

OCMP5048 - Week 2 Assignment

Part 1 - Data Types

1. What are the data/attributes/elements of this dataset?

The dataset provided in this assignment comprises [spotify](#) tracks. There are numerous variables, however we will be discussing the most significant ones and categorise them.

- **Genre** - describes the musical style or category to which a particular music track belongs
- **Popularity** - reflects the extent to which a song is known, liked or consumed within the given population or over a specified period
- **Tempo** - represents the speed of a musical track in beats per minute usually. It measures how fast or slow the music feels when played
- **Valence** - represents a measure of musical positiveness or happiness of a track. Usually is measured between 0.0 to 1.0
- **Time_signature** - indicates the musical time signature of a musical track

2. What are their data types?

3. Justify your answers. Why/how do you classify them?

Genre - consists of distinct labels or categories without any inherent order or meaningful numerical value. Hence this is **nominal**

Popularity - Although this variable is represented as numerical values, these values are ordinal in nature. Popularity running may not have uniform intervals between them, and the exact meaning of popularity can be subjective. However, tracks are ranked in a particular order, making it **ordinal** data

Tempo - While it has numerical value, it falls into the interval data type category because it has a consistent interval between values. For example, the difference between 120 beats per minute (BPM) and 130 BPM is the same as between 140 BPM and 150 BPM, however there is no true zero point for tempo

Valence - Typically ranges from 0 to 1, with 0 indicating a negative or sad mood, 1 indicating a positive or happy mood, and values in between representing varying degrees of mood positivity. Valence is considered **ratio** data because it has true zero point (0 represents the absence of mood positivity), and you can perform all mathematical operations on it

Time_signature - categorical labels that define the rhythmic structure of a song. While there may be some subjective complexity associated with different time signatures, they are primarily **nominal** as they represent distinct categories without a clear ordinal relationship

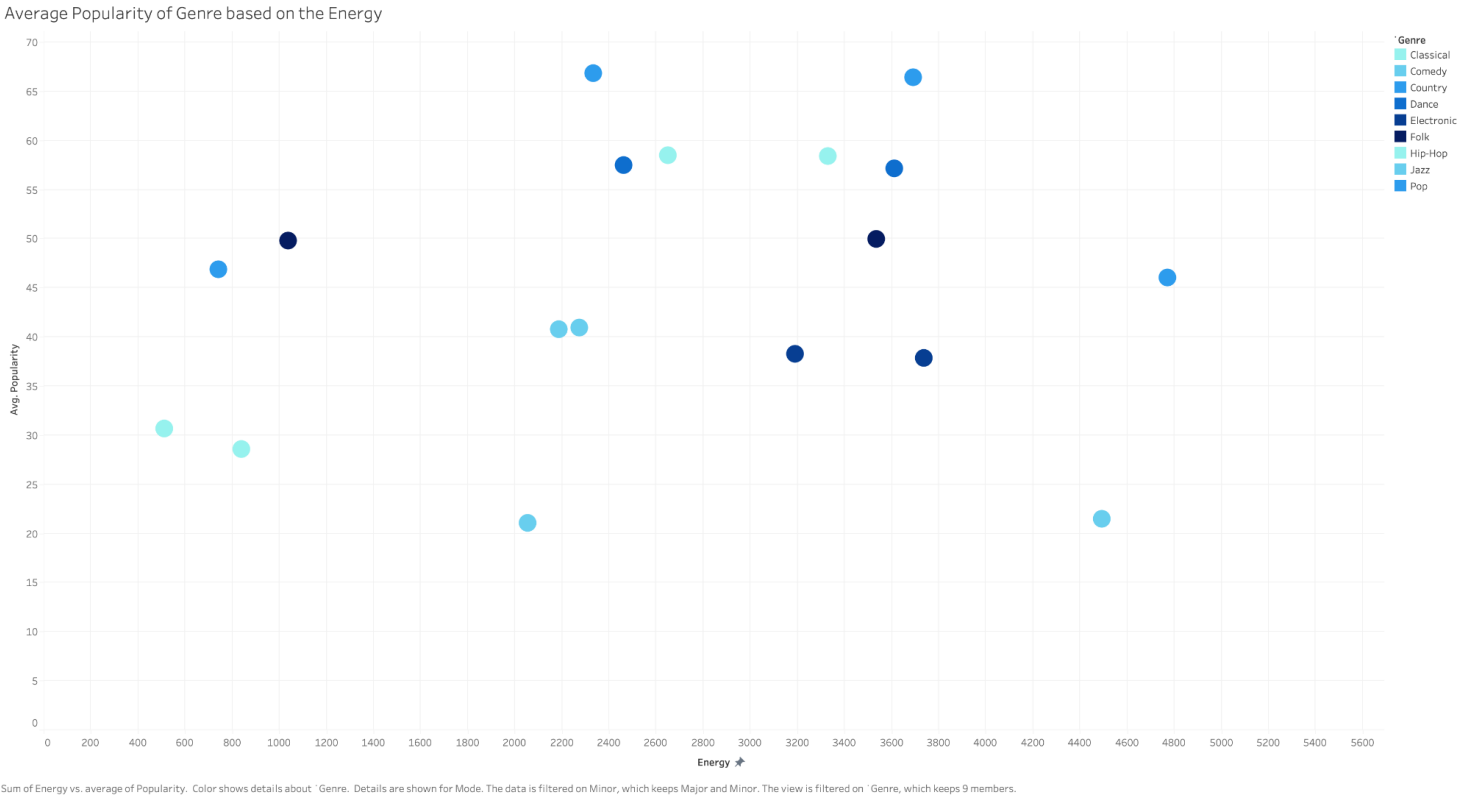
Part 2 - Colour

2.1: Colour & Continuous data

Plot 1: Scatter plot with Tableau 10 colour palette

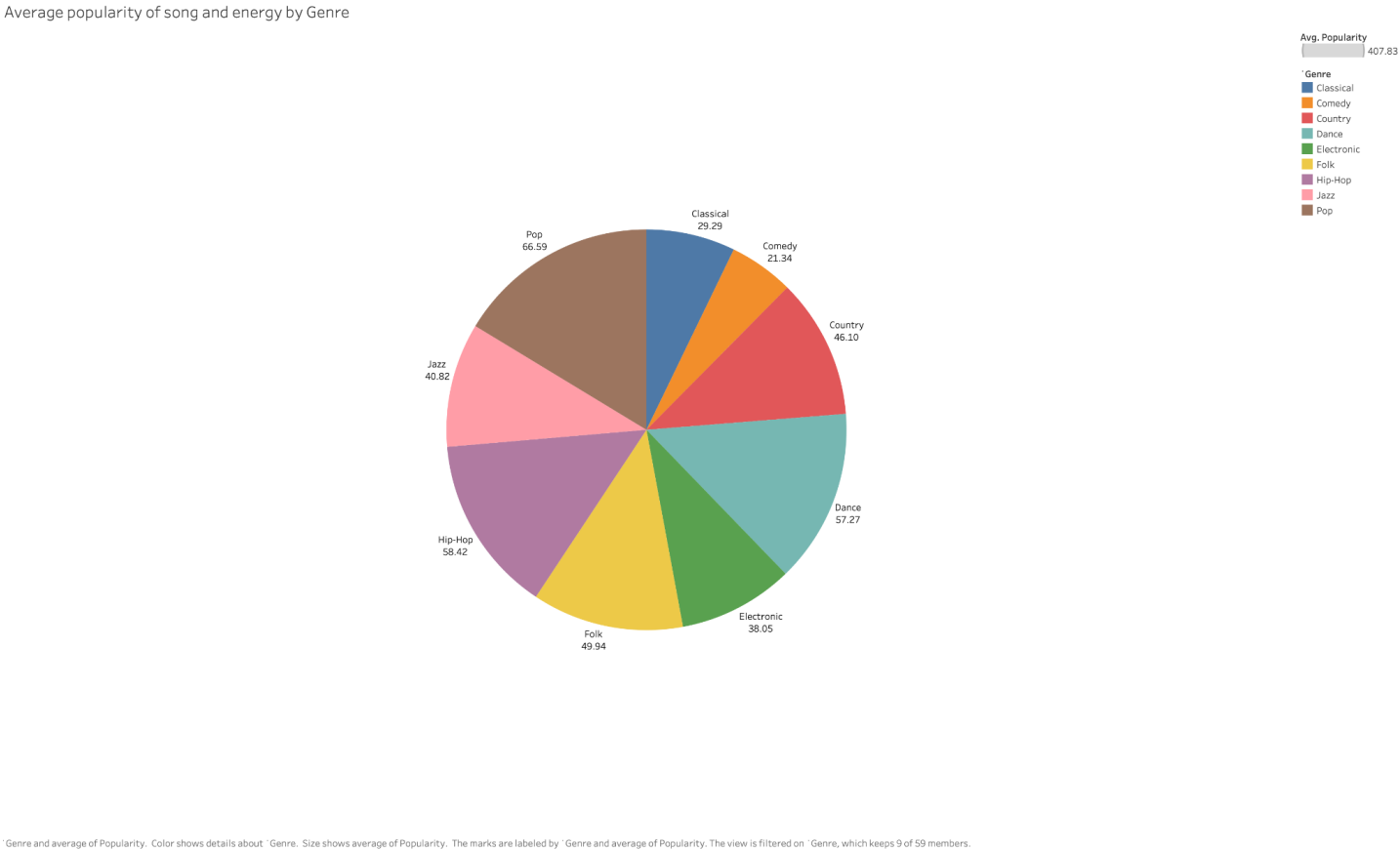


Plot 2: Scatter plot with colorbrewer2 colour palette



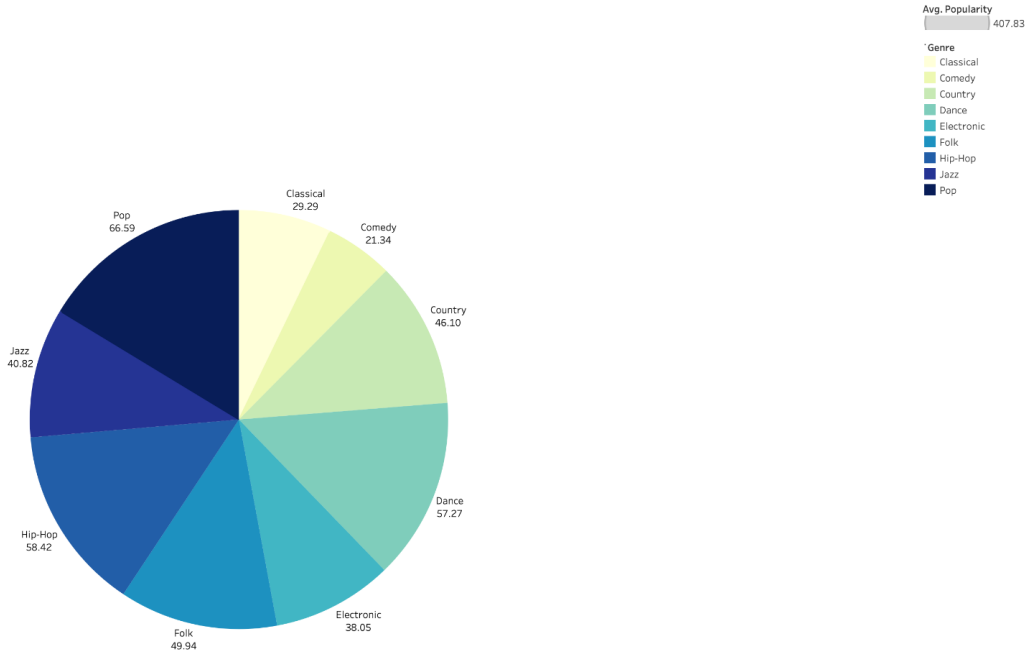
2.2: Colour & Categorical data

Plot 3: Pie chart with Tableau 10 colour palette



Plot 4: Pie chart with colorbrewer2 colour palette

Average popularity of song and energy by Genre



Genre and average of Popularity. Color shows details about Genre. Size shows average of Popularity. The marks are labeled by Genre and average of Popularity. The view is filtered on Genre, which keeps 9 of 59 members.

2.3: What do you observe?

The main observation between the Tableau 10 colour palette and the colorbrewer2 colour palette in plot 1 and 2 is the variety of colours. In plot 1 we observe a wide variety of different colours for each genre, whereas in plot 2, the palette chosen indicates different shades of blue. Although in the context of this visualisation there is no significance in the difference in the shades of blue used in plot 2, it is the differentiator from plot 1. From a visualisation perspective, it is much easier to differentiate the data points in plot 1 than plot 2. When comparing plot 3 and 4, we see that similarly plot 3 has a wider variety of colours, however plot 4 there are not only different shades of blue but also the other colours chosen in the colorbrewer2 colour palette is shown. This is largely due to the 9 stepped colours chosen.