

Lab 1: Basic Data Exploration Using Software Jamovi

Your name

Department name, Institution name

Course number: Course name

Instructor's name

Due date

Data clean-up

Variable Continuous measurement.

The variable “*loudness*” is a variable with continuous measurement. It represents the overall loudness of a track in decibels (dB). It is a continuous variable because it can take on an unlimited number of values between the range, which is -60 and 0 dB.

Variable with Nominal measurement.

The variable “*mode*” is a variable with nominal measurement. A nominal variable is a type of variable that is used to name, label or categorize particular attributes that are being measured. *Mode* is a nominal variable because it is used to categorize the modality of a track into Major and Minor.

Variable with ordinal measurement

The variable “*energy*” should be a variable with ordinal measurement. An ordinal variable is a type of measurement variable that takes values with an order or rank. Energy represents a perceptual measure of intensity and activity from 0 to 1. Energetic tracks feel fast, loud and noisy. High energetic tracks are ranked with a value close to 1 while less energetic tracks are ranked with a value close to 0, hence it is of ordinal measurement.

Frequency tables.

Frequency table of *genre*.

Frequencies of genre

Levels	Counts	% of Total	Cumulative %
alternative	2160	8.1 %	8.1 %
blues	2050	7.7 %	15.7 %
hiphop	2581	9.6 %	25.4 %
indie	4338	16.2 %	41.6 %
metal	3045	11.4 %	53.0 %
pop	3831	14.3 %	67.3 %

Frequencies of genre

Levels	Counts	% of Total	Cumulative %
rock	8747	32.7 %	100.0 %

The genre with the highest frequency is rock whereas the genre with the lowest frequency is blues.

Frequency table of *mode*.

Frequencies of mode

Levels	Counts	% of Total	Cumulative %
Minor	10082	37.7 %	37.7 %
Major	16670	62.3 %	100.0 %

The major mode category has higher frequency compared to the minor mode.

Joint frequency table of *mode* and *genre*.

Frequencies of mode

mode	genre						
	alternative	blues	hiphop	indie	metal	pop	rock
Minor	737	681	1204	1491	1275	1761	2933
Major	1423	1369	1377	2847	1770	2070	5814

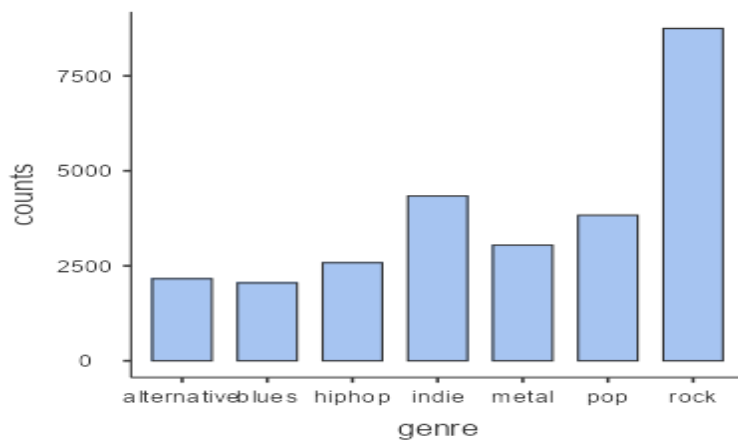
The minor mode has high frequencies across all genres. Moreover, the rock genre has highest frequencies in both modes and blues genre has the lowest frequencies in both modes.

Generally, frequency tables record the number of observation falling on each interval.

They are useful for analysing categorical data and for screening data for data entry errors

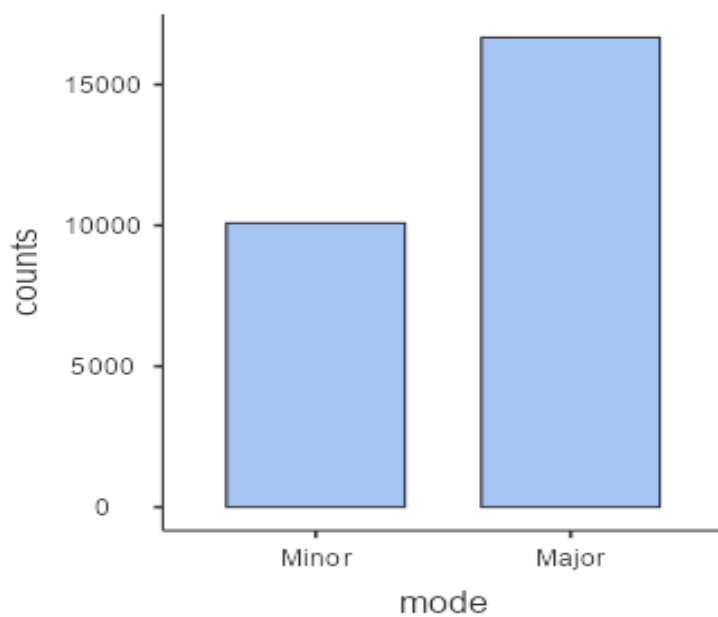
Bar charts

Bar chart for the distribution of *genre*



As from the frequency table, it is easily noticeable that the blues genre has the lowest frequency and the rock genre has the highest frequency.

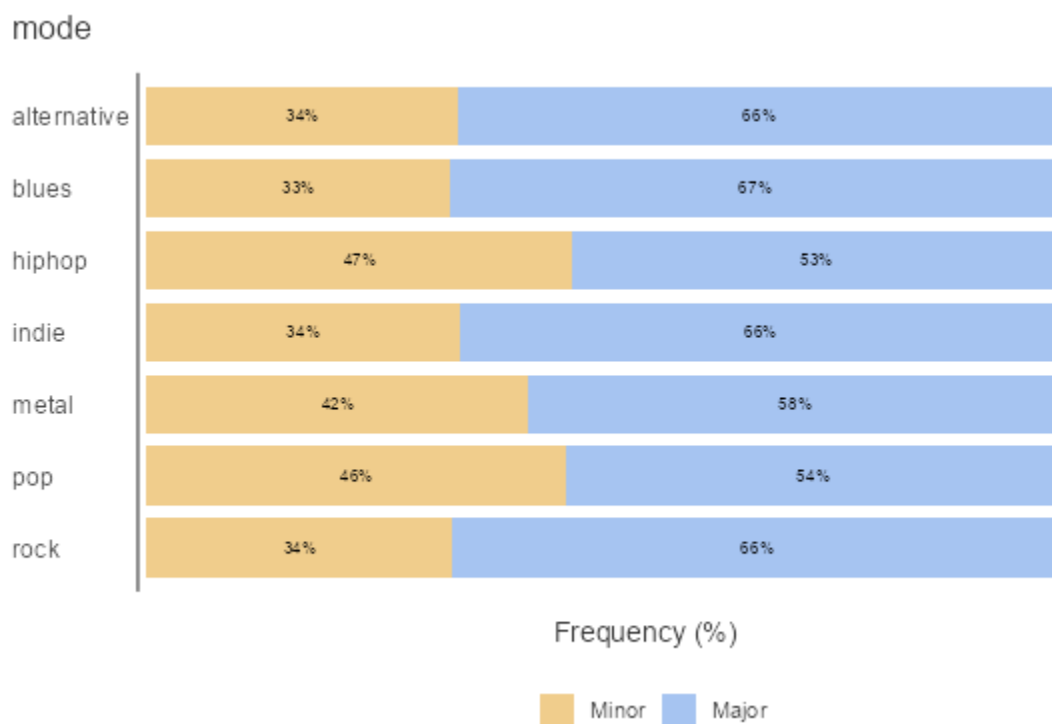
Bar chart for the distribution of *mode*



It is noticeable that the major mode has the higher frequency compared to the minor mode, as detected from the frequency table.

Bar chart for the distribution of *mode* and *genre*

One with proportions with bars stacked

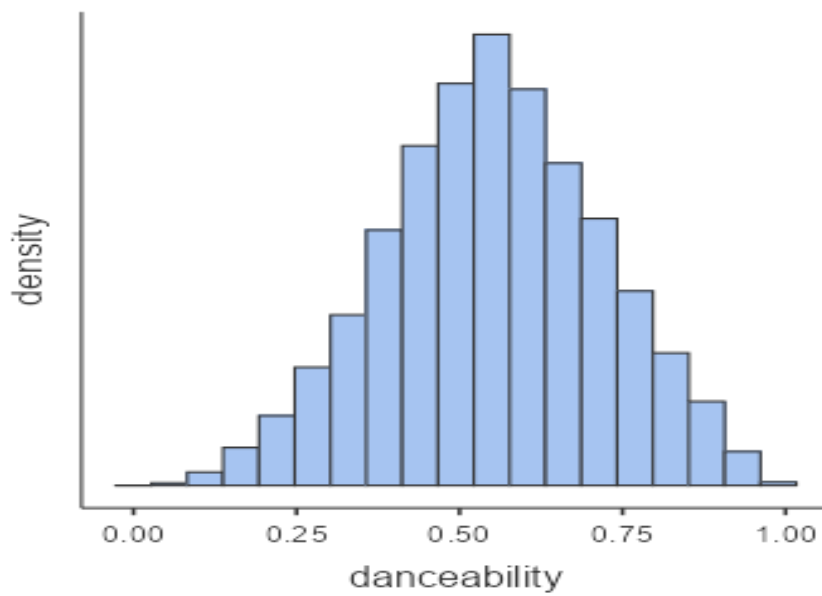


It is noticeable that the minor mode has high frequencies across all genres, the rock genre has highest frequencies in both modes and blues genre has the lowest frequencies in both modes.

One with proportions with bar stacked is easier to analyze and interpret because it displays the proportion of both modes for each genre in one column.

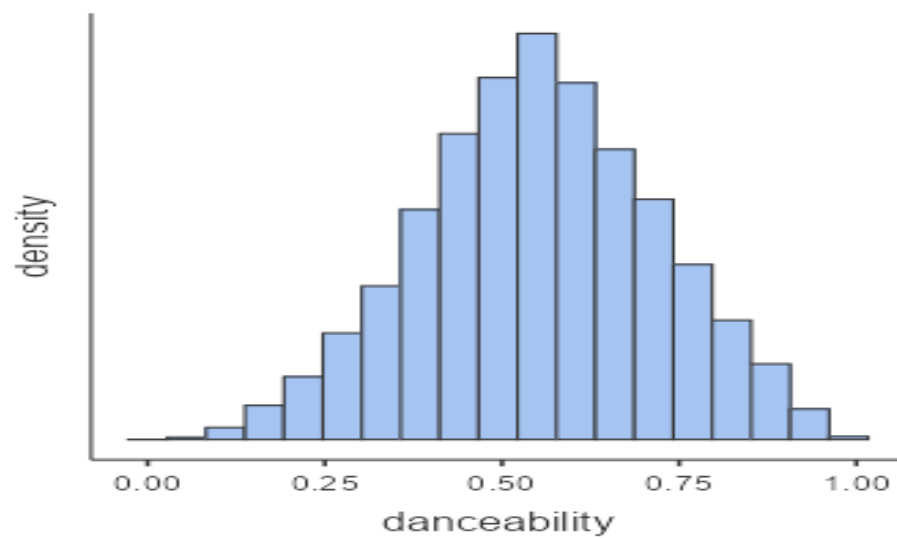
Histograms

Histogram for *danceability*



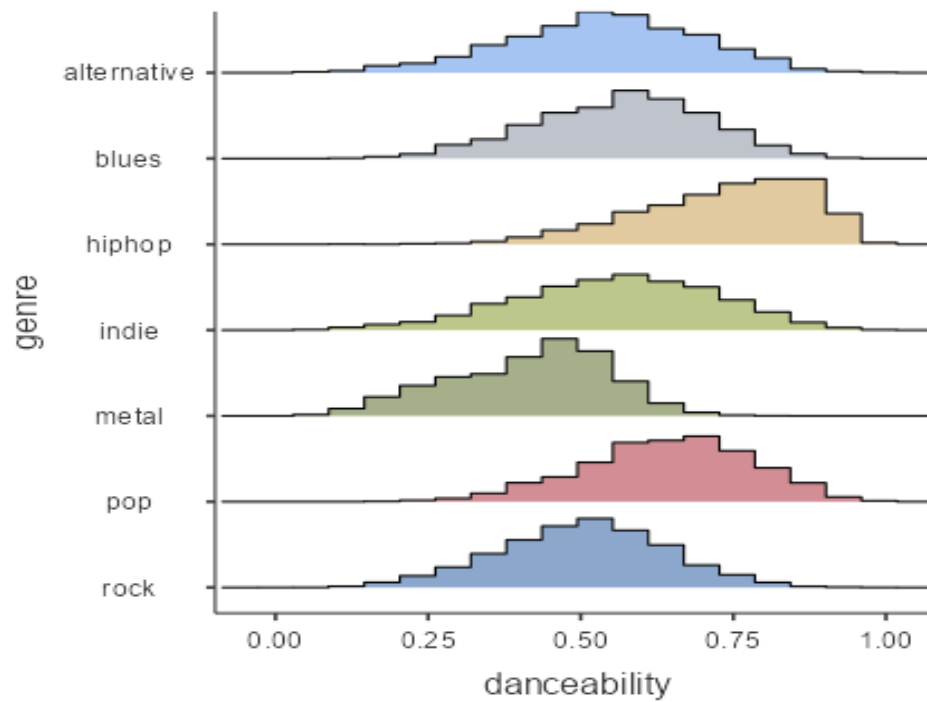
The densities are normally distributed

Histogram for *acousticness*.

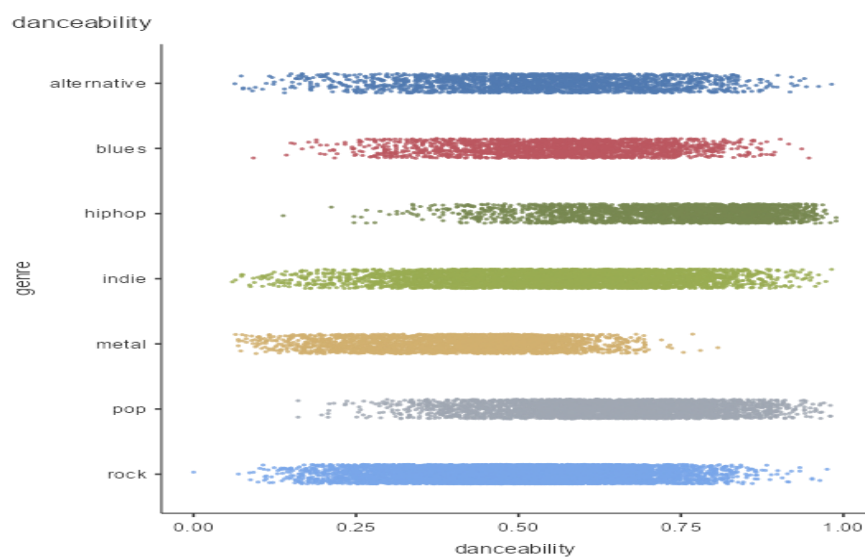


The densities are normally distributed

Histogram for *danceability* grouped by *Genre*.



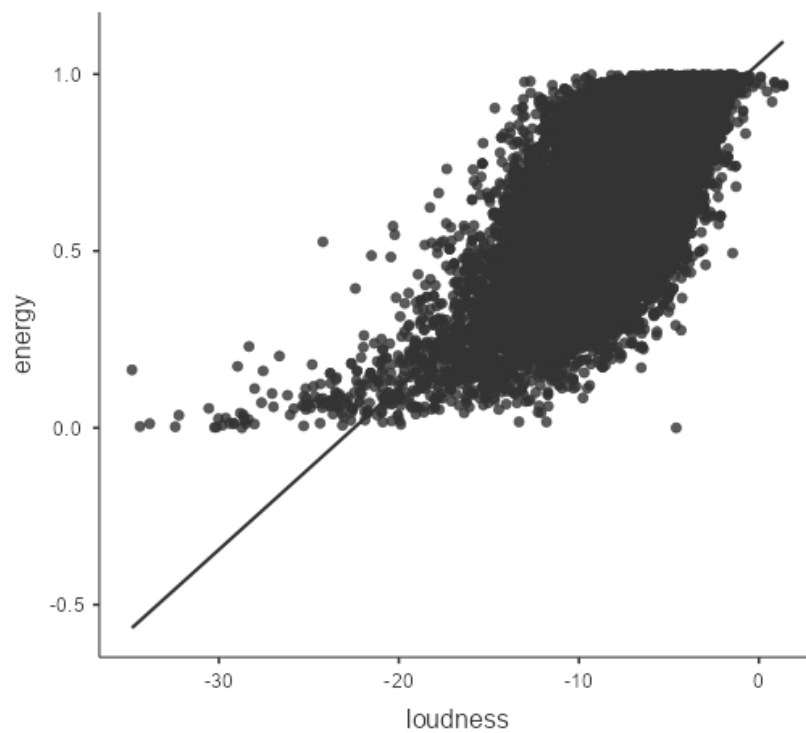
Jittered point plot for *danceability* grouped by *Genre*.



The jittered point plot shows the differences between genres better. It provides better visualization of the density of the data and the relationship between the variables. It also helps in finding clusters in the data.

Student's choice

Scatter plot of energy against loudness.



The above plotting is a scatter plot of energy against loudness. It can be observed that there is a linear relationship between the two variables. I included a linear regression line for prediction of energy when the loudness is known.