



An appropriate label for your cases should be chosen carefully. In our iTunes example, a natural choice of a label would be the name of the song. However, if you have more than one artist performing the same song, or the same artist performing the same song on different albums, then the name of the song would not uniquely label each of the songs in your playlist.



A quantitative variable such as the time in the iTunes playlist requires some special attention before we can do arithmetic with its values. The first song in the playlist has time equal to 3:29, that is 3 minutes and 29 seconds. To do arithmetic with this variable, we should first convert all of the values so that they have a single unit. We could convert to seconds; 3 minutes is 180 seconds, so the total time is $180 + 29$ or 209 seconds. An alternative would be to convert to minutes; 29 seconds is .483 minutes, so time in this way is 3.483 minutes.

USE YOUR KNOWLEDGE

1.1 Time in the iTunes playlist. In the iTunes playlist, do you prefer to convert the time to seconds or minutes? Give a reason for your answer.

In practice, any set of data is accompanied by background information that helps us understand the data. When you plan a statistical study or explore data from someone else's work, ask yourself the following questions:

1. **Who?** What **cases** do the data describe? **How many** cases appear in the data?
2. **What?** How many **variables** do the data contain? What are the **exact definitions** of these variables? In what **unit of measurement** is each variable recorded?
3. **Why? What purpose** do the data have? Do we hope to answer some specific questions? Do we want to draw conclusions about cases other than the ones we actually have data for? Are the variables that are recorded suitable for the intended purpose?



EXAMPLE

1.3 Data for students in a statistics class. Figure 1.2 shows part of a data set for students enrolled in an introductory statistics class. Each row gives the data on one student. The values for the different variables are in the columns. This data set has eight variables. ID is an identifier for each student. Exam1, Exam2, Homework, Final, and Project give the points earned, out of a total of 100 possible, for each of these course requirements. Final grades are based on a possible 200 points for each exam and the final, 300 points for Homework, and 100 points for Project. TotalPoints is the variable that gives the composite score. It is computed by adding 2 times Exam1, Exam2, and Final, 3 times Homework plus 1 times Project. Grade is the grade earned in the course. This instructor used cut-offs of 900, 800, 700, etc. for the letter grades.