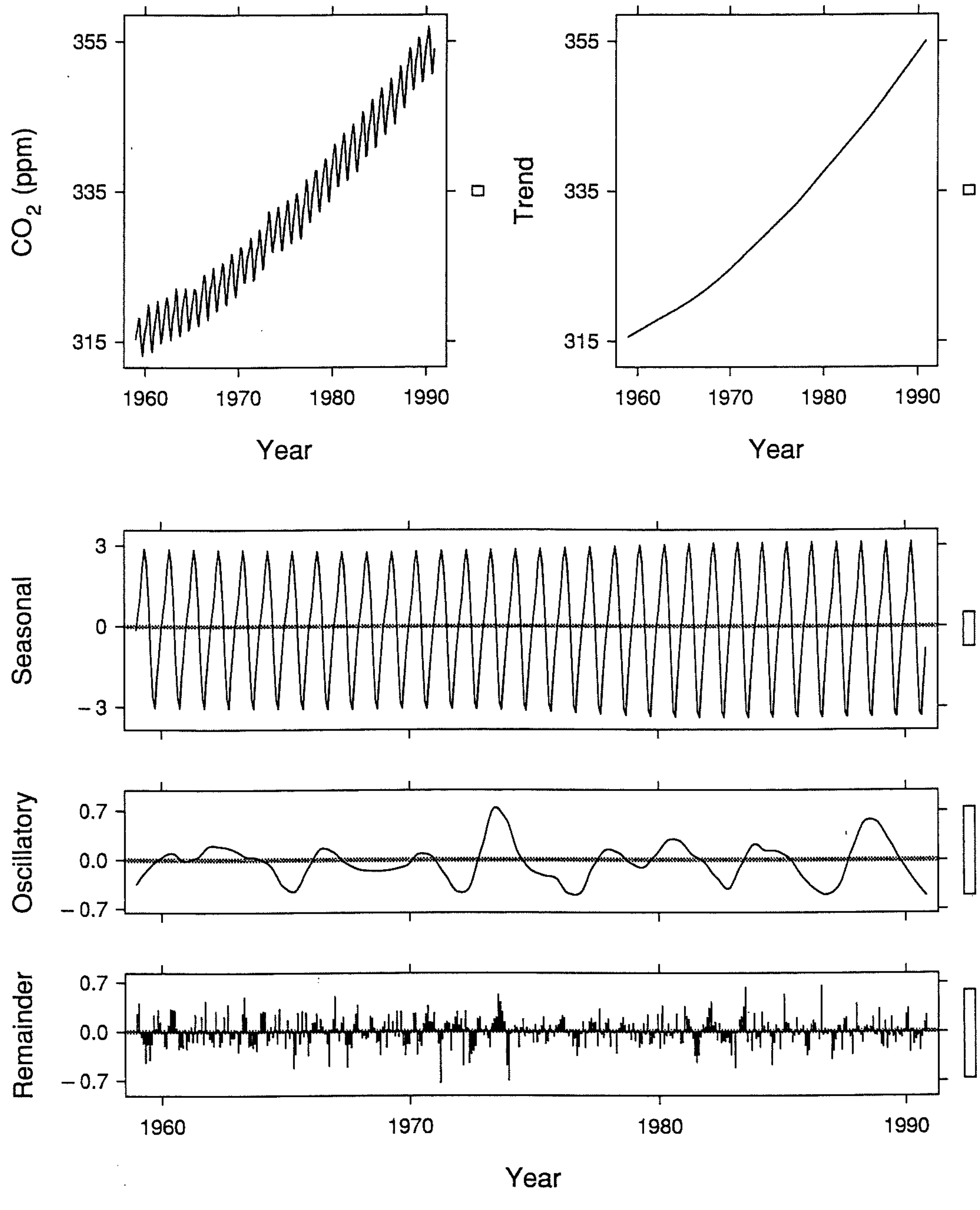
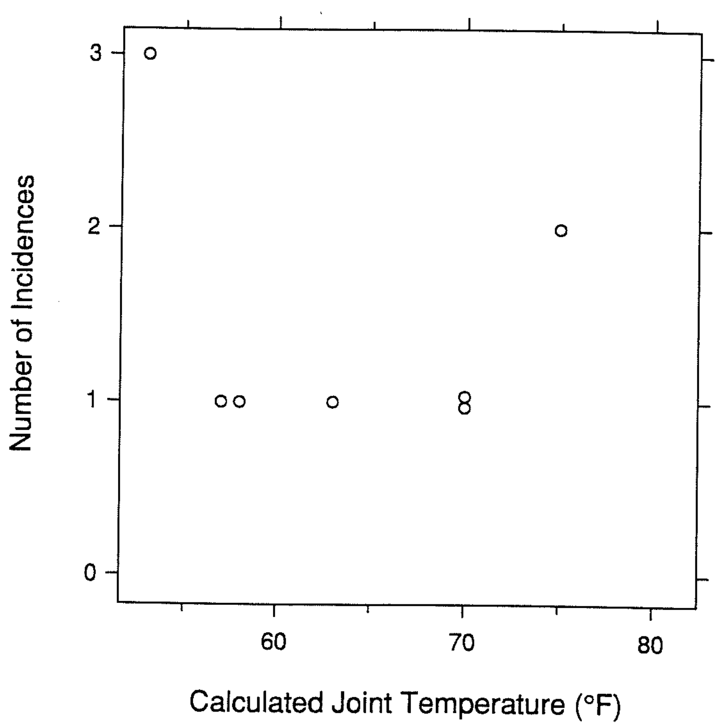


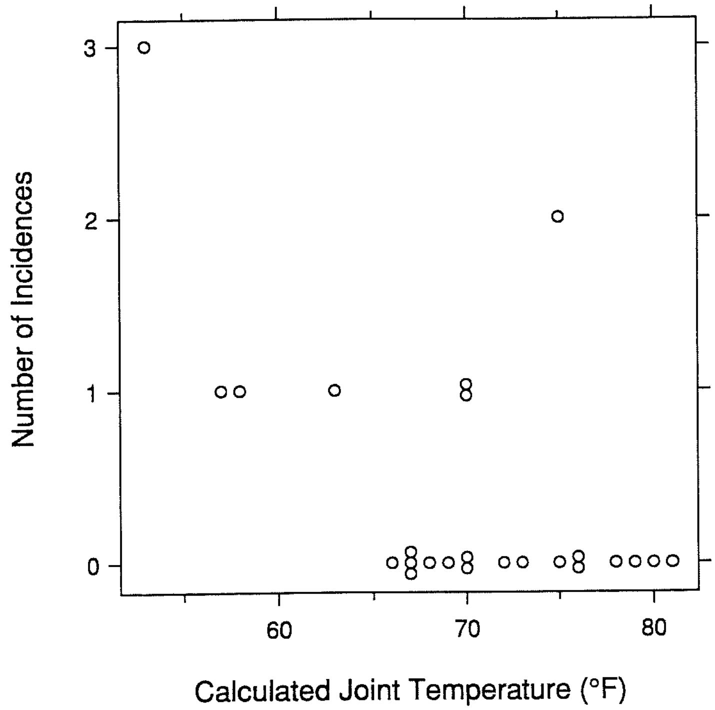
1.2 THE POWER OF GRAPHICAL DATA DISPLAY. Visualization provides insight that cannot be appreciated by any other approach to learning from data. On this graph, the top left panel displays monthly average CO_2 concentrations from Mauna Loa, Hawaii. The remaining panels show frequency components of variation in the data. The heights of the five bars on the right sides of the panels portray the same changes in ppm on the five vertical scales.

To assess the issue, the engineers studied a graph of the data shown in Figure 1.4. Each data point was from a shuttle flight in which the O-rings had experienced thermal distress. The horizontal scale is O-ring temperature, and the vertical scale is the number of O-rings experiencing distress. The graph revealed no effect of temperature on the number of stress problems, and Morton Thiokel, the rocket manufacturer, communicated to NASA the conclusion that the "temperature data [are] not conclusive on predicting primary O-ring blowby" [43]. The next day Challenger took off, the O-rings failed, and the shuttle exploded, killing the seven people on board.



1.4 STATISTICAL REASONING. These data were graphed by space shuttle engineers the evening before the Challenger accident to determine the dependence of O-ring failure on temperature. Data for no failures was not graphed in the mistaken belief that it was irrelevant to the issue of dependence. The engineers concluded from the graph that there is no dependence.

The conclusion of the January 27 analysis was incorrect, in part, because the analysis of the data by the graph in Figure 1.4 was faulty. It omitted data for flights in which no O-rings experienced thermal distress. Figure 1.5 shows a graph with all data included. Now a pattern emerges. The Rogers Commission, a group that intensively studied the Challenger mission afterward, concluded that the engineers had omitted the no-stress data in the mistaken belief that they would contribute no information to the thermal-stress question [43].

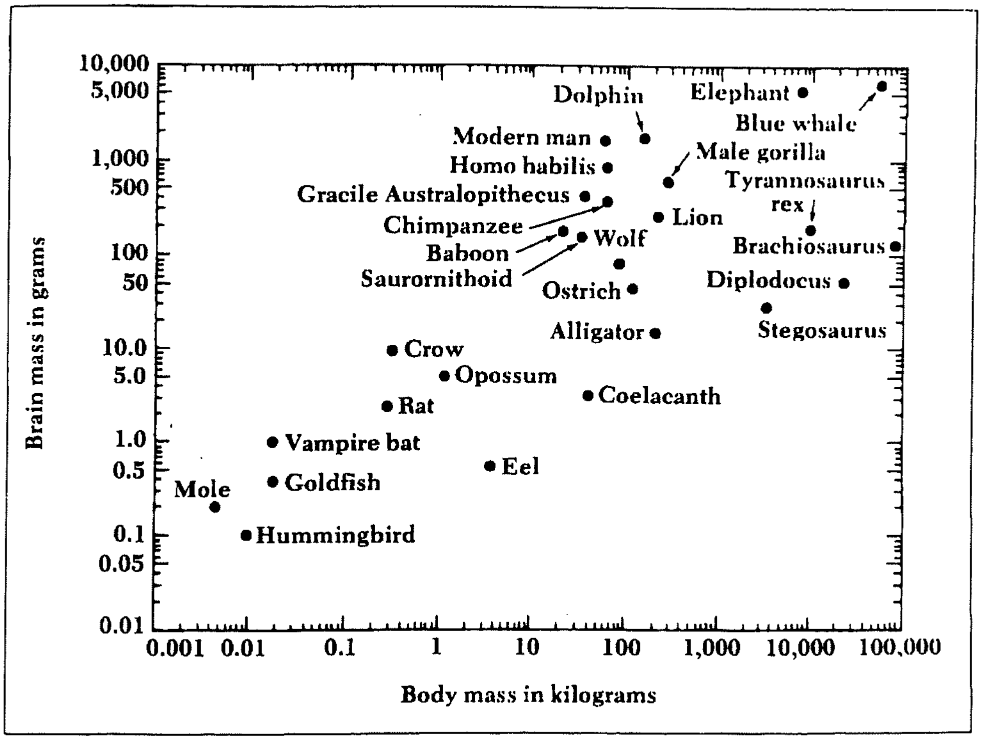


1.5 STATISTICAL REASONING. The complete set of O-ring data is now graphed, including the observations with no failures. A dependence of failure on temperature is revealed.

The graphical analysis of the O-ring data failed, not because of the display method used, as with the aerosol data, but rather because of a poor choice of the statistical information selected for the graph. This arose because of a flaw in the statistical reasoning that underlay the graph. The flaw violated a basic statistical principle: in the analysis of failure data, the values of a causal variable when no failures occur are as relevant to the analysis as the values when failures occur. Statistical thinking is vital to data display. A number of statistical principles are discussed in Chapters 2 and 3.

Brain Masses and Body Masses of Animal Species

Figure 1.6 is a graph from Carl Sagan's intriguing book, *The Dragons of Eden* [107]. The graph shows the brain masses and body masses, both on a log scale, of a collection of animal species. We can see that log brain mass and log body mass are correlated, but this was not the main reason for making the graph.



1.6 THE CHALLENGE OF GRAPHICAL DATA DISPLAY. This graph shows brain and body masses of animal species. The intent was for viewers to judge an intelligence measure, but the judgments require a visual operation that is too difficult.

What Sagan wanted to describe was an intelligence scale that has been investigated extensively by Harry J. Jerison [65]. Sagan writes that this measure of intelligence is "the *ratio* of the mass of the brain to the total mass of the organism." Later he adds, referring the reader to the graph, "of all the organisms shown, the beast with the largest brain mass for its body weight is a creature called *Homo sapiens*. Next in such a ranking are dolphins."

The first problem is that Sagan has made a mistake in describing the intelligence measure; it is not the ratio of brain to body mass but rather is (brain mass)/(body mass)$^{2/3}$. If we study a group of related species, such as all mammals, brain mass tends to increase as a function of body mass. The general pattern of the data is reasonably well described by the equation

$$\text{brain mass} = c\,(\text{body mass})^{2/3}.$$

Since the densities of different species do not vary radically, we may think of the masses as being surrogate measures for volume, and volume to the 2/3 power behaves like a surface area. Thus the empirical relationship says that brain mass depends on the surface area of the body; Stephen Jay Gould conjectures that this is so because body surfaces serve as end points for so many nerve channels [52]. Now suppose a given species has a greater brain mass than other species with the same body mass; what this means is that

$$(\text{brain mass})/(\text{body mass})^{2/3}$$

is greater. We might expect that the big-brained species would be more intelligent since it has an excess of brain capacity given its body surface. This idea leads to measuring intelligence by this ratio.

Let us now return to Figure 1.6 and consider the graphical problem, which is a serious one. How do we judge the intelligence measure from the graph? Suppose two species have the same intelligence measure; then both have the same value of

$$\frac{(\text{brain mass})}{(\text{body mass})^{2/3}} = r\ .$$

Thus

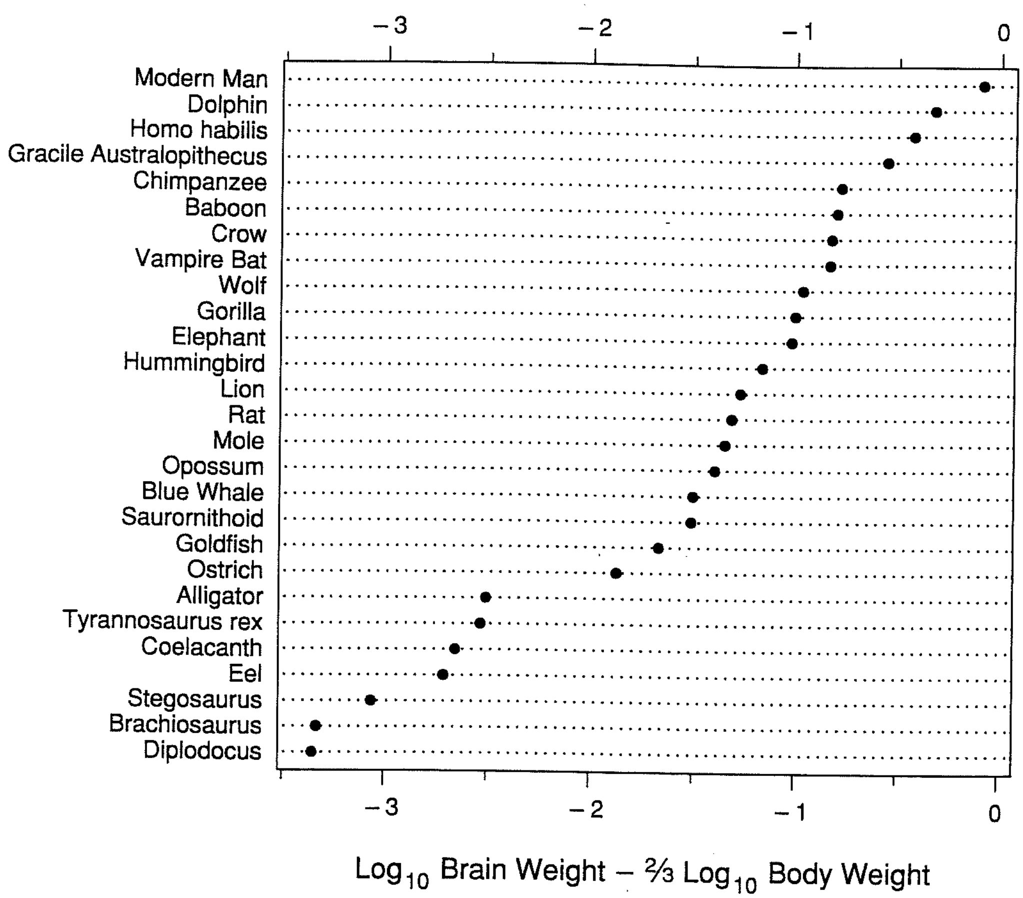
$$\log(\text{brain mass}) = 2/3 \log(\text{body mass}) + \log(r)$$

for both species. This means that in Figure 1.6, the two equally intelligent species lie on a line with slope 2/3. Suppose one species has a greater value of r than another; then the smarter one lies on a line with slope 2/3 that is to the northwest of the line on which the less intelligent one lies. In other words, to judge the intelligence measure from Figure 1.6 we must mentally superpose a set of parallel lines with slope 2/3. (If we attempt to judge Sagan's mistaken ratios, we must superpose lines with slope 1.) This visual operation is simply too hard.

Figure 1.6 can be greatly improved, at least for the purpose of showing the intelligence measure, by graphing the measure directly on a log scale, as is done in the dot plot of Figure 1.7. Now we can see strikingly many things not so apparent from Figure 1.6. Happily, modern man is at the top. Dolphins are next; interestingly, they are ahead of our ancestor *Homo habilis*.

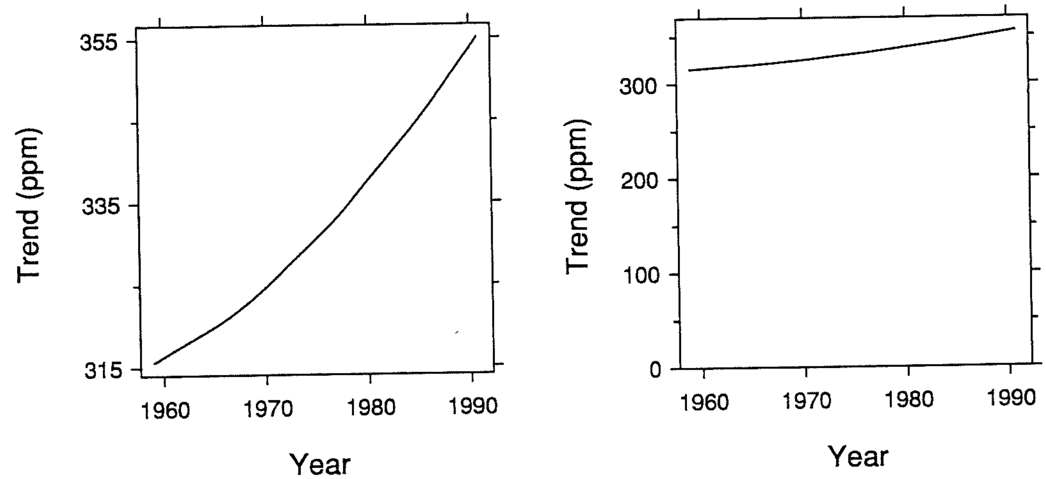
The problems with Figure 1.6 do not stop here. Five of the labels are wrong. The following are the corrections: "saurornithoid" should be "wolf," "wolf" should be "saurornithoid," "hummingbird" should be "goldfish," "goldfish" should be "mole," and "mole" should be "hummingbird." The correct labels yield the satisfying result that a hummingbird is smaller than a mole.

It should be emphasized that for some purposes, a corrected version of Figure 1.6 is a useful graph. For example, it shows the values of the brain and body masses and gives us information about their relationship. The point is that it does a poor job of showing the intelligence measure.



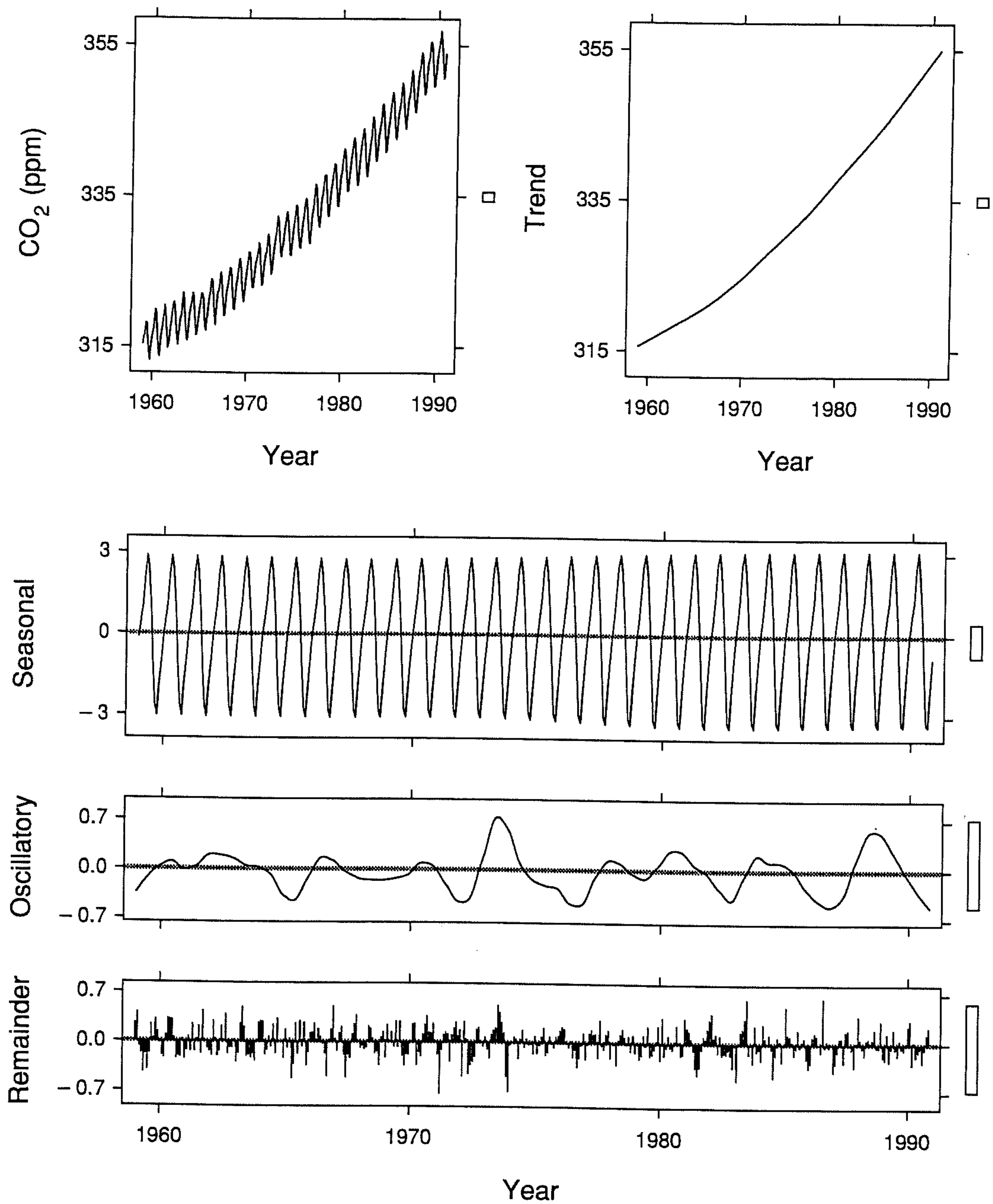
1.7 DOT PLOT. The intelligence measure is shown directly by a dot plot. (Both masses are expressed in grams for this computation.) The values of the measure can be judged far more readily than in Figure 1.6. For example, we can see modern man is at the top, even ahead of our very clever fellow mammals, the dolphins. Incorrect labels on Figure 1.6 have been corrected here.

The left panel of Figure 2.59 graphs the CO_2 trend curve from Figure 2.56. The sensible thing has been done; there is no zero and the segments are banked to 45°. The right panel includes zero and the result is ridiculous, even worse, misleading because the increase in the rate of change of CO_2 with time is not readily perceived because the orientations of the segments that make up the curve are so close to 0°. Were we to attempt both banking to 45° and including 0, keeping the width of the data rectangle the same, the height of the scale-line rectangle would be 32 cm, clearly impractical.

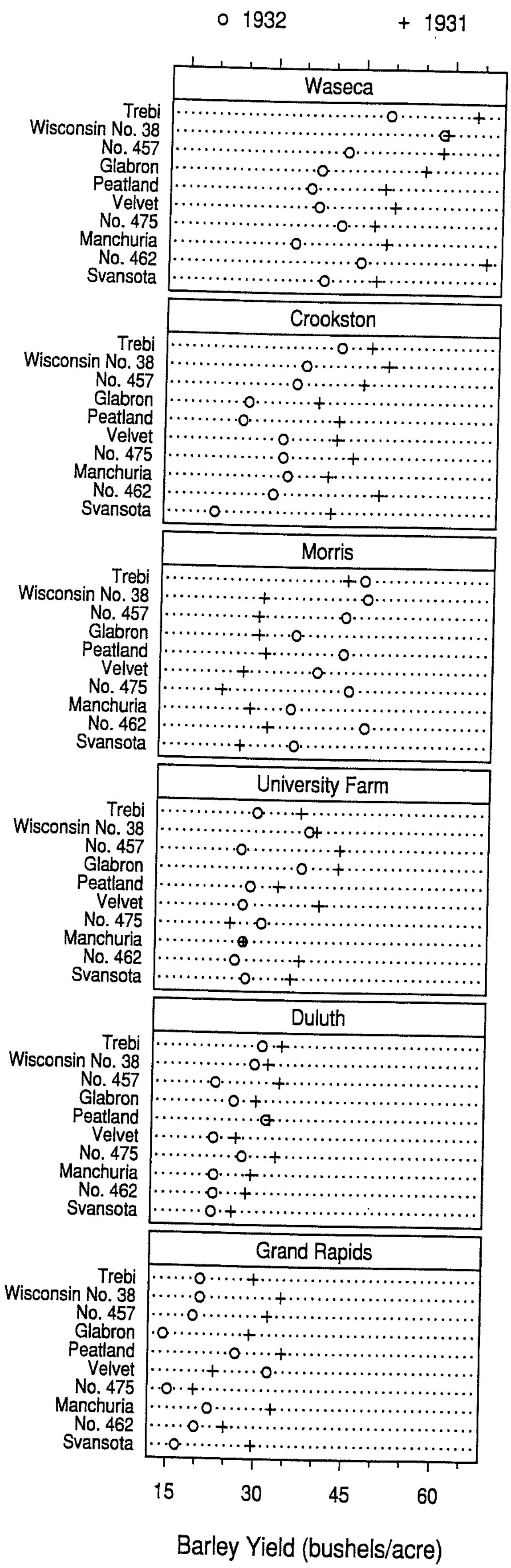


2.59 ZERO. *Do not insist that zero always be included on a scale showing magnitude.* The left panel displays the trend curve sensibly; the curve is banked to 45°. The right panel, which includes zero, does not allow effective judgment of the change through time because the aspect ratio is too small.

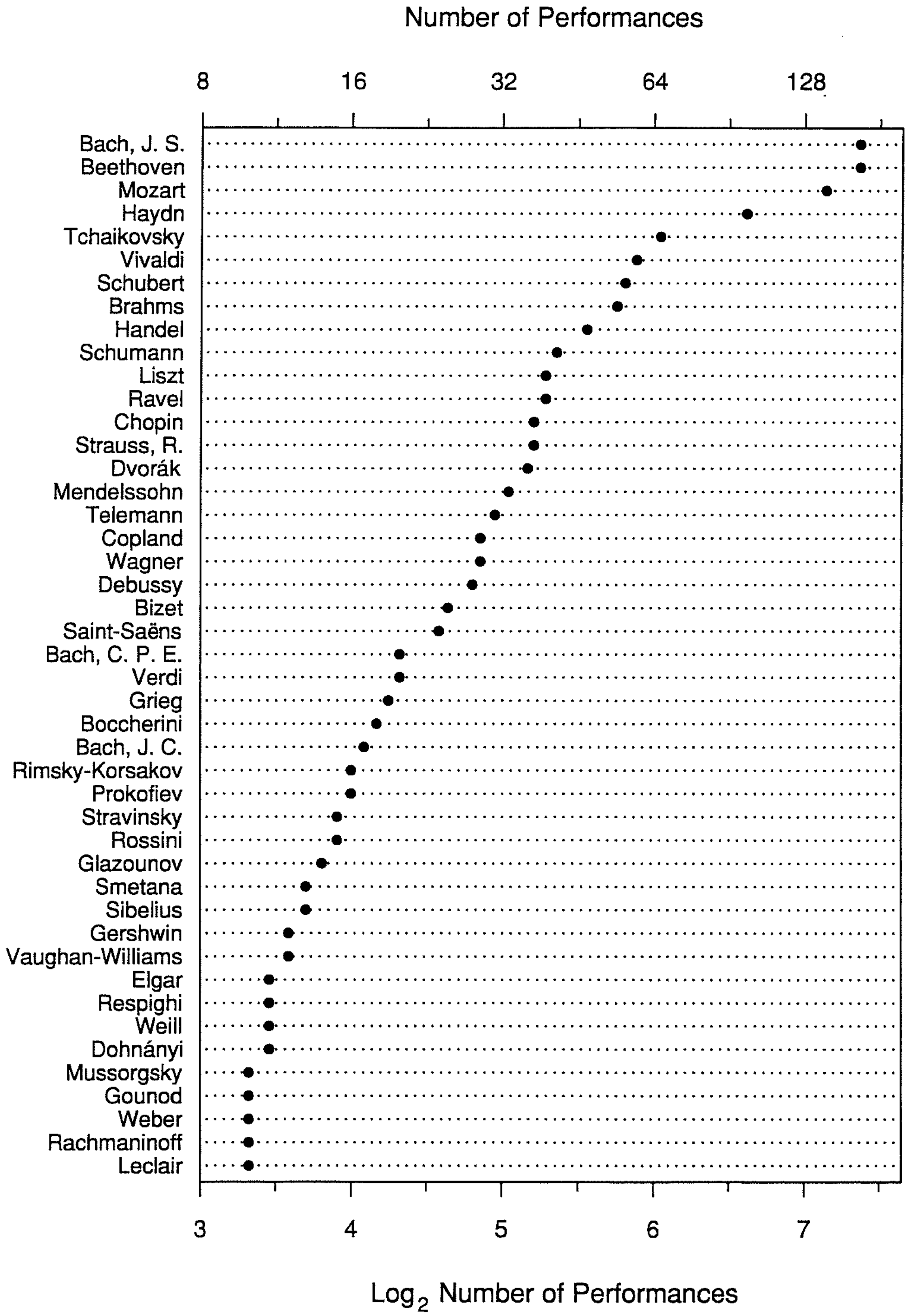
When this zero issue is contemplated calmly, and examples such as the previous ones are given, it all seems quite simple. For the CO_2 data, one would expect the zero-line issue not to arise since the right panel of Figure 2.59 is such a preposterous graph. But it did arise, and in a forum of great importance. In March 1981 members of the U.S. Senate convened scientists for testimony on global warming. There was an exchange between U.S. Senator Albert Gore, Jr., trying to galvanize scientists and politicians into action on global warming, and N. Douglas Pewitt, a witness for the U.S. Department of Energy who was resisting action. The exchange is described by Stephen Schneider in his book *Global Warming* [108]. It needs no comment:



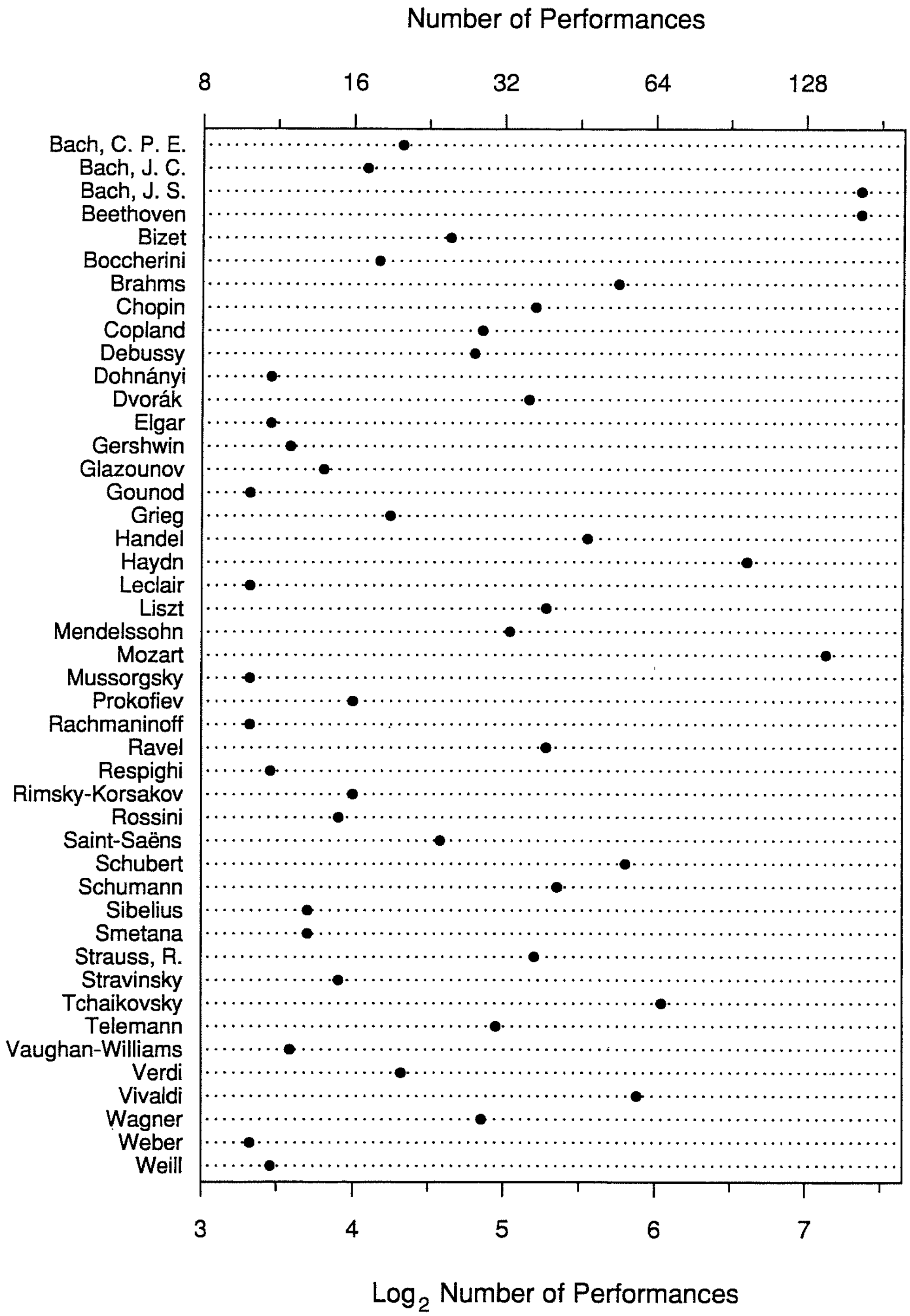
2.76 PACKING DATA. *A large amount of quantitative information can be packed into a small region.* The computer graphics revolution has given us the capability to graph a large amount of quantitative information in a small space. There are 1920 data points on this graph; each portrays two numerical values, so 3840 numbers are shown.



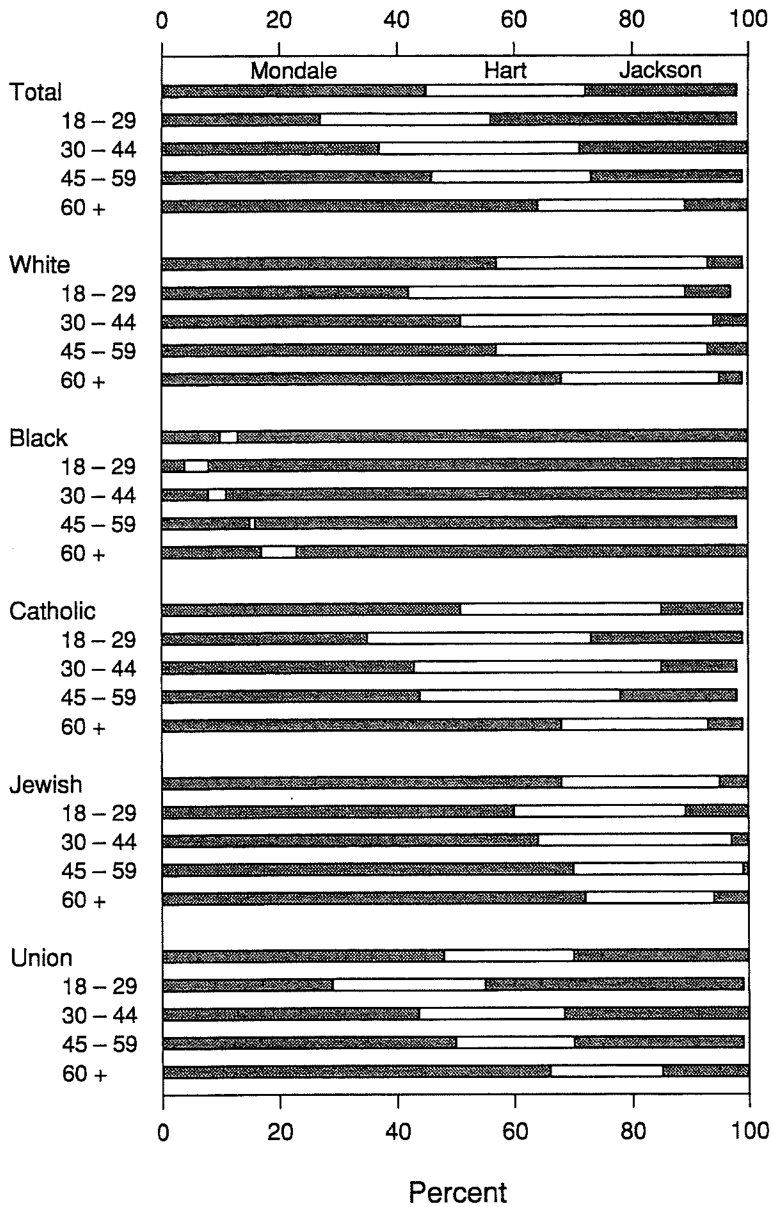
4.1 A MODEL FOR GRAPHICAL PERCEPTION. The model for graphical perception in this chapter provides a framework for studies of display methods. The model divides visual operations of graphical perception into pattern perception and table look-up.



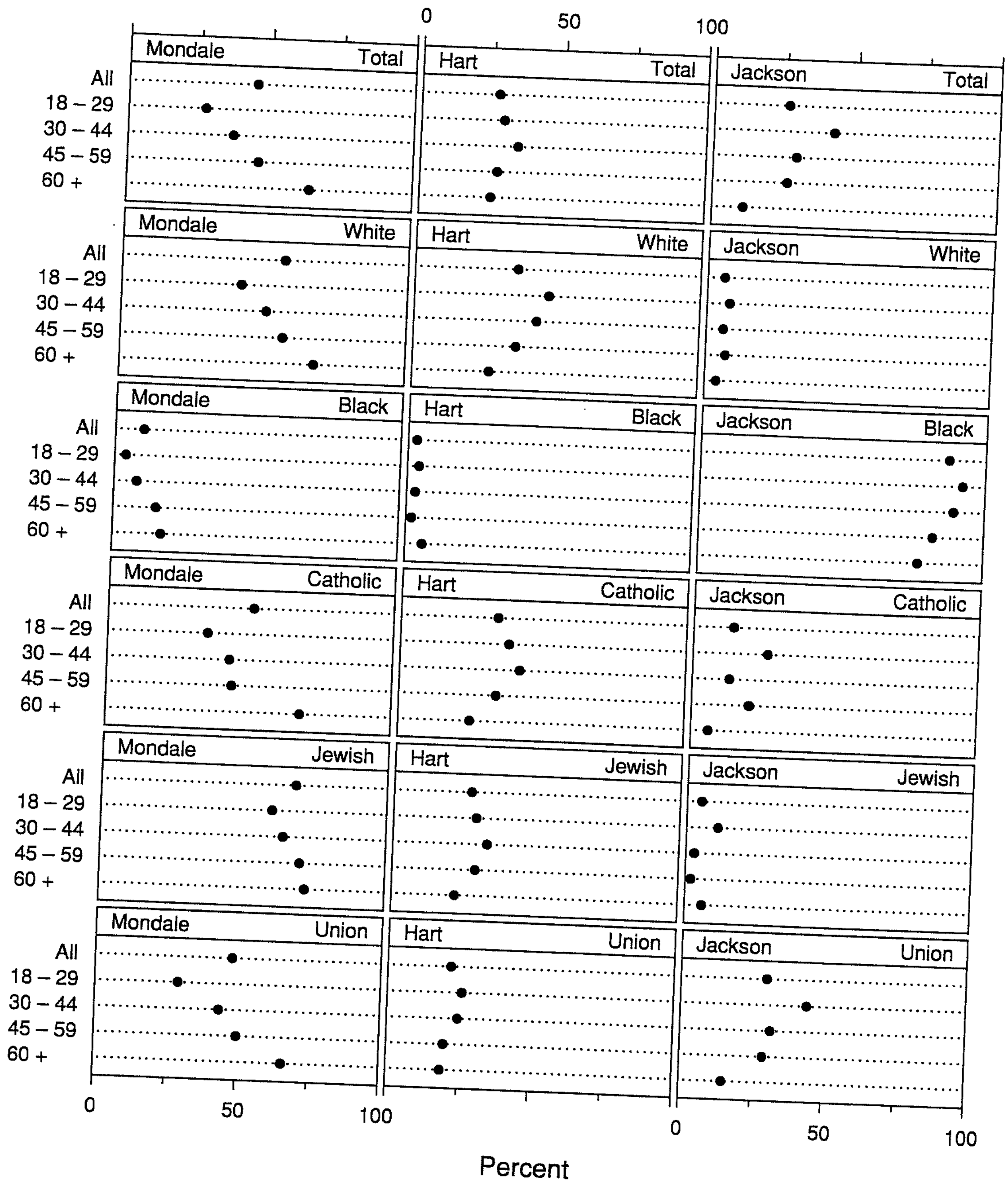
4.9 ORDER FOR DOT PLOTS. The data on this dot plot are ordered from smallest to largest. This enhances our visual decoding of the distribution of the values along the measurement scale.



4.10 ORDER FOR DOT PLOTS. The data on this dot plot are ordered alphabetically. This degrades our visual decoding of the distribution of the values along the measurement scale.



4.21 DIVIDED BAR CHART. A divided bar chart is used to show the percentage of the vote for three candidates in the 1984 New York Democratic primary election. The Mondale values are graphed by position along a common scale, but the Hart values and the Jackson values are not and our visual decoding of these latter two sets of values is less accurate than for the Mondale values.



4.22 MULTIWAY DOT PLOT. The data from Figure 4.21 are graphed by a multiway dot plot. Now the Hart values and the Jackson values are encoded by position along a common scale. Now we can perceive a Hart age pattern.