CHAPTER 9

Data Presentation

INTRODUCTION

Probably the most critical stage in a project is when data are collected and are ready to organize and interpret. The final step—and the one that can have the most impact—is presentation of the data in a form that clearly shows functions, physical properties or dependencies. Often, scientists and engineers spend great amounts of time and energy designing and executing tests and experiments. They often fail, however, to apply the same spirit when tabulating the data into a form that can readily explore and effectively communicate inter-relationships and anomalies. Unexpected or “rogue” data points can signal an experimental or recording error, or they can herald an effect worthy of additional study. In addition, the data preparation and presentation stage often lacks sufficient exploration of the various possible displays that can make results clear to an audience.

Today, wonderful computer packages easily tabulate and display data. However, the great number of options and ease-of-use can create the illusion that great looking and meaningful tables and displays can be produced with little effort. Invariably, there remains a need to explore data and make wise, analytical choices concerning the types of presentations best suited to demonstrate important features.

A medical doctor performs tests on a patient, integrates the information and finally develops a diagnosis. Likewise, an engineer designs an experiment, records data, and finally tabulates and interprets the information before presenting it.

Therefore, just as a medical doctor works towards a diagnosis as a clear goal, an engineer must keep his/her goals in mind as s/he enters the data analysis and presentation phase of a project. In some ways, an engineer’s or scientist’s task can be made more difficult because the function or process represented by the data may not be understood. Papers are often published that clearly present data showing a complex, new relationship, but without an explanation. Such papers can be seminal in a field and lead to a major breakthrough. In other words, the creation and presentation of a data set can represent an important accomplishment in its own right.

In a series of books [1][2][3], E.R. Tufte addresses the visual display of quantitative information. His definition of the ideal data presentation is represented by the following quote [1]: “Excellence in statistical graphics consists of complex ideas communicated with clarity, precision and efficiency.” Tufte’s goals of graphic displays (slightly modified) are:

- Present the data.
- Get the viewer to think about substance rather than methods or graphic design.
- Do not distort the message of the data.
- Encourage the eye to compare different pieces of data.
- Reveal levels of detail from a broad overview to a fine structure.
- Have a clear purpose (e.g., description, comparison, exploration).
- Integrate with other descriptions of the data set.

The following sections present data in tabular form and explore methods of display, illustrating the strengths and weaknesses of different presentation formats.

DATA SHOWING HOW A QUANTITY IS DISTRIBUTED

To show how a resource is divided, a pie chart presentation is very effective. At a glance, the relative sizes of the segments provide the approximate distribution, and accompanying legends or tables provide precise numerical data, if required. For instance, a series of pie charts showing consecutive yearly distributions may indicate trend information. As an example, assume that a design class has time allotted to various subjects, as shown in Table 9.1.

Table 9.1. Distribution of course time by topic.

Topic	**Time Allotted (in minutes)**	**Percentage**
Mystery Artifact Challenge	330	8.4%
Design Loop	330	8.4%
Major Design Project	1760	44.6%
Social Styles & Team Dynamics	220	5.6%
Guest Lectures	250	6.3%
Design Lectures	150	3.8%
Other (e.g., Technical Writing)	310	7.8%
Reverse Engineering	600	15.2%
Total	**3950**	**100%**

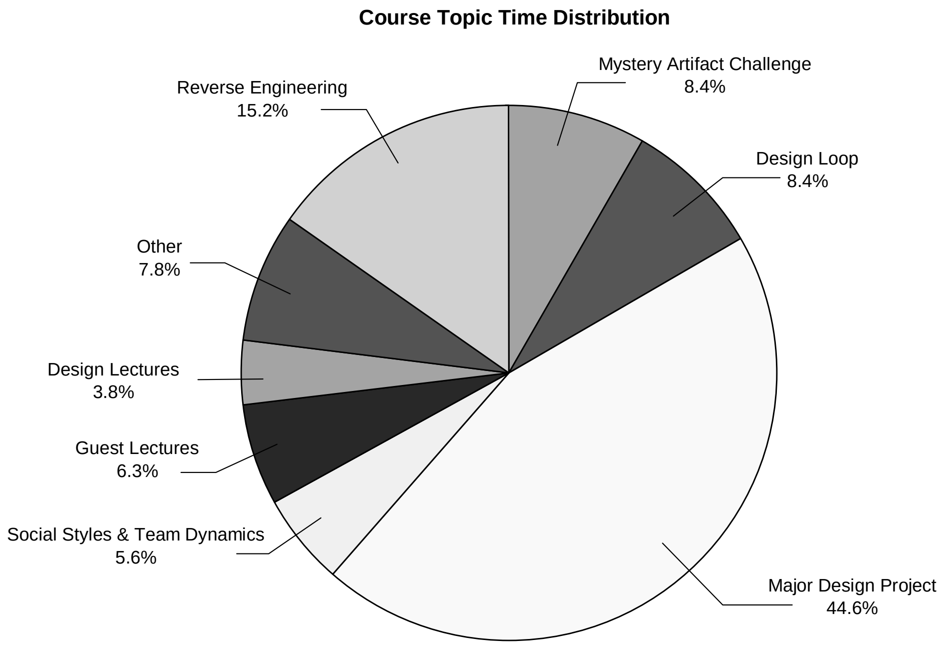

Figure 9.1. Pie chart of course topic distribution.

In Figure 9.1, the distribution by percent of time is clearly shown in a pie chart plot. At a glance, a plot like this provides guidance on how to apportion time among course topics.

COMPARISONS OF VALUES FOR DIFFERENT CATEGORIES

An article in the *NY Times* [4] compares pairs of companies that had the same ratings from a financial viewpoint, but were rated differently on a scale of environmental performance. Table 9.2, adapted from the *NY Times* article and covering five industries, summarizes the performance of “green” and ”not green” companies. It is evident from the table that the green companies showed a better return. Perhaps this is because engineers who are challenged to design products that are environmentally friendly are also challenged to think in a global way about safety, manufacturing, cost and other issues that impact efficiency and profitability.

Table 9.2. Green Begets Green: percent earnings 1994–97.

Industry	Company	Green	Not Green
Communications (COM)	NT	53.3%	
	AC		19.5%
Drugs (DRGS)	RER	59.2%	
	AL		33.7%
Utilities (UTIL)	EI	30.5%	
	DR		10.4%
Forest Products (FOR)	WAY	18%	
	LP		6.5%
Oil (OIL)	MOB	26%	
	UNC		17%

A bar chart display effectively presents these data. Although color can be expensive to print—relative to black and white—it can be worthwhile because of its impact on a presentation. (In this case, using green would be helpful.) Figure 9.2 and Figure 9.3 display the table information using vertically-oriented bars. There are two difficulties with the presentation of these data. The first is that two different plots are used, making a direct comparison difficult. The second is that two different vertical scales are used, creating the false impression that the "not green" companies provide greater or similar earnings.

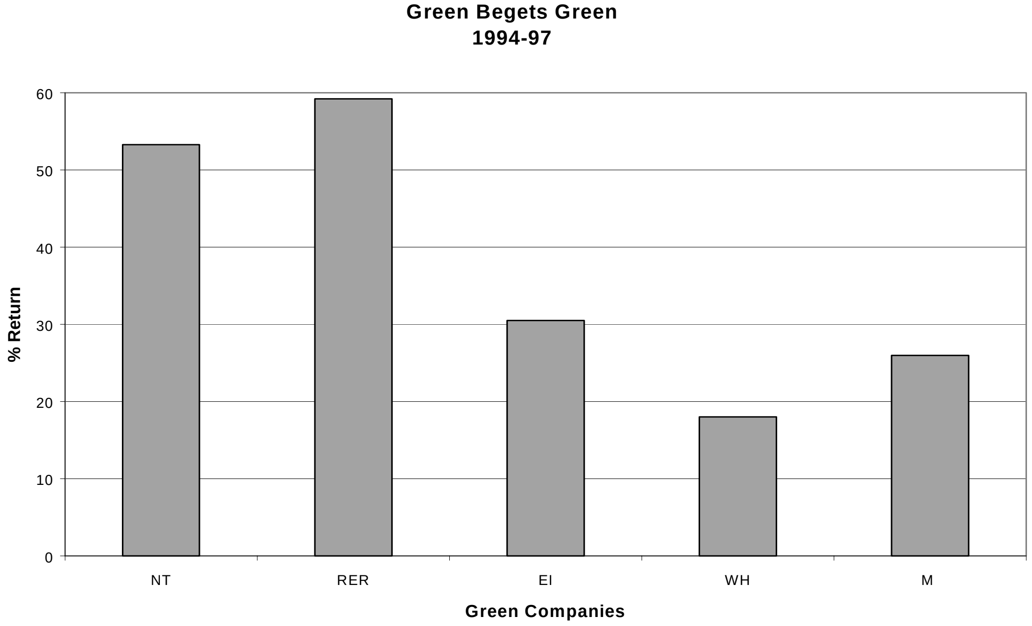

Figure 9.2. Vertical bar chart of "green" companies' financial performance.

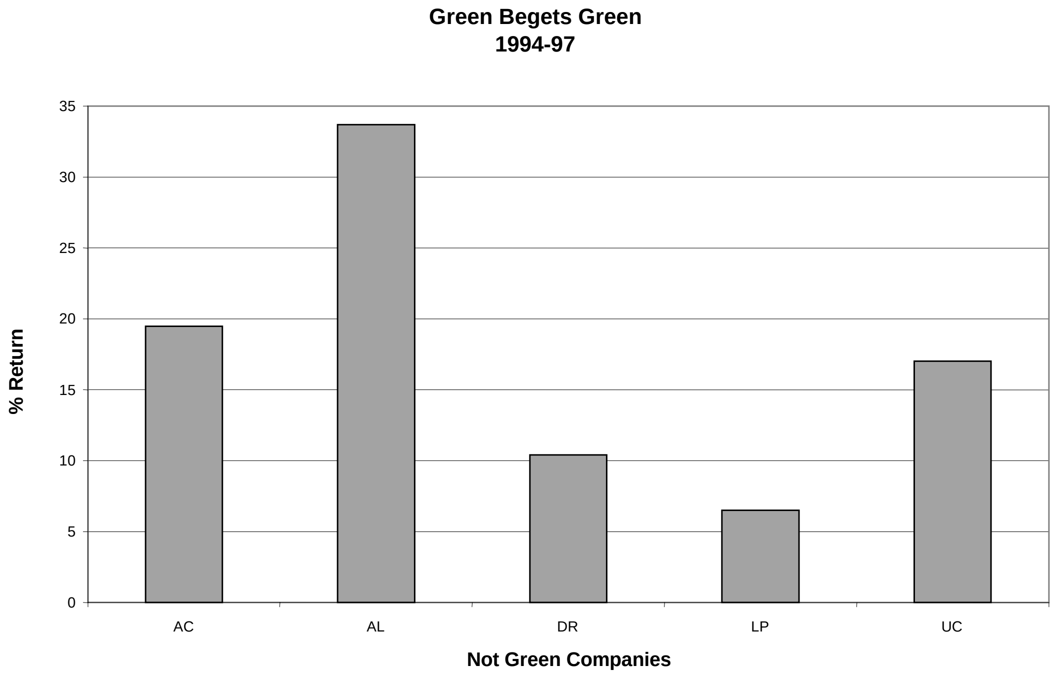

Figure 9.3. Vertical bar chart of “not green” companies’ financial performance.

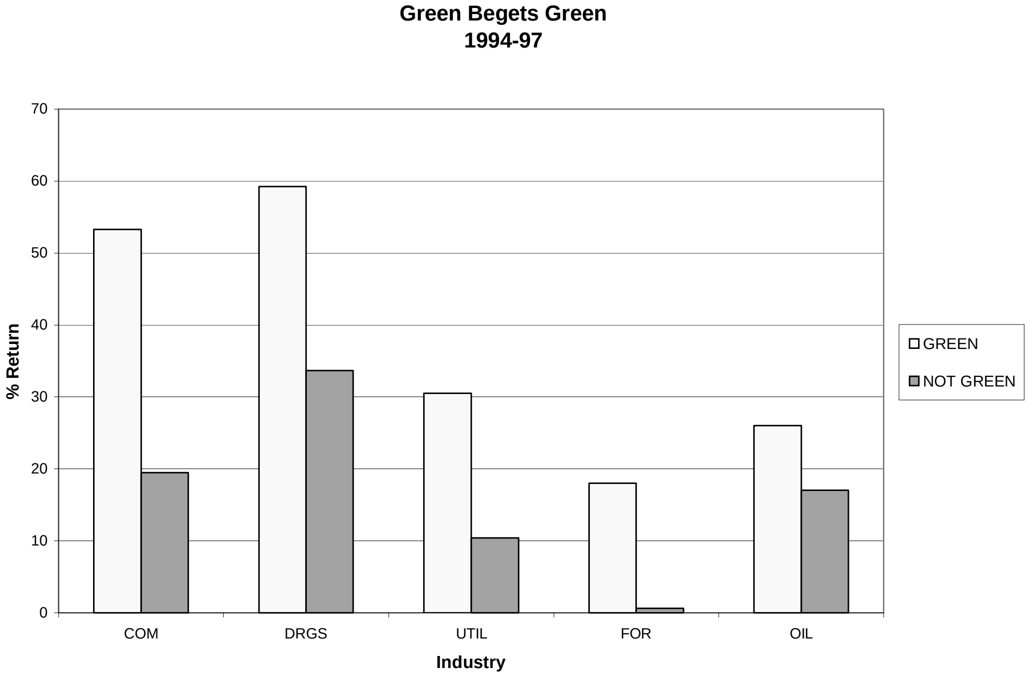

Figure 9.4. Vertical bar chart of “green” and “not green” companies’ financial performance.

In Figure 9.4, a combined plot directly compares the two classes of companies. This display has far more impact: the point that it pays to be environmentally friendly is made clear.

TWO PARAMETER DATASETS SPANNING LARGE RANGES

Table 9.3 lists data for heartbeat frequency (beats per minute) and mammal mass (in kilograms). The series of plots that follow explore these data in a number of ways.

Table 9.3. Heartbeat frequency for mammals.

Mammal	**Heartbeats per minute**	**Mass (kilograms)**
Bat	750	0.0009
Least shrew	700	0.002
Door mouse	635	0.014
Mouse	650	0.025
Hedgehog	300	0.25
Guinea pig	210	0.4
Rabbit	140	2.5
House cat	120	2.27
Dog	150	12.5
Man	72	68
Hog	75	100
Donkey	47.5	400
Horse	35	400
Elephant	37	3,636
Whale	17.5	26,636

Note that the data ranges are quite large. There is nearly a factor of 50 in the range of heartbeat frequencies and over 10^8 in mass. The first plot shown in Figure 9.5 is a scatter plot (XY plot) with linear scales. Both data points and straight lines connecting the points are shown. This allows the reader to evaluate the data visually, while pointing out the relationship between data points. To present both measured data and model predictions, engineers commonly use this type of plot. Typically, when one variable is dependent on the other, such as y = f(x), the independent variable is plotted on the horizontal axis, and the dependent variable is plotted on the vertical.

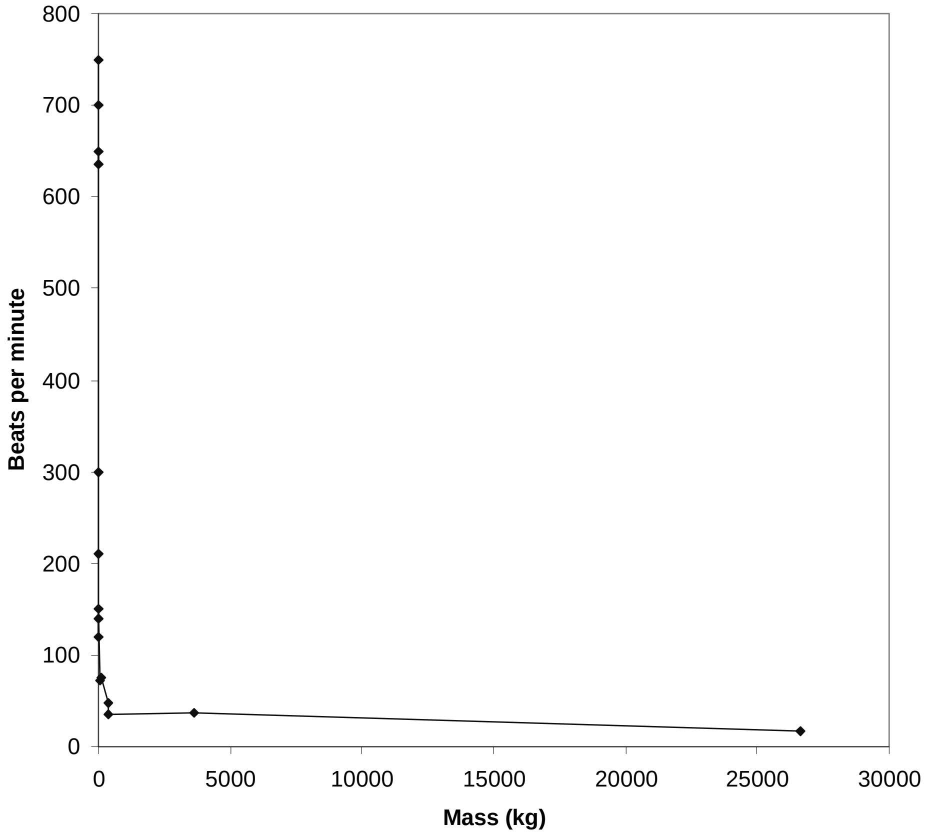

Figure 9.5. Scatter plot (using linear scales) of the heartbeat frequency of mammals.

Note that most data points in Figure 9.5 cannot be resolved visually, and only four points show their mass values clearly. This plot is certainly not a good choice for presentation or exploration of this data set. In Figure 9.6, a display using log axes is far more effective at presenting the data points so that all of the details are evident.

Further, a “best fit” model is presented in Figure 9.6, in the form of a plotted equation that approximates the data. “Best fit” means that the constants of the equation have been calculated to minimize the differences between the data and the best fit line. Such an equation is also known as a “linear regression” (whether or not it is a straight line), “least-squares fit” or “trendline.” The form of the equation must be selected first; i.e., a straight line ($y = ax + b$), an exponential ($y = ae^{bx}$), a polynomial ($y = ax^2 + bx + c$), etc. Most data analysis and spreadsheet programs can then calculate the constants. Either the form of the equation can be suggested by a mathematical model of the physics represented by the data, or the form of the data can be used to suggest a mathematical model. If there is no model, it is *not* appropriate to present a best fit line just to make the plot look nicer. In the case of the data shown in Figure 9.6, a number of researchers have noticed the power law relationship ($y = ax^b$) as a first step towards a mathematical model.

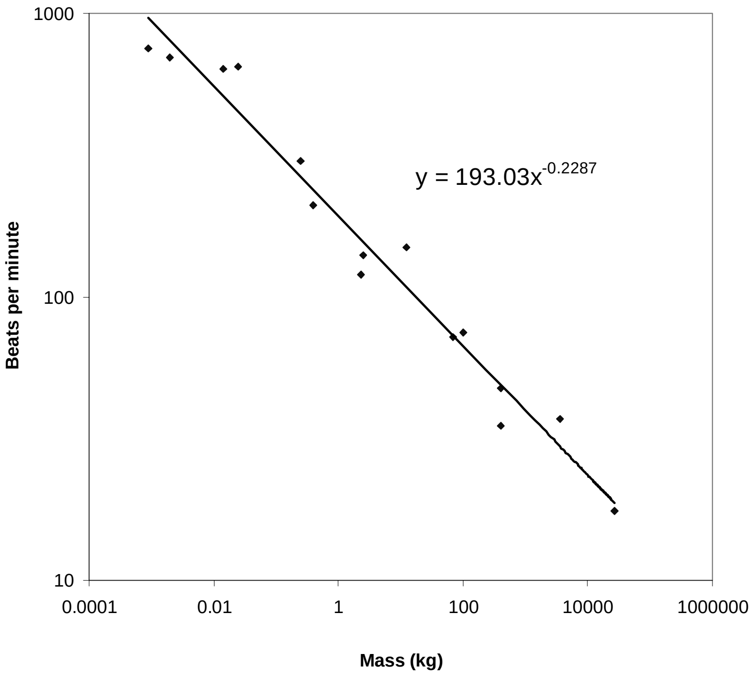

Figure 9.6. Scatter plot (using log-log axes) of the heartbeat frequency of mammals.

It may be of interest to try a semi-log plot, in which one axis uses a log scale and one axis uses a linear scale. The result is shown in Figure 9.7. This semi-log plot shows a best fit to the data for the frequency proportional to mass to the 1/4th power, which is consistent with the results from the earlier two figures. However, because of limitations in the way the best fit line is plotted by the program used to make the plot, the model appears to be a poor fit. Thus, a semi-log plot is not a good choice to best illustrate these data.

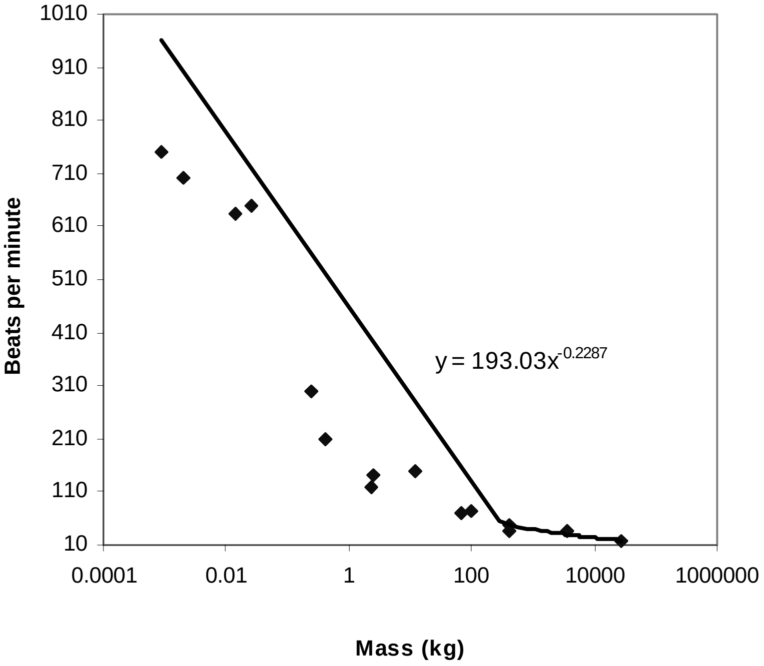

Figure 9.7. Scatter plot (using semi-log axes) of the heartbeat frequency of mammals.

COMPARE AN ANALYTICAL EXPRESSION WITH EXPERIMENTAL DATA

Often a theory exists for predicting how a process changes with time, distance or some other variable. A frequent goal is to explore the limits of a theory or determine how accurately it applies over some range. In this example, there is an accepted expression for relating the dynamic pressure produced by a flow impacting a flat surface:

$$p = \frac{1}{2} \rho V^2$$

where p is the pressure, ρ is the density and V is the velocity.

This expression is Bernoulli's equation and has been derived from fluid physics, in contrast to the previous section where the equation was dependent on the experimental data. Measurements of pressure and velocity have been made using a highly accurate thermistor probe that can measure the flow speed to better than a centimeter per second. The pressure was measured using a transducer that has some noise estimated to account for an error of plus or minus a few microbars. A wind tunnel experiment provided data over a range of wind speeds up to four meters per second. A comparison of the measured pressure/ velocity relationship with what is expected from Bernoulli's equation will be used to validate the use of Bernoulli's equation for this flow configuration. Then, Bernoulli's equation can be used to guide the design of a simple mechanical indicator of wind speed. Table 9.4 contains computed and experimental values of pressure as a function of wind speed. Incidentally, tables represent the most accurate way of presenting data.

Table 9.4. Theoretical and experimental data on pressure as a function of wind speed.

Wind Speed (cm/sec)	Experimental Dynamic Pressure (microbars)	Theoretical Dynamic Pressure (microbars)
10	0.55	0.06
25	0.04	0.0375
50	1.45	1.5
100	5.8	6
150	13.3	13.5
200	25.5	24
250	40	37.5
300	70	54
350	72.5	73.5
400	98	96

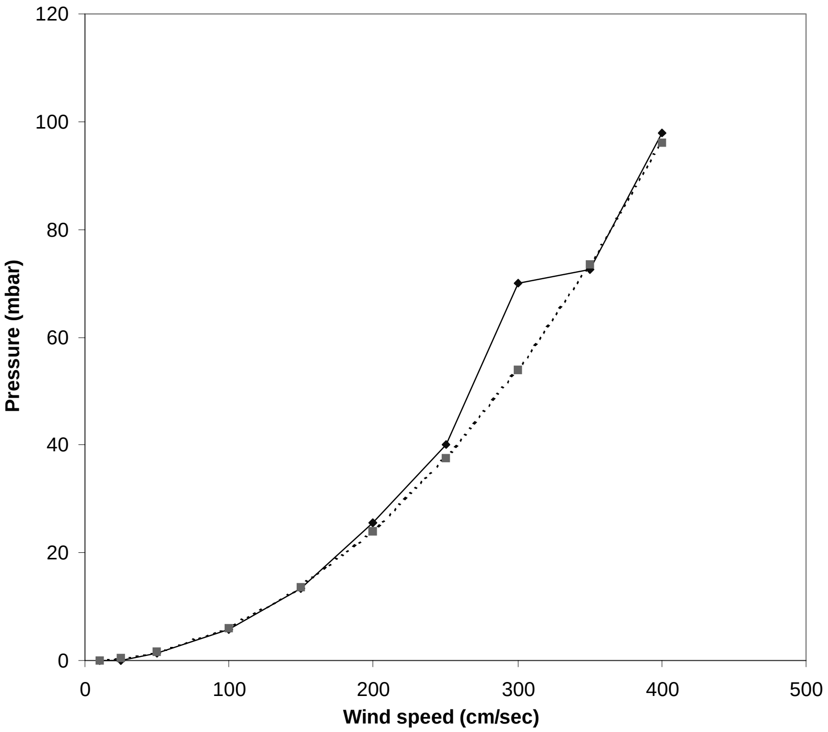

Figure 9.8. Scatter plot of predicted (dashed line) and experimental values (♦) of pressure vs. speed.

In Figure 9.8, the dashed line is the result of theory and the experimental data points are shown with error bars of plus or minus four microbars. The data point near 300-cm/sec-wind speed seems anomalous: the difference between the data and the prediction is larger than the uncertainty for which the measurements can account. In this case, the data should be checked. The deviation of this data point was not as evident from the tabular presentation. Even though the experimental data at very low wind speeds were not very reliable, the results can be used to defend extrapolating the theoretical curve with confidence. The conclusion that the equation is a good representation of the physics seems reasonable from this plot.

REFERENCES

1. Tufte, E.R., *The Visual Display of Quantitative Information,* Graphics Press, Cheshire, Connecticut, 1983, 197 pp.
2. Tufte, E.R., *Envisioning Information, Graphics Press,* Cheshire, Connecticut, 1990, 126 pp.
3. Tufte, E.R., *Visual Explanations,* Graphics Press, Cheshire, Connecticut, 1997, 156 pp.
4. Deutsch, C.H., “For Wall Street, Increasing Evidence that Green Begets Green,” *NY Sunday Times*, Financial Section, July 19, 1998, p 7.