Ch 1

1. **d** 20% of 50 states is 10 states.
2. **c** The vertical scale exaggerates small differences between the three brands, all of which are in the mid-90% range.
3. **b** Stems should include all but the last meaningful digit. In this case, the last digit is ones, so the stems should be multiples of 10 (09 represents 90, 10 represents 100, etc.)
4. **c** The larger values in the long tail of a right-skewed distribution will cause the mean to be larger that it would be in a symmetric distribution, but will not affect the median. This will make the mean larger than the median.
5. **b** The five number summary for these data is 2.6 3.8 4.5 5.6 6.3, which corresponds to boxplot (b).
6. **c** Proportion of males with "Environment" response is 29/110 = .264. Same proportion for females is 33/140 = .236.
7. **b** Adding $30,000 to the total earned by 6 people will increase the mean by $30,000/6 or $5,000. Since the owner's salary is already the maximum value in the distribution, increasing it will not change the median.
8. **a.** The mean speed of all vehicles will be a weighted average of the speeds of cars on both sets of lanes. This average must be between 64 and 68 miles per hour.
9. **d** Standard deviation is a type of average of how far each individual infant's weight is from the mean weight.
10. **c** 110 is the first quartile and 200 is the maximum. This range of weights includes all individuals except the 25% below the first quartile.

Ch 2

1. **b** 95% of the heights should be within 2 standard deviations of the mean, which is the interval 68 – 5 to 68 + 5 or 63 to 73. Thus 5% are outside that interval.
2. **c** 70.5 is one standard deviation above the mean. Since 68% are within one standard deviation, the remaining 32% are evenly divided into the regions above and below that interval.
3. **a** The area under any density curve is equal to 1. This curve is skewed left, so it isn't symmetric, Normal, or skewed right.
4. **c** Since the curve is skewed left, the mean will be "pulled down" by the values in the lower tail, making it less than the median.
5. **c** Percentile of $z = 1.6$ minus percentile of $z = -0.3$ is .9452 – .3821 = .5631.
6. **e** The distinct curve in the plot indicates that the distribution is strongly skewed (right).
7. **d** The distribution of z-scores is a linear transformation, so it won't change the shape of the distribution. (See explanation at Quiz 2.1A).
8. **d** $Q_1 \approx 5.5$ and $Q_3 \approx 6.7$, so IQR is approximately 1.2.
9. **c** All Normal curves are symmetric and peak at the mean, but 95% of the area is within $\underline{2}$ standard deviations of the mean, not 1.
10. **c** $z_{30} = \dfrac{30 - 22}{11.9} = 0.67$, which is the 75th percentile of the Standard Normal curve, so about 25% of the times are above that value.

1. a We expect fuel efficiency to be (at least partially) determined by engine size, with larger engines consuming more fuel. Hence the gas mileage should go down as engine size goes up.

2. d $\hat{y} = 10 + 0.9(95) = 95.5$

3. b $y - \hat{y} = 93 - (10 + 0.9(90)) = 93 - 91 = 2$

4. c Correlation is not affected by the units in which the variables are expressed.

5. e (a) and (b) are incorrect because one or more of the variables is categorical; (c) is incorrect because r cannot have units, such as meters per second; (d) is incorrect because r cannot be greater than 1.

6. e Since the point $(\bar{x}, \bar{y})$ must be on the least-squares regression line, knowing that $\bar{x} = 1.2$ means $\bar{y} = 1400 + 2000(1.2) = \$3800$.

7. c A linear model is only appropriate if there is no discernible pattern in the residual plot.

8. b The least squares line minimizes the squares of the vertical distances between the points and the line, which is the difference between observed and predicted values of the response variable, yield.

9. c (c) interprets the equation's slope. (a) would be true if it said *decrease*; (b) mixes up the explanatory and response variables; (d) predicted yield when UV = 20 is $\hat{y} = 3.98 - .046285(20) \approx 3.05$.

10. d Statement I is incorrect because the line would minimize residuals for the other variable. II and III are both correct (see "Correlation and Regression wisdom" in Section 3.2).

1. c Absence of control makes it impossible to distinguish the effect of the remedy from other variables, such as time.

2. b $56, 92, 96$, and 79 are ignored (unassigned); repeat of 23 is ignored.

3. d The Sugar Maples in each plot are a cluster. The manufacturer is assuming that they mirror the characteristics of the entire population of Sugar Maples in the forest.

4. c Answer (b) is a common misperception: while the creation of similar groups by randomization can reduce the impact of some confounding variables, there needs to be control as well.

5. b Each worker acts as his own "pair."

6. d Samples taken from the same population by exactly the same method will vary because of chance variation.

7. d Treatment refers to the specific experimental characteristics of a single group, experimental units are the individual items that are randomly assigned to groups and upon which the response variable is measured, response is the variable we think will vary in different treatments.

8. c The variable city size is confounded with presence or absence of a Major League Baseball team.

9. a A control group that is treated in the same way as the experimental group in every way except for the designated treatment will isolate the effect of the treatment.

10. d Since this is a random sample of the population of all nursing graduates at this school, we can make inferences about all the nursing graduates at the school. Since the data will come from a survey, not a controlled experiment, we cannot infer any cause-and-effect relationships.

1. a The probability remains 1/3 every day regardless of previous days' events.

2. c This is the only option that sums to 1 and has all probabilities between 0 and 1, inclusive.

3. b $P(\text{bus}) = \dfrac{170}{550} \approx 0.309$

4. b $P(\text{walks} \mid \text{junior}) = \dfrac{48}{300} = 0.16$

5. b Independent because P(walks) = (walks| junior) = 0.16, not mutually exclusive because P(Walks and Junior) $\neq$ 0

6. a Correct by general addition rule. [(b) is not correct since the events are disjoint].

7. c $P(A \text{ or } B) = 0.24 + 0.52 - (0.24)(0.52) = 0.6352$

8. c $P(\text{at least one type O-neg}) = 1 - P(\text{no type O-neg in 10 people}) = 1 - (.9280)^{10} \approx 0.526$

9. e $P(diabetes) = (0.86)(0.028) + (0.12)(0.044) + (0.02)(0.035) \approx 0.030$

10. d $P(B \cap B) + P(G \cap G) = \dfrac{6}{16} \cdot \dfrac{5}{15} + \dfrac{10}{16} \cdot \dfrac{9}{15} = 0.5$