

report discussing pros and cons for each of the seven displays

- Correlation plot:
  - Pro: Majority of data is numerical, and is expected to have some kind of correlation... this is a great way to see how variables correlate to each other... for example streams to danceability to energy... if all 3 have good correlation we can start to form some better hypothesis about what leads to good streams and what are some factors included together.
  - Con: pearsons correlation doesnt do well with categorial variables, so werent able to include them, and missed out on interpreting that data... like key or month released..
- Scatter plot matrix
  - Pro: was able to understand the distribution of each variable relationship much better... a lot of non linear relationships, which helps interpret the previous correlation plots better too.
  - Con: Slightly messy display, variable scales were also very different in some cases, even when standardizing.
- Parallel Coordinates
  - Pro: Was able to clearly show that acousticness% is inversely proportional to energy%, and we are able to see a similar but less strong relationship with the song descriptions going forth, by observing a zig zag motion.
  - Con: A lot of values in the dataset contained 0 especially in the streams and playlists categories, either by purpose or not, which affected this plot a lot, and shows that it probably affected the other as well, a little difficult ot interpret but it is manageable.
- PCA plot
  - Pros: Shows the distribution of PC1 and PC2 clearly. PC1 seems to capture more of the outlier variance, but PC2 shows that it captures a stable variance among the more frequent values. Scree plot also clearly shows that the first 3-4 PCs capture the most amount of variance.
  - Cons: Scatter plot of PCA isnt the best shape.
- Biplot
  - Pros: Once again shows that energy and acousticness run in completely different directions, but now clearly shows us that energy and valance and inverse to acousticness as well, and they follow eachother. Also shows us that streams, spotify playlists, deezer and apply playlists are cause for the outliers in PC1...
  - Cons: Some of the vectors are close to each other, and is hard to read. Streansm spotify, apple, deezer are basically 90 degrees to the song descriptors, which shows that they have little correlation with each other.
- MDS Plot (data)
  - Pros: The data seems to be close together, indicating that the data points are all somewhat similar. There are outliers, but there is not obvious distribution.

- Cons: Since everything is somewhat similar, there are no clusters, and we are unable to group datapoints into certain clusters...
- MDS Plot (attributes)
  - Pros: more defined clusters than the data plot. We can clearly see that apple, spotify, deezer playlists are correlated with streams, and that the char variables are clustered. We can also see that artist\_count, bpm, and released day are kind of close to each other, something to think about. We can also see how close each of the song descriptions are.
  - Cons: Graph distances between the points are a little difficult to interpret results, unsure how impactful / meaningful the distances are between the points without scale.

are there any interesting findings you can make?

- We can clearly see that there is an inverse relationship between acousticness and energy, danceability, and valence, while these 3 are closely correlated. Furthermore, the amount of streams do not seem to be dependent on these, meaning that the quality of the song or how popular does not solely depend on the type of music it is. It can be a song that people dance to at the club, or it can be a song that someone listens to that is slow paced or calming.
- We see that there is not much of a difference in popularity between charts, meaning people among different music distribution platforms reach the same songs similarly, and that the platform itself pays little difference in making a song popular.
- The more important factor of streams is being in more playlists, not necessarily what number on the chart it made it on.

what information of your data do these displays show well

- Shows the information between the continuous variables well, such as the % indicators of song descriptions... and the correlations between chart and playlist numbers to streams

what information can't they show

- Struggles to show the relationship between categorical variables. Also, some of the visualizations does not account for non-linear relationships, so although the overall picture is similar and supports each other across the visualizations, some finer detail is lost. Really would like to investigate more about how different key combinations, and modes, release month and days, diving into releasing on Mondays vs weekends, play a difference in streams or even song descriptors...