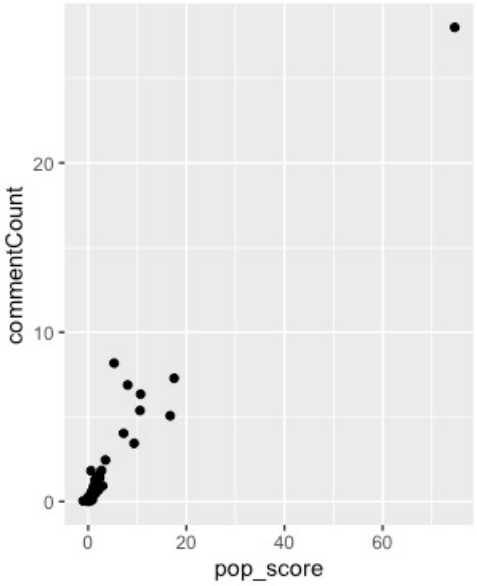
Scatterplots for Video-stats



These plots show how likes, dislikes, and comments relate to the number of views. Each plot shows that there is a positive linear relationship between each of the variables and views. In each of the plots there is an outlier, which will require further exploration. However, there is a degree of consistency among the three plots, with much of user response concentrated around the smaller increments of views. As views increase, liking, disliking and commenting activity seems to cease.

Stephanie Sears

Since the the distribution between commentCount and viewCount appears to have more nonlinearity, I thought it would be interesting to see what patterns emerge between comments and the pop_score. The plot appears as follows:



Of course, there is that outlier that requires further investigation. However, the plot seems to suggest that the more popular the artist is, the more likely it is to have commenting on their videos. Further investigation into the nature of the popularity of each artist is needed to understand these variables more deeply.

Examining Pop_Score via Histogram

In the histogram, the data is skewed right. On the right side of the graph are the less popular videos. To understand popularity even more deeply, the histogram helps to identify where the videos with the highest concentration of pop_score values fall among the overall distribution of videos. It is now possible to delve further into this subset of the data to explore any sentiments associated with it.