INF4000 Data Visualisation

240175267

Analysing how Key Audio Features vary across different categories of Valence in the Spotify Tracks Dataset

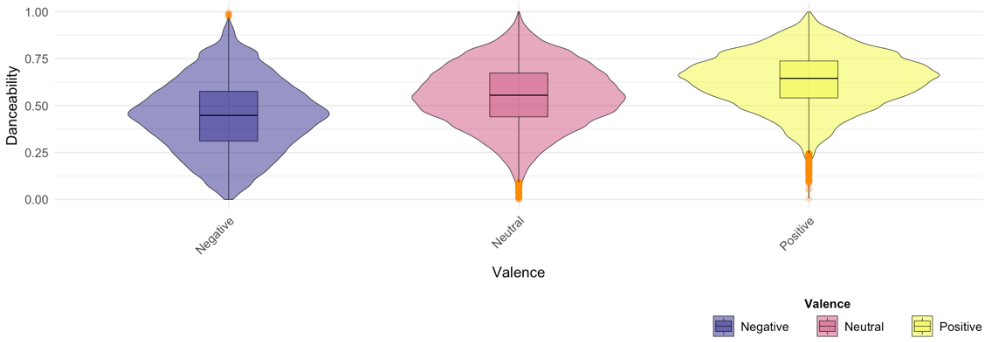
Word Count: 3289

How do Key Audio Features vary across difference categories of Valence?

Investigating relationships in the Spotify Tracks Dataset

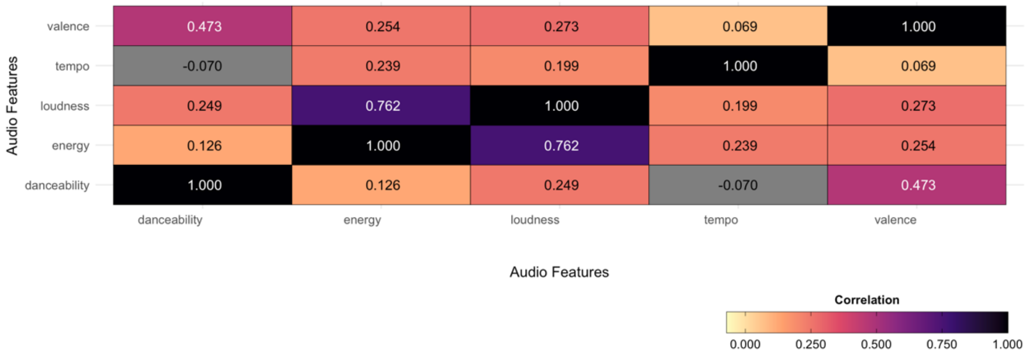
Plot 1.

Distribution of Danceability across each Valence category



Plot 2.

Correlation Matrix of Key Audio Features and Valence

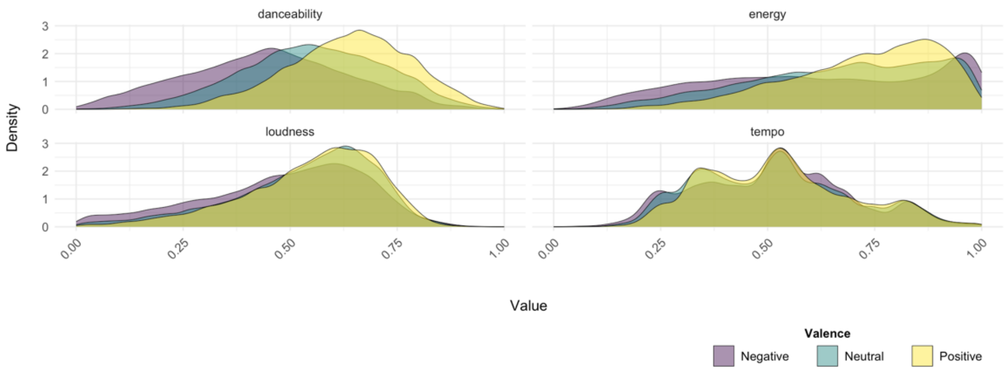


PLOT 1. KEY TAKEAWAYS:
This distribution suggests that positive tracks are consistently more danceable while negative tracks tend to be less danceable, but with greater variation.

PLOT 2. KEY TAKEAWAYS:
The matrix shows that the strongest relationships exist between energy and loudness, valence and danceability, and valence and loudness, suggesting that valence and energy influence the danceability and loudness of tracks.

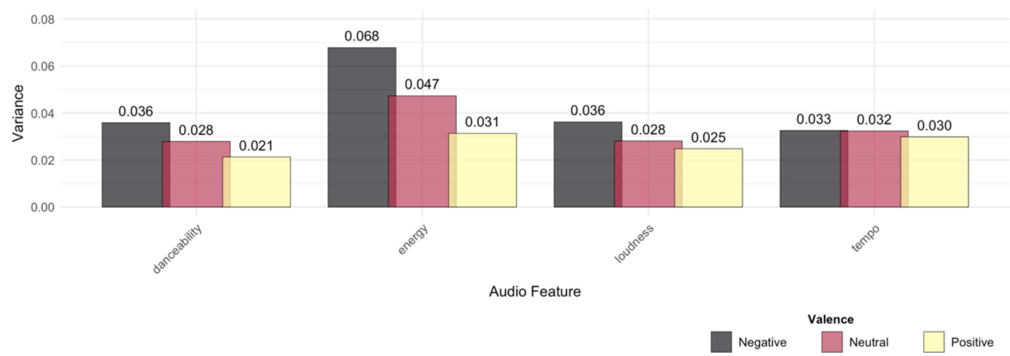
Plot 3.

Density Distributions of Key Audio Features by Valence category



Plot 4.

Variance of Key Audio Features by Valence category



PLOT 3. KEY TAKEAWAYS:
The density distributions show that positive tracks tend to have higher danceability, energy and loudness, while tempo is less affected by valence.

PLOT 4. KEY TAKEAWAYS:
The bar charts indicate that energy exhibits the highest variance in tracks with negative valence. As valence increases variance in energy decreases, suggesting that more positive tracks have more consistent energy levels.

Valence is a measure of musical positiveness and is a measurable indicator of the mood of a track. Tracks which sound positive score higher and those which sound negative score lower. Source: Spotify Tracks Dataset

1. Knowledge Building

The research question we aimed to address as part of this project, a specific side question of interest in follow up to the INF6027 research project analysing the Spotify Tracks Dataset, was:

How do key audio features vary across different categories of valence?

This question holds significance, particularly in the context of music streaming platforms, such as Spotify, where music recommender systems are now integral to helping users navigate vast music catalogues more efficiently (Afchar et al., 2022). A better understanding of the relationships between key audio features such as *danceability*, *energy*, *loudness* and *tempo*, and valence (mood) could ultimately result in the improvement of these systems and streaming platforms, who already employ music recommendation algorithms to curate personalised playlists for subscribers and could facilitate deeper personalisation through emotionally relevant playlist curation based on individual user preferences and mood.
The research question was considered narrow enough in scope that it could be answered effectively with several visualisations, but not so narrow that those visualisations would essentially each be an illustration of the same information presented in a slightly different way. To answer the question, a composite visualisation combining four individual plots was developed to explore how these key audio features vary across different valence categories. Prior to visualisation the continuous *valence* variable was transformed into a categorical form by defining three categories of equal size denoting Negative, Neutral and Positive valence, followed by the definition of ordered levels for the categories using the `factor()` function in R:

Plot 1, violin plots overlayed with boxplots, focussed specifically on the key audio feature which correlated most strongly with valence, *danceability*. This combined plot visualised the distribution of *danceability* across negative, neutral and positive levels of valence and facilitated a comprehensive illustration of both distribution and summary statistics across valence categories, in addition to identifying outliers,
Plot 2, a correlation matrix, was designed to visualise the strength and direction of the of the relationship between valence (in its untransformed, continuous form) and each key audio feature in addition to that between each key audio feature pair,
Plot 3, a faceted density plot, centred on the distribution of each of the four key audio features within each categorical level of valence. This plot would be able to provide insight into the shape, spread and skew of each feature,
Plot 4 comprised bar charts representing the variance of each key audio feature in each of the three valence categories. These charts would be able to highlight which key audio features were more consistent or variable for particular categories of valence.

Collectively, the four plots were able to answer the research question by exploring the behaviour of each key audio feature across the three categories of valence. The key findings which resulted from the creation of the composite visualisation and its component parts were:

- The distribution of *danceability* in Plot 1 suggests that positive tracks are consistently more danceable while negative tracks tend to be less danceable, but with greater variation,
- The correlation matrix which forms Plot 2 demonstrates that the strongest relationships exist between *energy* and *loudness*, *valence* and *danceability*, and *valence* and *loudness*, suggesting that valence and energy influence the danceability and loudness of tracks,
- The density distributions represented in Plot 3 show that positive tracks tend to have higher danceability, energy and loudness, while tempo is less affected by valence,
- The bar charts which comprise Plot 4 indicate that energy exhibits the highest variance in tracks with negative valence. As valence increases, variance in energy decreases, suggesting that more positive tracks have more consistent energy levels.

These key findings concur in part with the work of Panda et al. (2021) who identified energy, valence and acousticness as being highly relevant to Music Emotion Recognition (MER) but do not align with the findings of Pyrovolakis et al. (2022) who noted that tempo was an audio feature which was central to mood classification. The findings suggest that personalised music recommendations could be enhanced through mood-based categorisation of tracks.

2. Theoretical Frameworks

The aim of Plot 1 was to explore how *danceability* varies across negative, neutral and positive valence categories. The *danceability* key audio feature was chosen for this initial visualisation as it was the feature correlated most strongly with *valence*. Understanding the distribution of *danceability* across different valence levels could provide insight as to whether the underlying mood of a music track makes listeners want to dance.

While formulating ideas regarding the creation of appropriate visualisations, it was important to consider the ASSERT model. This model is a framework for applying traditional research methodologies to the development of visualisations (Ferster, 2023) and is composed of six components: **A**sk, **S**earch, **S**tructure, **E**nvision, **R**epresent, **T**ell.

The first element of the framework, **Ask**, precedes visualisation design, and considers possible topics and questions. During this stage, it was important to consider the scope of any questions and ensure that they could be meaningfully answered. In the context of Plot 1, in addition to the overarching research question, the pertinent question was:

How does the distribution of danceability differ across the three valence categories?

This question is directly related to the main research goal and its aim of exploring how a key music feature, in this case *danceability,* correlates with mood, categorised by *valence*. It is more focussed than the broader research question, and could be answered with a single, appropriate visualisation, but is an element key to being able to answer the research question.

The **Search** component, which again preceded design, involved gathering relevant information and conducting analysis to identify the key audio features that were most strongly correlated with *valence*. This analysis ultimately identified that *danceability* was most strongly correlated with *valence*, providing the foundation for the visualisation in Plot 1 of the composite.

When following the **Structure** component of the ASSERT model, it was necessary to organise the raw data in a way which made it suitable for analysis and visualisation. During this stage of the process, the data was cleaned to ensure that missing values were addressed, and then subset and pre-processed to remove significant outliers and normalise the key audio features with one another. The continuous *valence* variable was transformed into three, equally sized categories denoting negative, neutral and positive valence with their resulting ordered levels defined using the `factor()` function in R.

The next step, **Envision**, was an experimental process used to establish which visualisation type would be most suited to conveying the data to the audience and in doing so answer the question considered earlier in the process. Multiple visualisation types were prototyped through an iterative process, including boxplots and histograms, before a final design was selected for Plot 1 which would clearly represent the distribution of *danceability* across negative, neutral and positive valence levels.

The **Represent** phase of the process involved refinement of the chosen visualisation to ensure that its design was able to convey its message optimally. The chosen visualisation for Plot 1, a set of violin plots overlayed with box plots, although relatively simple, was able to provide both detailed and summarised insight and offer a suitable trade-off between information clarity and information load. The visualisation was created using ggplot2, an R package used for data visualisation, which utilises a layered grammar of graphics to build graphics, rather than following a data flow. The grammar of graphics is a layered framework, where each layer represents a distinct and independent component. These components are combined, with every possible combination creating a meaningful visualisation (Wilkinson, 2012). In the case of ggplot2, the approach to structuring the grammar focusses on constructing a graphic through multiple layers of data (Wickham, 2010). For Plot 1, the subset key audio feature data was defined as the plot object that would serve as the basis for the visualisation, followed by mapping the variables to the plot aesthetics to define the x- and y-axis variables and the valence categories which would be used to fill the violins and boxes with colours to differentiate them. The next step was to add further layers by defining the geometric objects `geom_violin()` and `geom_boxplot()` which ggplot2 would use to represent the defined key audio feature data. This procedure was followed to add plot titles, subtitles, captions and axis labels, to define the colour palette which was to be used and to define a theme and its associated elements. Plot 1 utilised a traditional cartesian co-ordinate system with its x-axis representing the valence categories and the y-axis representing *danceability*, but as this co-ordinate system was the default for this visualisation type, ggplot2's layered grammar of graphics did not require that this was explicitly defined.

The final element of the ASSERT model, **Tell** was an opportunity to ensure that the end-user could effectively interpret the data presented in the visualisation. The visualisation successfully communicated the story of how the distribution of *danceability* in Plot 1 suggests that positive tracks are consistently more danceable than negative tracks, but that negative tracks show greater variation in danceability than positive tracks do. Additionally, it highlighted that tracks within the neutral valence category occupied the range in between the positive and negatives categories.

Following the steps described in the ASSERT model resulted in the creation of a comprehensive visualisation which was well placed to answer the question at hand. The grammar of graphics utilised in ggplot2 guided the selection of specific plot elements and ensured that a clear, well-structured visualisation was produced.

3. Accessibility

In the context of data visualisation, accessibility is a requirement to ensure all intended end-users of a visualisation can clearly interpret the visualisation and, in doing so, understand its message unhindered. Accessibility considerations should be central to visualisation design as users with additional accessibility requirements, such as a disability or visual impairment should be able to successfully interact with a visualisation in just the same way as those without these requirements.

Plot 2, a correlation heatmap, was created with accessibility in mind, to facilitate simple visualisation of the relationships between key audio feature pairs and *valence*, and incorporated design elements to meet these accessibility requirements. A gradient legend was included to allow the user to quickly reference and establish approximate correlations in addition to heatmap tiles annotated with precise numeric correlations corresponding to specific key audio feature pairs, for clarity and ease of interpretation.

Appropriate use of colour is an essential consideration when ensuring accessibility in design, and creators must ensure their visualisations are tailored to users with visual limitations such as colour vision deficiency. An iterative process was followed in creating Plot 2 before arriving at a choice of colour palette that was both accessible and visual appealing, using the `viridis` package in R, which provides perceptually uniform colour maps designed to improve graph readability for those with common forms of colour vision deficiency (Rudis et al., 2024).

Experimentation was necessary to ensure there were no issues with insufficient contrast between tile colours and numeric annotations which may have affected legibility, as reduced print contrast limits the accessibility of text for many people (Crossland & Rubin, 2012). Optimal contrast was achieved by rendering annotations in white for tile colours associated with correlations > 0.4. Narrow black borders were added to the tiles to demarcate the area of the plot corresponding to each key audio feature pair and clearly separate similarly coloured adjacent tiles.
When selecting text style for the plot, the ggplot2 default typeface, Helvetica was an appropriate choice due to its simple, well-spaced, sans-serif design. This ensured clarity and legibility and is a typeface known to be accessible for dyslexic individuals in terms of its readability (de Avelar et al., 2015). Emboldened text was utilised in conjunction with a variety of font sizes to clearly differentiate labelling of titles, subtitles, axis text, axis titles, legend elements and annotations. All textual elements of the visualisation were

rendered in black, to maximise contrast against the white plot background. Consideration was given to the orientation of the x-axis text labels as when oriented horizontally they crowded the plot. Ultimately, their orientation was adjusted to 45º resulting in a cleaner appearance which facilitated understanding and legibility.
A concise caption was added to the plot, outlining its “Key Takeaways”. Whilst this may have hindered accessibility from a design perspective as it cluttered the plot somewhat, its inclusion was considered essential to ensure the visualisation was accessible to as wide an audience as possible as it provided explanatory content to those with additional needs who may otherwise have struggled to interpret the presented information.
Plot 2, as a whole, was designed with a simple underlying structure and a neutral, minimalist theme so as not to detract from the data itself and with uniformity with the remaining component plots in mind to ensure that having engaged with one element of the visualisation, there was a familiarity of design which aided engagement with the subsequent elements. Extraneous visual flourishes which did not add value to the visualisation were kept to a minimum.

It is vital to consider the form in which the visualisation will be consumed by the end-user. If Plot 2 was to be displayed digitally, the accessibility features it incorporated would be unaffected. If the visualisation was to be printed and displayed, however, the printed version may not remain faithful to the original. In this situation, it would be essential to test the accessibility of the printed version relative to the digital original. To ensure accessibility across a range of media, performing tests with tools such as the WebAIM Contrast Checker (WebAIM: Contrast Checker, 2019) in addition to reviewing the visualisation against the Web Content Accessibility Guidelines (WCAG) (W3C, 2023) would have been extremely beneficial.

4. Visualisation Choice

Plot 3 was created to visualise the distribution of all four key audio features across each valence category. Given the nature of this visualisation, a faceted density plot was chosen. Density plots are smoothed out histograms and are well suited to visualising the distribution of continuous variables which facilitates simple comparison of the shape of distributions between categories (Wickham & Grolemund, 2016).

A faceted plot was chosen, as it would provide clear, comprehensive, side-by-side comparisons of the distributions of all four key audio features without overcrowding. This would offer the end-user the clarity to focus on single key audio features while making cross-valence category comparisons.

Plot 3, as with the other three component plots, used a minimalist design focussing on the plot elements needed to convey the data to the end-user and minimising unnecessary visual embellishments. This aligns with the work of Tufte, who defined the data-ink ratio as the proportion of ink used to represent data compared to the total ink in a plot (Inbar et al., 2007). Tufte aimed to maximise this ratio by minimising non-data ink and redundant data-ink, resulting in a cleaner visualisation that focuses on the data itself (Inbar et al., 2007). As each faceted plot included three density curves representing the distribution of one key audio feature across each of the three valence categories, a degree of transparency was applied to the fill colour. This prevented the curves, where they overlapped, from obscuring one another and allowed a clearer view of the underlying curves which made comparison of the three distributions simpler, minimising non-data ink and enhancing the plot’s interpretability.

The choice of density plots also provided an aesthetic benefit to the visualisation as this plot type incorporates visually appealing smooth curves which can be more interpretable when compared with other visualisation types such as histograms. These curves are able to illustrate the spread and shape of the data across different valence categories, as well as the presence of unimodal, bimodal or multimodal distributions.

Although a density plot was the final visualisation choice for Plot 3, these plots are not without their limitations, and alternative visualisation options were also considered. As Plot 3 was a faceted plot with multiple categories, there was significant overlap of the density curves. Although use of transparency was able to mitigate against this, there was an inevitable loss of clarity when multiple curves were similar. Additionally, while providing the benefits of a smooth distribution curve, density plots do not show exact values for each observation, with each point on the curse representing an estimate of the probability density at that particular value (Jain, 2024) making it difficult to quantify the exact number of data points for a specific value. Histograms could have been used to represent the distributions, but whilst they are effective at visualising frequency distributions and identifying patterns, the smooth, continuous nature of the key audio feature data would have been lost, particularly if inappropriate bin widths had been chosen. Additionally, while histograms are able to visualise the distribution of single variables extremely well, they are not particularly well suited to the side-to-side comparison offered by faceted density plots. Box plots are another means of visualising data distributions which may have been a suitable alternative to faceted density plots and unlike with histograms, a faceted plot could have been created incorporating a box plot for each of the four key audio features across the three categories of valence. Boxplots, however, while illustrating summary statistics and highlighting outliers well, do not capture the shape of a distribution, nor the nuance of the smooth distribution provided by a density plot.

5. Implications and Improvements

When designing a visualisation, it is essential to consider the ethical implications of a particular design as well as those associated with the use of a particular dataset to produce the visualisations. If these considerations are not borne in mine, a visualisation may unintentionally misrepresent the data and mislead the end-user, distorting their view. Whilst no evidence of obvious bias was identified in the Spotify Tracks Dataset used in this project during Exploratory Data Analysis, it is important to consider that the data may not have been a complete representation of artists or music genres as a whole. Furthermore, if this is the case, this may limit the usefulness of derived insights in other contexts. An expansion of the dataset to include a broader diversity of artists and music genres, would facilitate analysis capable of providing a more representative view.

The chosen design for Plot 4 was a grouped bar chart, which visualised the variance in each key audio feature for each level of categorical valence, with annotations above each bar representing the precise numeric variance for each key audio feature across the three valence categories. While this visualisation was clear in its presentation of the distribution of variance, it is important to acknowledge that the potential for bias in the dataset could have influenced the variance results. Additionally, the method used to categorise valence into negative, neutral and positive groups was arbitrary, with no consideration given to the complexity of musical mood. This may have resulting in oversimplified output.

In order to prevent the end-user from overestimating the reliability of the presented data, it may have been beneficial to attach confidence intervals and / or error bars to the chart bars as a representation of the uncertainty of the annotations. Additionally, while a caption was included at the bottom of the composite visualisation, this could have been extended with additional contextual information about the arbitrary nature of the valence categorisation method and how the categories could be compared with the original continuous valence variable. Care was taken to ensure that the visualisation design was kept as neutral as possible and without design cues which might unintentionally bias the end-user, and to further mitigate against any risk of misinterpretation and ensure that the visualisation did not overstate its claims, the “Key Takeaways” information caption adopted a cautious tone. The provision of additional, supplementary information within this section to signpost the end-user to other literature around the data may have increased perceived transparency further.

6. References

Afchar, D., Melchiorre, A., Schedl, M., Hennequin, R., Epure, E., & Moussallam, M. (2022). Explainability in music recommender systems. *AI Magazine*, *43*(2), 190-208.

Crossland, M. D., & Rubin, G. S. (2012). Text accessibility by people with reduced contrast sensitivity. *Optometry and Vision Science*, *89*(9), 1276-1281.

de Avelar, L. O., Rezende, G. C., & Freire, A. P. (2015). WebHelpDyslexia: a browser extension to adapt web content for people with dyslexia. *Procedia Computer Science*, *67*, 150-159.

Ferster, B. (2023). *Interactive visualization: Insight through inquiry*. MIT Press.

Inbar, O., Tractinsky, N., & Meyer, J. (2007, August). Minimalism in information visualization: attitudes towards maximizing the data-ink ratio. In *Proceedings of the 14th European conference on Cognitive ergonomics: invent! explore!* (pp. 185-188).

Jain, A. (2024, January 23). *Everything about Density Plot*. Medium. https://medium.com/@abhishekjainindore24/everything-about-density-plot-3e74bb664d98

Panda, R., Redinho, H., Gonçalves, C., Malheiro, R., & Paiva, R. P. (2021, July). How does the spotify api compare to the music emotion recognition state-of-the-art?. In *18th sound and music computing conference (SMC 2021)* (pp. 238-245). Axea sas/SMC Network.

Pyrovolakis, K., Tzouveli, P., & Stamou, G. (2022). Multi-modal song mood detection with deep learning. *Sensors*, *22*(3), 1065.

Rudis, B., Ross, N., & Garnier, S. (2024, January 28). *The viridis color palettes*. Cran.r-Project.org https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html

WebAIM: Contrast Checker. (2019). Webaim. https://webaim.org/resources/contrastchecker/

Wickham, H. (2010). A layered grammar of graphics. *Journal of computational and graphical statistics*, *19*(1), 3-28.

Wickham, H., & Grolemund, G. (2016). R for Data science: import, tidy, transform, visualize, and model data. https://openlibrary.org/books/OL26836898M/R_for_Data_Science_Import_Tidy_Transform_Visualize_and_Model_Data

Wilkinson, L. (2012). *The grammar of graphics* (pp. 375-414). Springer Berlin Heidelberg.

W3C. (2023, October 5). Web Content Accessibility Guidelines (WCAG) 2.2. Www.w3.org; W3C. https://www.w3.org/TR/WCAG22/

Wickham, H., & Grolemund, G. (2016). R for Data science: import, tidy, transform, visualize, and model data. https://openlibrary.org/books/OL26836898M/R_for_Data_Science_Import_Tidy_Transform_Visualize_and_Model_Data