1. (a) The height of a rectangle in a histogram is the percent of data in that block divided by the width of the base of that block.

brightness unit range	how many light bulbs	percent of data	height of rectangle
50–90	100	$100/600 = 1/6 \approx 16.7\%$	$16.7/(90-50) \approx .417$
90–120	100	$100/600 = 1/6 \approx 16.7\%$	$16.7/(120-90) \approx .556$
120–140	200	$200/600 = 1/3 \approx 33.3\%$	$33.3/(140-120) \approx 1.67$
140–150	200	$200/600 = 1/3 \approx 33.3\%$	$33.3/(150-140) \approx 3.33$

(b) The median is the dividing point between the lower 50% and upper 50% of data values. Assuming date is evenly distributed within each of the 4 blocks, we estimate the median to be 130 brightness units, because the midpoint of the 3rd block has $1/6+1/6+1/6 = 1/2$ of the data below it, and has $1/6 + 1/3 = 1/2$ of the data above it.

(c) This computation is not valid because the data clearly do not follow a normal distribution. This computation would be correct if the data were normal. For the given data, we can see that $67\% \approx 2/3$ of the data lies in the range 50–140, so the 67th percentile score is about 140.

2. (a) The range "within 1 degree of average" corresponds to the range $-z$ to $+z$ on the normal table, with $z = 1/1.3 \approx .77$. The normal table gives a corresponding area of about 56%.

(b) We convert temperatures 100, 103 to standard units.

$$z = \frac{100 - 102.4}{1.3} =\approx -1.85$$
$$z = \frac{103 - 102.4}{1.3} =\approx .46$$

The normal table gives areas 93.57% and about 35% for these z values, respectively. Thus the total percent of measurements in the given range is estimated as $\frac{1}{2}(93 + 35) \approx 65\%$.

(c) We use the same calculation as given in 1 (c), but substitute the average and SD for the temperature data.

$$\text{67th percentile } \approx (.44)(1.3) + 102.4 \approx 103.0 \text{ degrees Fahrenheit}$$

(d) When rescaling a list of measurements by adding (or subtracting) a constant C and then multiplying by a constant factor k, the average of the list changes by adding (or subtracting) the constant C and then multiplying by the constant k, while the SD of the list changes only by multiplying by k. The SD is not affected by the sliding of the data right or left by adding (or subtracting) the constant C. Thus the new average is $(5/9)(102.4 - 32) \approx 39.1$ degrees Celsius, and the new SD is $(5/9)(1.3) = .72$ degrees Celsius.

3. The story describes an **association** between treatment with drug X and the rate of disease Y, but it does not establish **causation**. The issue is that, without further controls, the treatment and control groups could differ significantly in one or more characteristics that are also associated with disease Y. Such characteristics are called **confounders**. This is the main lesson of Chapters 1 and 2.

   A complete answer will begin by saying "no", there is not enough information given in this story to conclude that drug X is effective. The complete response will explain that conclusion by saying that while there is an **association** between treatment with drug X and rates of disease Y, this is not enough to establish **causation**, because there could be **confounders**. Finally, the complete response will say that the evidence for causation would be strong if the design of the drug trial was **randomized** (random assignment of subjects to treatment and control groups) and **double-blind** (neither the subjects nor the researchers should know whether a given subject is in the treatment or the control group during the trial). A complete response should use all of the vocabulary terms in bold type in the previous sentences.