1. (a) The cases are water python eggs.

(b) There are 187 cases, eggs about which information is known.

(c) What is known about each egg is

nest environment, a categorical variable with values cold, neutral & hot

fate, a categorical variable with values hatched and lost.

(d) They come from an experiment, in which researchers assigned values of nest environment, the explanatory variable.

(e) $\Pr(\text{cold nest}) = 27/187$.

(f)
$$\Pr(\text{cold nest or lost}) = \Pr(\text{cold nest}) + \Pr(\text{lost}) - \Pr(\text{cold nest and lost})$$
$$= \frac{27}{187} + \frac{58}{187} - \frac{11}{187} = \frac{74}{187}.$$

(g) $\Pr(\text{neutral and lost}) = 18/187$.

(h)
$$\Pr(\text{cold nest} \mid \text{hatched}) = \frac{\Pr(\text{cold nest and hatched})}{\Pr(\text{hatched})} = \frac{16/187}{129/187} = \frac{16}{129}$$

(i) Of the options here, only bar graphs are for categorical data. Option (ii) is appropriate.

2. Only in Option (a) do we see the salient features of a binomial setting:

Success = "hit the target", and X counts occurrences of Successes

Pr(Success) is reasonably the same on each trial/shot

Set number of trials is fixed at the number of arrows in her quiver, 10.

3. Correlation can be calculated when two quantitative variables are present, but the concept is meaningful/sensible only when the variables exhibit a linear pattern in their scatterplot.

4. (a) Both variables under consideration here — whether you used a night light, and whether you develop myopia — are categorical. What statisticians mean by the term "correlation" cannot be calculated in this context.

(b) Probability must be a number between 0 and 1. It cannot be 100.

(c) Pam's evidence for her statement is observational data. Even if she were able to observe the same phenomenon in hundreds of football players and establish that these variables have an association, the claim of causation is not supported.

5. The commands I provide work, and are ones you should know. They aren't the only options.

(a) nrow(cars)

(b) tally(~ numSiblings, data = survey)

(c) mean(~ weight, data = cars)

(d) coefficients(lm(gasMileage ~ weight, data = cars))

(e) gf_boxplot(~ sqFt, data = houses)

(f) names(houses)

(g) cor(gasMileage ~ weight, data = cars)

(h) filter(houses, sqFt > 2000)

6. The mode in the bar graph is specific to countries that have a high income level (whatever that means), and shows about 58% of adults in those countries are overweight or obese. This 58% need not translate into an actual number of people which is large when compared with the world's population. So, Option (b) is the only one supported.

7. The standard deviation, correlation and range are all sensitive to outliers.

8. (a) is an instance of blocking, and is common practice.

(c) is so deeply woven into doing experiments that it is difficult to imagine any experiment not employing it.

(d) is an instance of matched-pairs, and is common.

(f) is what we mean when we modify "blind" with the word "double". It helps to eliminate researcher bias.

In contrast to the aforementioned common practices, (b) and (e) appear to be attempts to bias results in favor of making a treatment appear to be effective.

9. First, $$\bar{x} = \frac{1}{4}(11 + 36 + 23 + 34) = \frac{1}{4}(104) = 26.$$

So $$s^2 = \frac{1}{3}\left[(11-26)^2 + (36-26)^2 + (23-26)^2 + (34-26)^2\right] = \frac{1}{3}(225 + 100 + 9 + 64) = \frac{398}{3} = 132.67$$

$$\Rightarrow s = \sqrt{132.67} \doteq 11.52.$$

10.

Scatterplot	r
C	0.94
A	0.68
D	0.04
B	-0.86

11. (a)

Histogram	Boxplot
(i)	B
(ii)	C
(iii)	A
(iv)	D

(b) (ii) and (iv) are generally symmetric

(c) (i), (iii) and (iv) are unimodal

(d) (i) is positively-skewed, making the mean larger

(e) The first quartile is farthest to the left (smallest) for C

(f) A has the smallest IQR.