Using Tables and Figures to Present Data

Suppose you are studying the growth of a magic beanstalk as function of time and that you record the following data in your laboratory notebook:

August 28, 2004

A bean from lot AG12, identified as AG12a, was planted at 10:05 AM and promptly watered and fertilized with a solution of Alice's Miracle Grow. The beanstalk's height was measured using a Jack & Son Model X15 laser-based tape measure, with the height taken as the distance between the ground and the point where the topmost leaf is attached to the beanstalk.

- 10:15 am: the beanstalk emerged from the ground
- 10:30 am: height is 9.96 m
- 10:45 am: height is 107 m
- 11:00 am: height is 329 m
- 11:25 am: height is 763 m
- 12:05 pm: height is 1.21 km
- 12:35 pm: height is 1.32 km
- 1:00 pm: height is 1.34 km
- 2:00 pm: height is 1.34 km

When you prepare a report of this study, how should you present your data? Although it is tempting to transfer directly into your report what is in your notebook, this rarely is a good idea as what is of interest is not the raw data itself, but the information you extract from that data. For example, the raw data gives the time when you made a measurement, but what is of interest is how much time elapsed since you planted the bean.

Using a Table to Present Your Data

One way to present this data is in a table such as the one here:

Table 1: Results for Bean AG12a

elapsed time (min)[a]	height of beanstalk (km)[b]
0	0.00
10	0.00
25	0.01
40	0.11
55	0.33
80	0.76
120	1.21
150	1.32
175	1.34
235	1.34

[a] From time planted.
[b] From ground to topmost leaf.

Note that the table is numbered and includes an informative title (which is positioned above the table), that each column begins with a descriptive entry and includes units (when appropriate), that the data in each column are aligned using the decimal point and that footnotes are included below the table to supply additional information. Note, as well, that the less useful measurement of absolute real time is converted into an elapsed time and that the heights are converted to a common unit. This table contains all the information recorded in the lab notebook in a format that makes it easy for the reader to see the relationship between time and the height of the beanstalk.

Although not appropriate for the data in Table 1, it often is useful to include a statistical summary of your data. Table 2, for example, reports the maximum height of beanstalks grown using five beans from the same lot, along with their mean and standard deviation.

Table 2: Maximum Heights of Beanstalks

Sample ID[a]	Maximum Height (km)[b]
AG12a	1.34
AG12b	1.24
AG12c	1.48
AG12d	1.39
AG12f	1.31
mean	1.35
std dev	8.89×10^{-2}

[a] Sample AG12e did not germinate and is not included in this summary.

Using a Figure to Present Your Data

Another way to present the data in Table 1 is as a figure that shows height as a function of elapsed time (see Figure 1). Note that the figure is numbered and includes an informative title (positioned below the figure), that the data points are easy to see, that the scales for the axes spread the data over the figure's available space, and that the axes have descriptive titles (including units where appropriate).

One advantage of a figure is that it allows you to highlight your analysis of the data. For example, Figure 1 shows the result of modeling the data using the following equation

$$\text{height} = A \times e^{-(\text{elapsed time}-B)^2/C^2} + D \tag{1}$$

where A, B, C, and D, are variables used to fit the model to the data. A figure also is a useful way to display several related sets of data. For example, Figure 2 shows results for three trials. Note that the figure includes a legend that identifies each set of data. This figure also connects the points with line segments to further highlight the relative trend in each data set. Note, as well, that this figure includes a grid to aid in reading the graph. Connecting points with line segments and including grid lines are not common choices but they are effective in some situations; use them judiciously.

Figures 1 and 2 are examples of good figures. In Figure 1, for example, the data points are easy to identify because the plotting symbols are of sufficient size; this, in turn, makes it easy to evaluate how well Equation 1 fits the data (note, for example, that the model suggests that the height first decreases to a negative value before it increases, a prediction that makes no sense). In both figures the data occupy the available space, which makes it easy to appreciate the data's trend. For example, in Figure 2 we see that all three growth curves are sigmoidal in shape with a similar lag time between the planting of the seed and the emergence of the bean stalk. We also see that there is some variation between the three beans with respect to the growth rates and the maximum height of the mature plant.

A poorly prepared figure hides important information. Consider, for example, Figure 3, which shows the temperature of a sample of porridge as it cools. Including the thermometer's full scale on the y-axis (0°C to 240°C) compresses the data to a small portion of the y-axis, which makes it difficult to visualize how the temperature changes with time. Furthermore, the plotting symbols are so small that it is difficult to see the

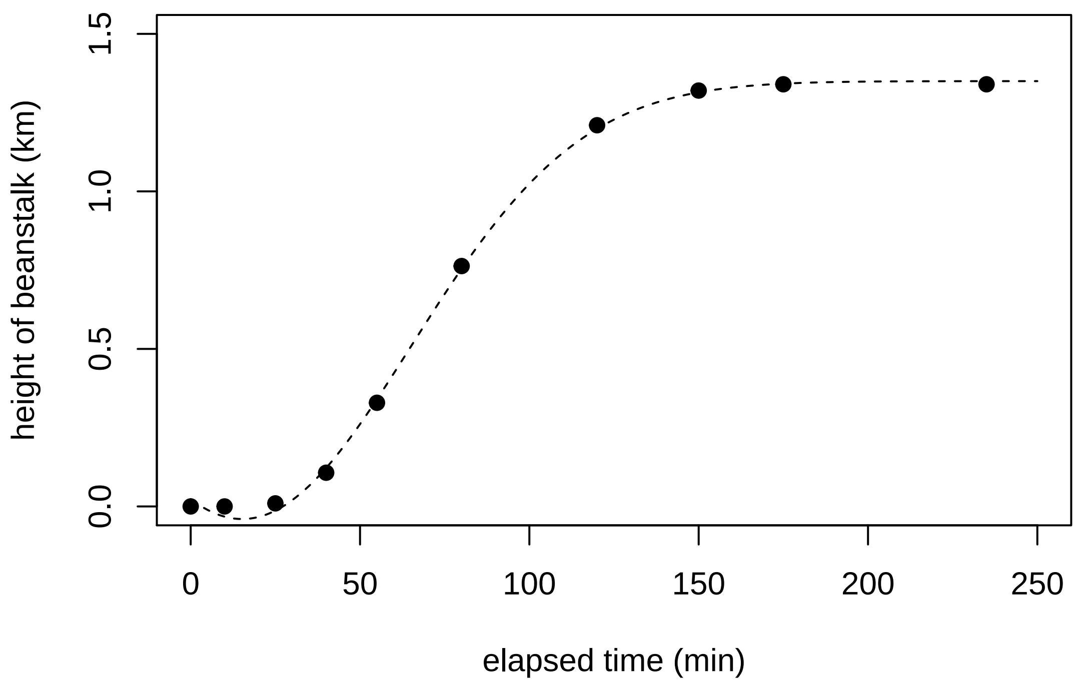

Figure 1: Height of a beanstalk from seed AG12a as a function of time after planting. The individual data points are drawn from Table 1; the dashed line shows the result of fitting Equation 1 to the data where A is -1.39, B is 15.2, C is 70.4, and D is 1.35.

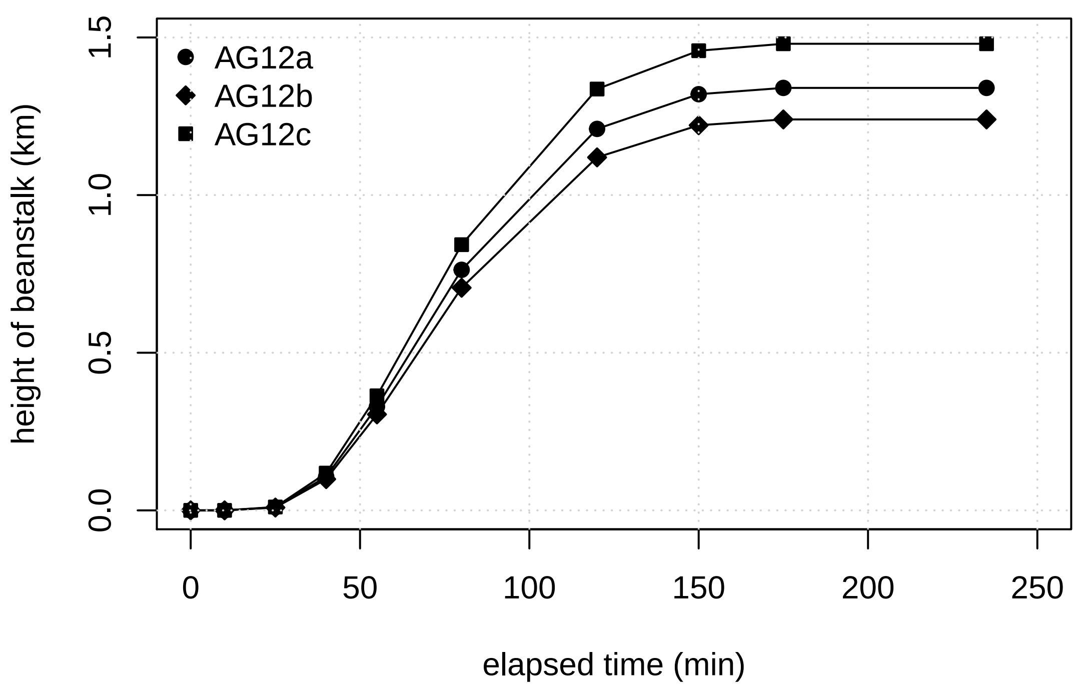

Figure 2: Heights of a beanstalk from three seeds selected at random from lot AG12 as a function of time after planting.

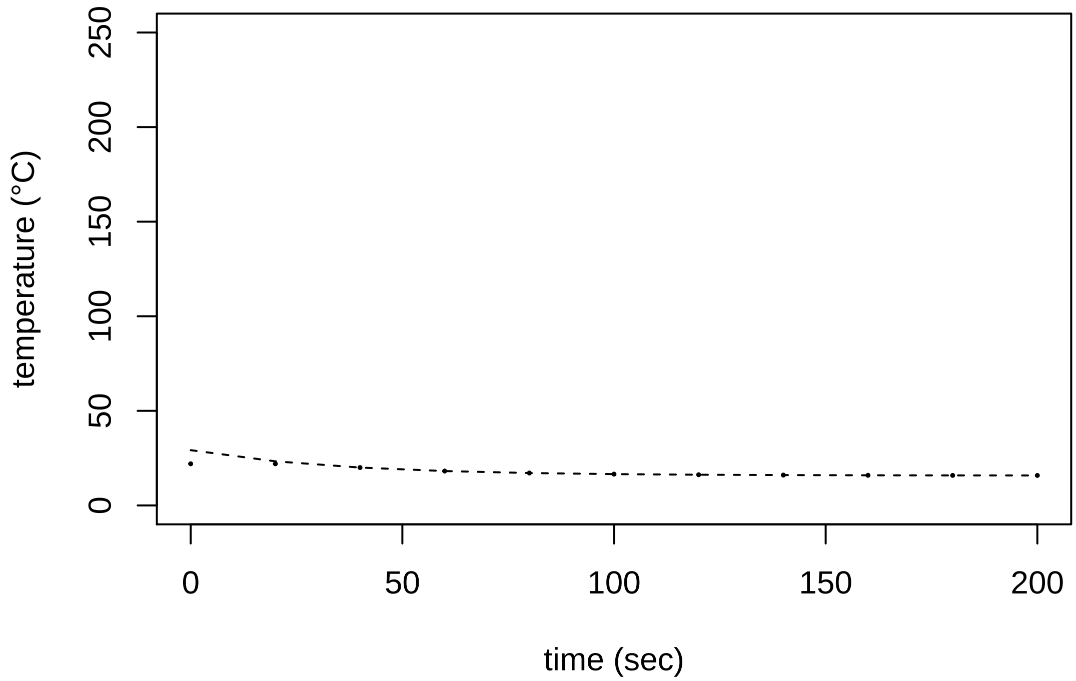

Figure 3: Cooling curve for Baby Bear's porridge.

individual data points. Finally, the caption does not suggest why the first part of the data was not included when modeling the data.

A better figure is shown in Figure 4. Note that in this figure the scale on the y-axis spans the relevant temperatures only (the scale for an axis does not need to begin at 0), which makes it easy to see the cooling curve. The greater size of the plotting symbols makes the data easier to see and improves the reader's ability to evaluate the quality of the fit between the data and the theoretical model. Note, as well, that the figure's caption clarifies why the first 25 seconds of data were not included when modeling the data and that the theoretical model is based on Newton's Law of Cooling. This figure is one that provides the reader with a useful summary of the data.

When to Use a Table and When to Use a Figure

So, which is the best method for presenting data: a table or a figure? If you wish to show a trend or to demonstrate that a theoretical model explains your data, then a figure usually is the best choice. On the other hand, when exact numerical values or a statistical summary are important, then a table is the better choice. In general, for any data set, you should use a figure or a table, but not both; however, let the needs of your argument guide your decision.

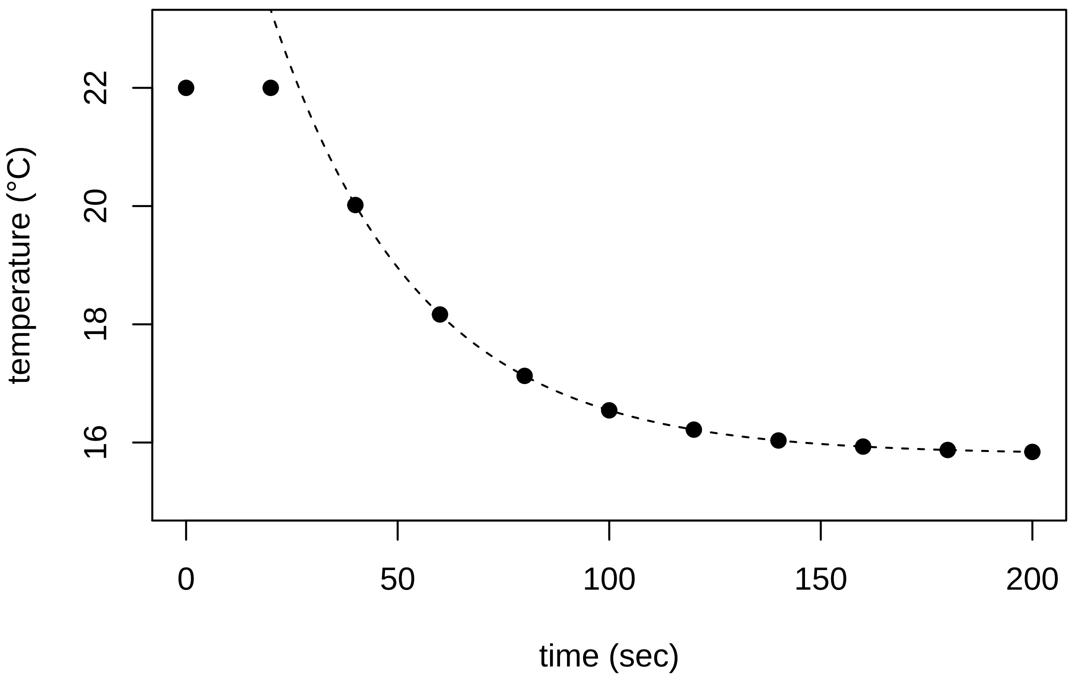

Figure 4: Cooling curve for Baby Bear's porridge. The sample was removed from the stove at $t = 25$ sec. The dashed line shows the modeling of cooling using Newton's Law of Cooling for times greater than 25 sec.