

## 7.12 Music

*Source:* Collected by Dianne Cook.

*Number of cases:* 62

*Number of variables:* 7

*Description:* Using an Apple computer, each track was read into the music editing software Amadeus II, and the first 40-second clip was snipped and saved as a WAV file. (WAV is an audio format developed by Microsoft®, commonly used on Windows but becoming less popular.) These files were read into R using the package `tuneR` (Ligges 2006), which converts the audio file into numeric data. All of the CDs contained left and right channels, and variables were calculated on both channels.

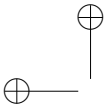
Variable	Explanation
artist	Abba, Beatles, Eels, Vivaldi, Mozart, Beethoven, Enya
type	rock, classical, or new wave
lvar, lave, lmax	average, variance, maximum of the frequencies of the left channel
lfener	an indicator of the amplitude or loudness of the sound
lfreq	median of the location of the 15 highest peak in the periodogram

*Primary question:* Can we distinguish between rock and classical tracks? Can we group the tracks into a small number of clusters according to their similarity on audio characteristics?

*Data restructuring:* This dataset is very clean and simplified. The original data contained 72 variables, most of which have been excluded.

*Analysis notes:* Answers to the primary question might be used to arrange tracks on a digital music player or to make recommendations. Other questions of interest might be:

- Do the rock tracks have different characteristics than classical tracks?
- How does Enya compare with rock and classical tracks?
- Are there differences between the tracks of different artists?



*Data files:*

<code>music-sub.csv</code> , <code>music-sub.xml</code>	Subset of data used in this book. The last five tracks in the data (58–62) have the artist and type of music loosely disguised so that they can be used to test classifiers that students built using the rest of the data.
<code>music-all.csv</code> , <code>music-all.xml</code>	Full datasets, 72 variables, and a few missing values.
<code>music-clust.csv</code> , <code>music-clust.xml</code>	Subset of data, augmented with results from different cluster analyses
<code>music-SOM1.xml</code> , <code>music-SOM2.xml</code>	Different SOM models appended to the data.

### 7.13 Cluster Challenge

*Source:* Simulated by Dianne Cook.

*Number of cases:* 250

*Number of variables:* 5

*Description:* Simulated data included as a challenge to find the number of clusters.

*Primary question:* How many clusters in this data?

*Data files:*

`clusters-unknown.csv`

### 7.14 Adjacent Transposition Graph

*Source:* Constructed by Deborah F. Swayne.

*Number of cases:* 24 nodes and 36 edges in the  $n = 4$  adjacent transposition graph; 120 nodes and 240 edges in the  $n = 5$  graph.

*Number of variables:* 3 variables in the  $n = 4$  graph; 4 variables in the  $n = 5$  graph.

*Description:* The  $n = N$  adjacent transposition graph is generated as follows. Start with all permutations of the sequence 1, 2, ...,  $N$ . There are  $N!$  such sequences; make each one a vertex in the graph. Connect two vertices by