Data Visualizations and Observations
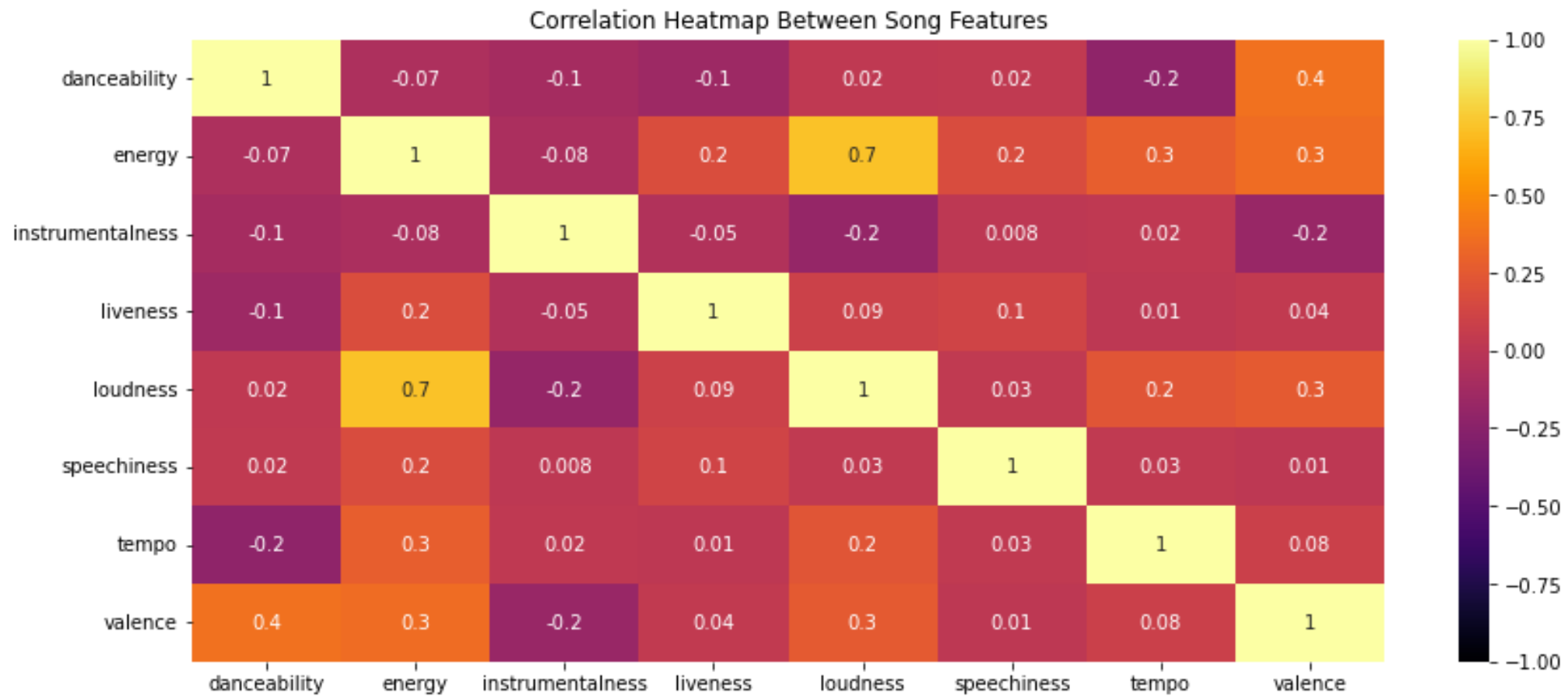
Sarah Schwartz

Semtember 29, 2022

Figure 1. The heatmap show the correlation between all of the numerical track features. The closer the color is to light yellow and the closer the value it to 1, the more positively correlated the two features are. Conversely, the closer the color is to black and the value is to negative one, the more negatively correlated the two values are. Loudness and energy appear to have a strong positive correlation, with an R-value of 0.7. Most other values have no correlation (close to zero) or are only somewhat correlated. The most interesting observation to me is that danceability and tempo have a slightly negative correlation.
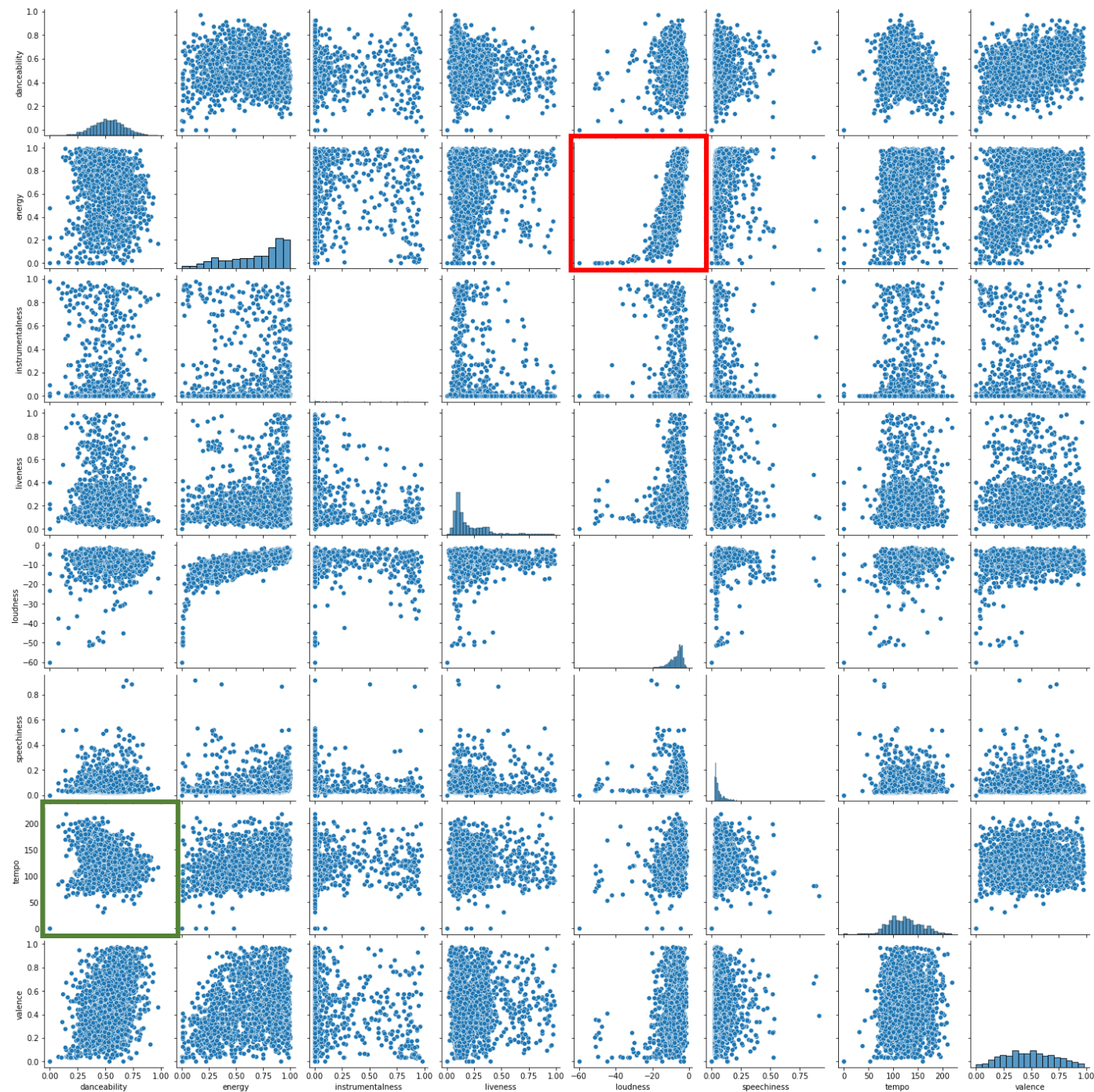
Figure 2. Here are the scatter plots for the same numerical track features from Figure 1. The box in green is the Tempo vs. Danceability that had a negative correlation with an R-value of -0.2 in the previous figure. While the red box highlights the Energy vs. Loudness scatterplot data the was shown to have a positive correlation of 0.7.
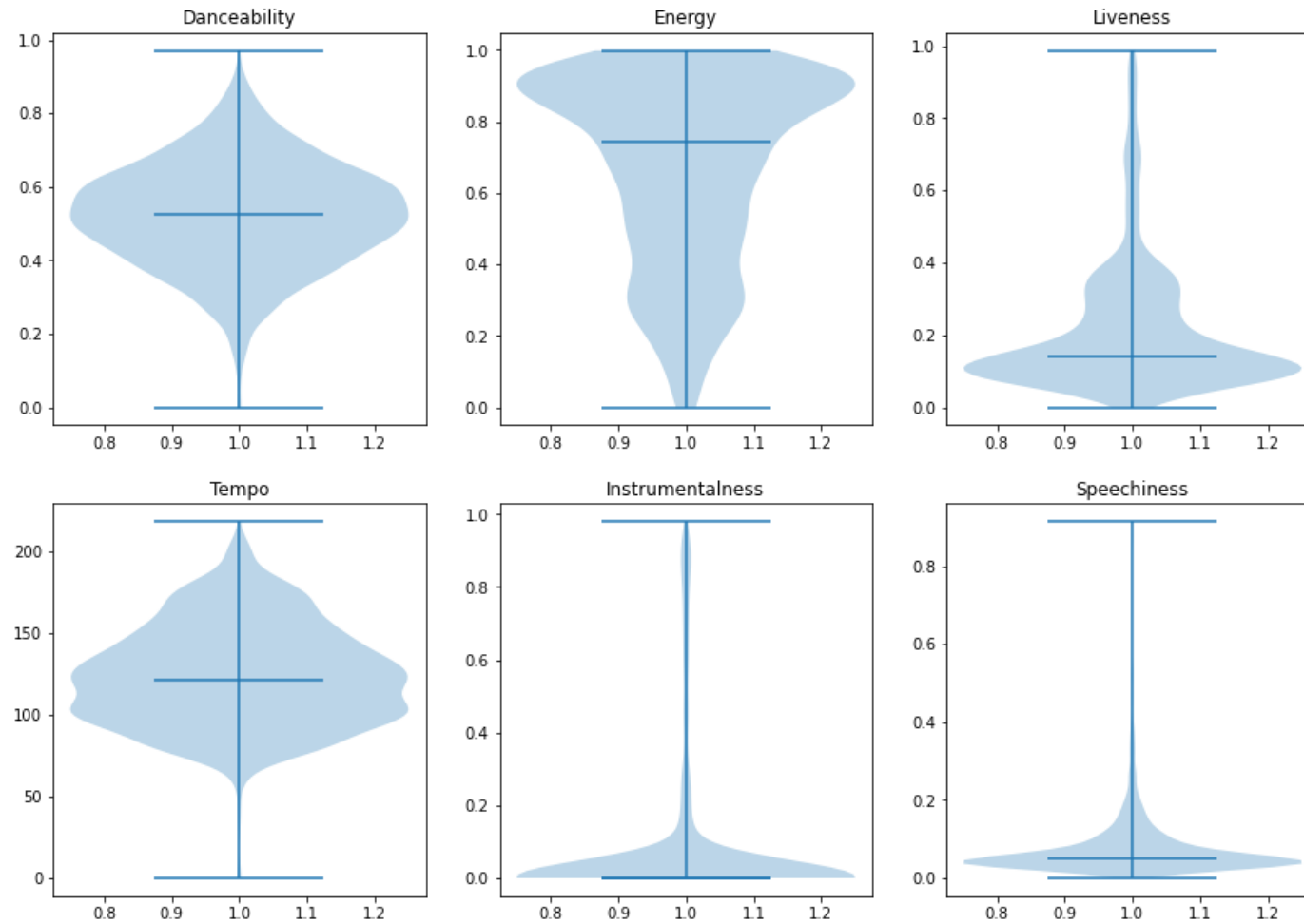
Figure 3. Above is the violin plots for Danceability, Energy, Liveness, Tempo, Intramentalness, and Speechiness. For each feature, The minimum, maximum and mean values are all represented as bars, while the overall distribution is shown by the light shading. Danceability is the most similar to a normal distribution. I find the distribution of the Intramentalness to be interesting since it appears that the mean and minimum values are very similar, but there is a small group of highly instrumental tracks.
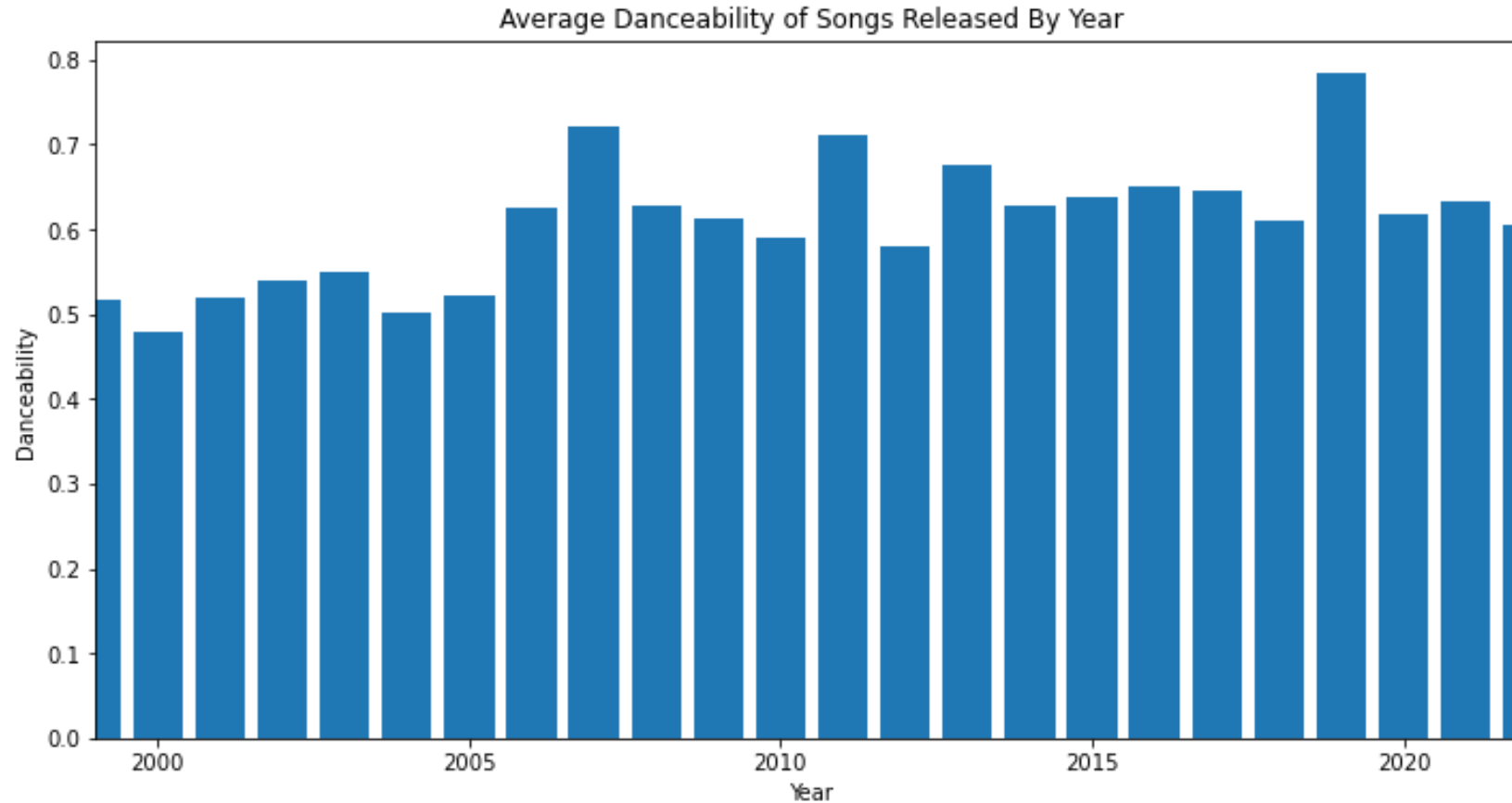
Figure 4. This last chart looks simple, but don't be deceived! There might be an influence from the Covid-19 pandemic in the data. Above is the average danceability of songs released each year. The Danceability for 2019 looks to be much higher than that of the following year. In fact, The average Danceability for 2019 and 2020 are 0.623873 +- 0.174231 and 0.560098 +- 0.12276 with sample sizes of 92 and 102 tracks respectively. The p-value between these two values is 0.034 or there is only a 3.4% chance that they are the same value. This can be considered a statistically significant different. However, comparing the values for 2018 and 2020, they are much more similar, so it is possible that 2019 was just a great year for dancing music.