

Lab 1

In this lab, we'll get a short introduction to doing some basic data exploration in using the software jamovi. You'll need:

- A computer to install the software on OR
- An internet connected device to run the software through a browser.

Getting Set Up

If you're on a computer/chromebook, you can download jamovi for free.

The manual has some links and explanations on how to download and install jamovi on your device: <https://www.jamovi.org/user-manual.html#installation>

If you'd rather use the web-browser version, here is the link: <https://cloud.jamovi.org/>

(NOTE: the web version is a demo, and so it is still being worked on. This means it sometimes experiences downtime or other weird bugs.)

Getting the Data

In this lab, we'll be looking at the Spotify data set. Download the file `spotify.csv` from Canvas.

Open the data set in jamovi, and we'll be ready to begin! Attached to this lab is some data documentation to explain what each column represents and how to read the measurement.

Tasks

Data Cleanup

- Once you've loaded the data into jamovi, look at it for a bit. In the Variables pane, click Edit. Investigate each variable's measure type.
- Note that `duration_ms` is measured in milliseconds, which can be difficult to interpret. Let's practice converting this to minutes. We'll create a new variable.
 - Under the Variables pane, we should see the option at the top to Add a variable, with a dropdown menu. Select "Computed Variable."
 - Edit the new variable. Name it `duration_mins`, and use the formula `duration_ms / 60000` (since there are 60,000 milliseconds in a minute).
 - Check your new variable in the Data pane: the values should be reasonable, and you can check by hand to make sure they are correct.
- The `mode` column is a qualitative variable that categorically notes whether a song is written in a Major key or Minor key. It currently uses 1's and 0's to represent that, but let's change it.
 - Under the Variables pane, find and edit `mode`.
 - On the right, there is a box with "Levels" that currently has 0 and 1 only. Edit those, so that 0 is "Minor" and 1 is "Major"

- Check your `mode` variable in the Data pane: it should be filled with "Major" and "Minor" instead of 0's and 1's.

In Your Writeup

Pick a variable that had a continuous measurement. Explain why it was continuous.

Pick a variable that had a nominal measurement. Explain why it was nominal.

None of the variables had an ordinal measurement by default. Are there any that you think *should* be ordinal? Which ones? Why? If you don't think any should be, explain why.

Frequency Tables

- Create a frequency table of `Genres`.
- Create a frequency table of `mode`.
- Create a joint frequency table of `mode` and `Genres` (`mode` grouped by `Genre`)

In Your Writeup

Include each table (you should be able to copy and paste into your favorite text editor). For each table, point out any important observations, and describe the general characteristics of the frequency tables.

What kind of information does this tell us about how genres and mode are spread across our data?

Bar Charts

- Create a bar chart for the distribution of `Genre`.
- Create a bar chart for the distribution of `mode`.
- Create bar charts for the distribution of `mode` grouped by `Genre`
 - create one with the frequencies/counts visualized, and one with proportions
 - create one with the bars grouped and one with the bars stacked

In Your Writeup

Include the bar charts for `Genre` and `mode` in your writeup (you can copy and paste them as pictures). For the bar chart of `Genre` and `mode`, compare the visual to the table you produced.

For the `mode` grouped by `Genre` charts, pick the one that you think illustrates the differences in modality by genre the clearest and include it. Given a description or summary of the chart, as well as reasons for inclusion: why did you think this one best illustrates modality by genre?

Histograms

- Create a histogram for `danceability`.
- Create a histogram for `acousticness`.
- Create a histogram for `danceability` grouped by `Genre`.
- Create a jittered point plot for `danceability` grouped by `Genre`.

In Your Writeup

Include your histograms for `danceability` and `acousticness`. Compare the two. What kinds of similarities or differences do you see, and how could we interpret those differences in terms of our data?

Include the histogram and jittered plot point for `danceability` grouped by `Genre`. Which of these two visuals do you think shows the differences between genre's better (in terms of danceability). Explain.

Student's Choice

- Investigate some different types of plots for different variables in this data set.
- Try to do some by groups, and some without.
- You can look at histograms, bar charts, pareto charts, scatterplots, and any sorts of options along the way, but it might be best to stay under the "Exploration" menu in the Analyses pane.

In Your Writeup

Pick one of the plots you looked at in this section, and include it. Explain what it is plotting (in specifics), what information we are able to see, and why you chose this plot.

Writeup

Your writeup should essentially be a small report or summary of the activities in this lab. You can organize it with a short introduction and headings for each task. Include the items and responses from the "In Your Writeup" parts in each section. Make it readable and professional. Submit it as a physical (printed) copy in class with your name on it.

Column Name	Description
Artist Name	Name of the artist
Track Name	Name of the track
Popularity	The popularity of the track. The value will be between 0 and 100, with 100 being the most popular. The popularity is calculated by algorithm and is based, in the most part, on the total number of plays the track has had and how recent those plays are.
Sub-Genres	A list of the sub-genres the artist is associated with. If not yet classified, the list is empty.
Playlist	The name of the most popular playlist that the song is featured in.
danceability	Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.
energy	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.
key	The key the track is in. Integers map to pitches using standard Pitch Class notation . E.g. 0 = C, 1 = C#/D♭, 2 = D, and so on.
loudness	The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typical range between -60 and 0 db.
mode	Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0.
speechiness	Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.
acousticness	A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.

Column Name	Description
<code>instrumentalness</code>	Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal". The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.
<code>liveness</code>	Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.
<code>valence</code>	A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).
<code>tempo</code>	The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.
<code>id</code>	The Spotify ID for the track.
<code>track_href</code>	A link to the Web API endpoint providing full details of the track.
<code>analysis_url</code>	A URL to access the full audio analysis of this track. An access token is required to access this data.
<code>duration_ms</code>	The duration of the track in milliseconds.
<code>time_signature</code>	An estimated overall time signature of a track. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure).