Chapter 2 Solutions

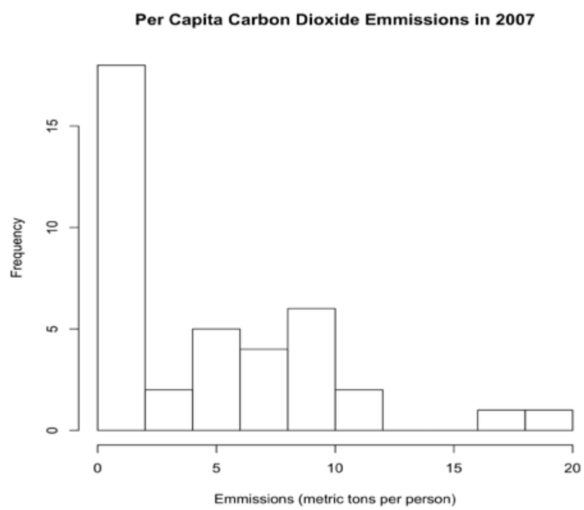
2.1: Mean breaking strength = 30841 pounds. Only 6 pieces have strengths less than the mean. The mean is so small relative to the data because of the sharp left skew (low outliers).

2.2: The mean expenditure for all countries including the United States is $2332.20. The mean when the United States is excluded is $2186.53. Hence the United States as an outlier increases the mean by about $145.67, even with as many as 34 other countries.

2.3: The mean travel time is 31.25 minutes. The median travel time is 22.5 minutes. The mean is significantly larger than the median due to the right skew in the distribution of times.

2.4: The mean is larger than the median, for surely the distribution of home prices is right skewed.

2.5: A histogram is given below. Note the right skew. Hence, the mean is larger than the median. Here, the mean is 4.61 and the median is 3.95 tons per person.



2.6: (a) and (b) A back to back stemplot is provided. The five-number summaries are tabulated. (c) It seems that the offensive line players are heavier. Perhaps there is one outlier—one 325-pound defensive lineman.

	Minimum	Q1	Median	Q3	Maximum
Offensive line	304	309.5	319	331.5	344 pounds
Defensive line	280	285	300	305	325 pounds

Offensive line		Defensive line
	28	0 5
	29	8
4 4	30	0 5 5
9 8 5	31	
5 4	32	5
8	33	
4	34	

2.7: (a) Minimum = 9, Q1 = 16, Median = 18, Q3 = 22, Maximum = 51. (b) The boxplot shows right skew in the distribution of MPG values.

2.8: For these data, Q1 = 10, Q3 = 30, and so IQR = 30 – 10 = 20 minutes. Hence, Q1 – 1.5 x IQR = 10 – 1.5 $\times$ 20 = –20 minutes. Obviously no times can be negative, so no outliers are in the left tail. Q3 + 1.5 $\times$ IQR = 30 + 1.5 $\times$ 20 = 60 minutes. Hence, the “60” would not be considered an outlier, but it’s close.

2.9: IQR = 22 – 16 = 6, so Q3 + 1.5 $\times$ IQR = 22 + 1.5 $\times$ 6 = 31. There are 5 values greater than 31 that would be identified as potential outliers (33, 35, 41, 41, 51). Since Q1 – 1.5 $\times$ IQR = 16 – 1.5 $\times$ 6 = 7, there are no potential outliers below 7.

2.10: (a) $\bar{x}$ = (5.2 + 13.8 + 8.6 + 16.8)/4 = 44.4/4 = 11.1 picocuries. (b) The standard deviation can be computed in steps:

x	5.2	13.8	8.6	16.8
$x - \bar{x}$	–5.9	2.7	–2.5	5.7
$(x - \bar{x})^2$	34.81	7.29	6.25	32.49

Hence, $s^2 = \dfrac{1}{n-1}\sum(x-\bar{x})^2 = \dfrac{1}{4-1}(34.81 + 7.29 + 6.25 + 32.49)$ = 26.94667

So $s = \sqrt{s^2} = \sqrt{29.94667}$ = 5.19 picocuries.

2.11: Both data sets have the same mean and standard deviation (about 7.5 and 2.0, respectively). However, construct simple stemplots to reveal that Data A have a very left-skewed distribution, while Data B have a slightly right-skewed distribution.

2.12: (a) No. The distribution isn't symmetric. (b) Yes. The distribution is symmetric and mound-shaped with no severe outliers. (c) No. The distribution is strongly right-skewed.

2.13: Group 1: $\bar{x}$ = 23.7500, s = 5.06548. Group 2: $\bar{x}$ = 14.0833, s = 4.98102. Group 3: $\bar{x}$ = 15.7778, s = 5.76146.

2.14: Both groups (developing countries and developed countries) have right-skewed distributions for unpaid parking tickets. Comparing, developing countries' diplomats tend to have more unpaid tickets. National income alone, however, does not explain countries whose diplomats have more or fewer unpaid tickets.

2.15: (b) 167.48

2.16: (b) 168.25

2.17: (b) 151.6, 163.5, 168.25, 174.3, 177.6

2.18: (c) the mean is greater than the median.

2.19: (b) 50%.

2.20: (c) the five-number summary.

2.21: (c) 8.2.

2.22: (a) $0 \le s$.

2.23: (b) seconds.

2.24: (a) the median.

2.25: The distribution of incomes in this group is almost certainly right-skewed, so the mean is $58,762 and the median is $46,931.

2.26: In both cases (for the under 35 crowd and for all families), the distribution of account sizes is right-skewed. Lots of people have very small retirement savings accounts.

2.27: With 842 colleges (an even number), the median location is (842 + 1)/2 = 421.5, so the median is computed by averaging the 421st and 422nd endowments sizes. The first quartile, Q1, is found by taking the median of the first 421 endowments (when sorted). This would be the (421+1)/2 = 211th endowment. Similarly, Q3 is found as the 632nd endowment (211 endowments above the median).

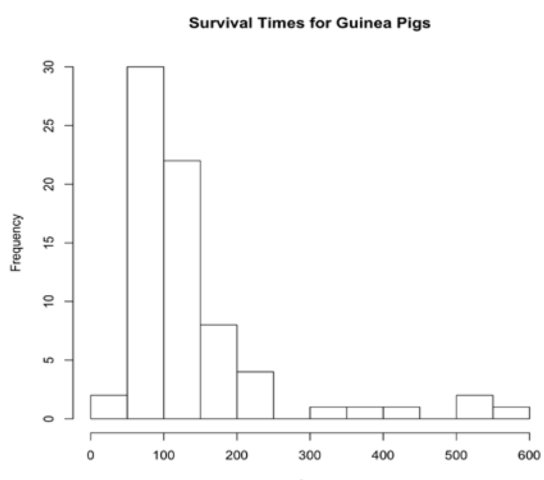
2.28: (a) Minimum = 23040, Q1 = 31975, Median = 31975, Q3 = 32710, Maximum = 33650. (b) Notice that the Minimum is much farther from Q1 than the Maximum is from Q3. This suggests a long left tail, consistent with a left-skewed distribution.

2.29: The five-number summaries for the three species are tabulated below. Boxplots don't add much information not already present in the stemplots.

	Minimum	Q1	Median	Q3	Maximum
Bihai	46.34	46.71	47.12	48.25	50.26
Red	37.4	38.07	39.16	41.69	43.09
Yellow	34.57	35.45	36.11	36.82	38.13

2.30: (a) Median = 2, Q1 = 1, Q3 = 4, (b) $\bar{x}$ = (15)(0) + (11)(1) + (15)(2) + (11)(3) + (8)(4) + (5)(5) + (3)(6) + (3)(7) + (3)(8)]/74 = 194/74 = 2.62 servings. This is larger than the median because the distribution is right-skewed.

2.31: A histogram of the survival times follows. The distribution is strongly right-skewed, with center around 100 days, and spread 0 to 600 days. (b) Because of the extreme right skew, we should use the five-number summary: 43, 82.5, 102.5, 151.5, 598 days. Notice that the median is closer to Q1 than to Q3.



2.32: (a) If countries or years have very different numbers of babies born, it would be unreasonable to compare across years or across countries by counts. (b) 4,243,333 babies. (c) The distribution is left-skewed. (d) The median is the 2,121,667th baby weight, and falls in the interval 3000 to 3499 grams. Q3 is in the interval 3500 to 3999 grams. Q1 is in the interval 2500 to 2999 grams.

2.33: (a) Symmetric distributions. (b) Removing the outliers reduces both means and both standard deviations.

2.34: (a) Mean (green arrow) moves along with moving point. Median (red arrow) points to middle point (rightmost nonmoving point). (b) Mean follows moving point. When moving point passes rightmost fixed point, median moves with it until moving point passes leftmost fixed point—then median stays there.

2.35: (a) The 6th observation must be placed at median for the original 5 observations. (b) No matter where you put the 7th observation, the median is one of the two repeated values above.

2.36: Both distributions are very similar: On weekdays more babies are born, and there is from weekday to weekday, though Mondays appear to have slightly fewer births. On weekends, fewer births take place. Of course, many more births take place in the United States.

2.37: The mean for all 51 entries is 8.4%, far from the national percentage of 12.5%. You can't average averages. Some states, like California and Florida, are larger and should carry more weight in the national percentage. Indeed, there are more people over the age of 65 living in Florida than there are residents in Wyoming.

2.38: More than half of all American households do not carry credit card debt.

2.39: (a) Pick any four numbers all the same: for instance, (4,4,4,4) or (6,6,6,6). (b) (0,0,10,10). (c) There is more than one possible answer for (a), but not for (b).

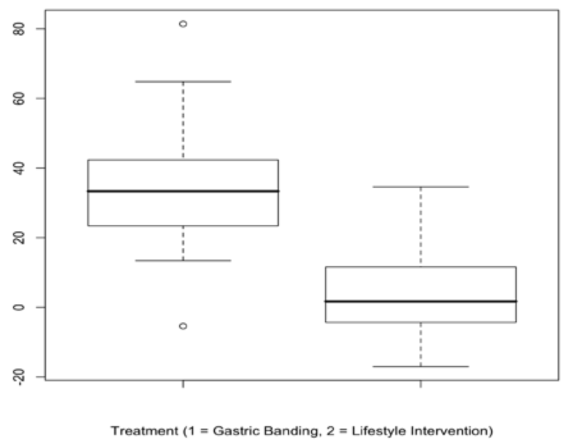
2.40: The TI-89 calculator used by the author reported $s = 1$ for the list

100,000,000,001 100,000,000,002 100,000,000,003. At some point, the calculator will fail… but for virtually any practical setting, a decent calculator will correctly compute.

2.41: Lots of answers are possible. Start by insuring that the median is 7, by "locking" 7 as the 3rd smallest value. Then, adjust the minimum or maximum accordingly to acquire a mean of 10 (so they sum to 50). One solution: 5 6 7 8 24.

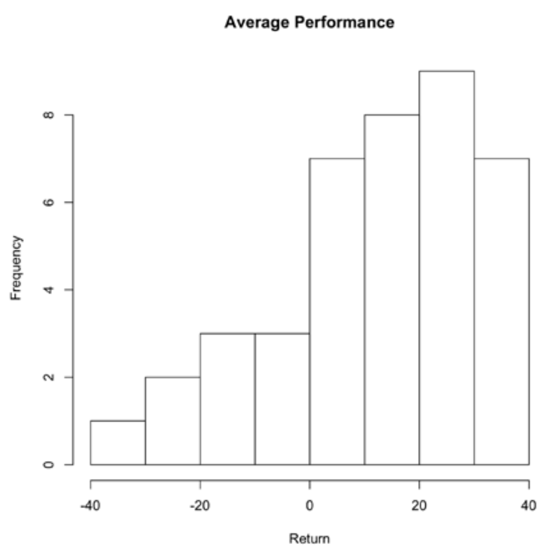
2.42: Lots of answers are possible. One solution: (–100, 1, 2, 3, 4, 5, 6).

2.43: (a) Weight losses that are negative correspond to weight *gains*. (b) A side-by-side boxplot (a version that reports suspected outliers using the 1.5 IQR rule) is provided below. Gastric banding seems to produce higher weight losses, typically. (c) It's better to measure weight loss relative to initial weight. (d) If the subjects that dropped out had continued, the difference between these groups would be as great or greater because many of the "lifestyle" dropouts had negative weight losses (i.e., weight gains), which would pull that group down.



2.44: The distribution of Candiens players' salaries is very right-skewed. The median salary is $1,425,000 (while the mean is $2,520,646, consistent with a strong right-skew). The middle half of players earn between $756,250 and $4,416,500, although a handful earn more than $5,000,000.

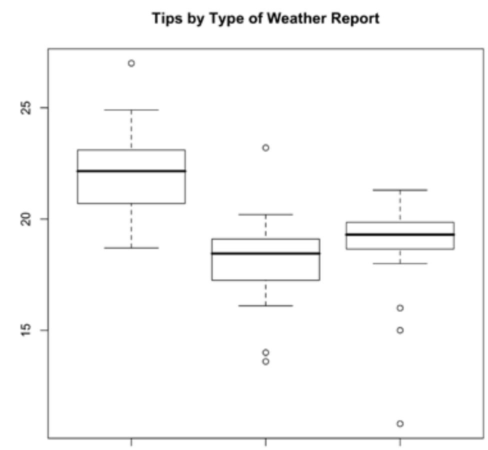
2.45: The distribution of average returns is skewed-left. Most years, average return is positive. Returns range from about –40% to 40%, with the median return about 16%.



2.46: Comparing side-by-side boxplots, Lavender seems to produce the highest customer expenditures.

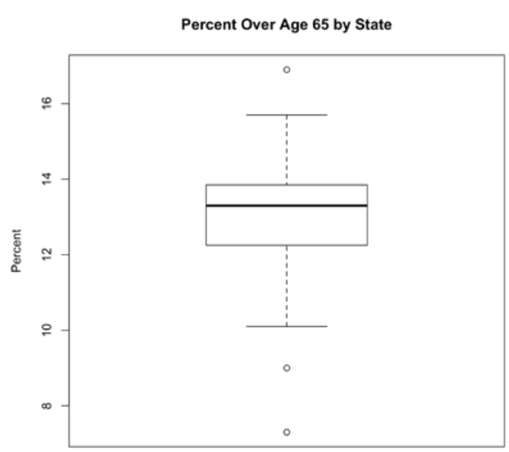
2.47: Based on side-by-side boxplots, lean people spend relatively more time active, but there is little difference in the time these groups spend lying down.

2.48: A side-by-side boxplot of tip results follows. Good weather forecasts generally yielded better tips, while there was little to no difference between a bad forecast and no forecast.



2.49: The distribution is also right-skewed. The median salary was $300, and the middle half of salaries were between $167.50 and $450. A handful of Canadians made $1000 or more. One earned $2200.

2.50: (a) 7.0, 12.1, 13.0, 13.6, 16.9. (b) A boxplot follows and suggests rough symmetry. There are three outliers (7.0% for Alaska, 8.8% for Utah, and 16.9% for Florida) using the 1.5 IQR rule.



2.51: (a) Min = 0.0272, Q1 = 0.6449, Median = 3.954, Q3 = 8.1555, Max = 18.9144. Notice that the maximum is farther from Q3 than the minimum is from Q1. This suggests right skew. (b) IQR = 8.1555 – 0.6449 = 7.5106. Hence, 1.5 $\times$ IQR = 11.2659. Now Q1 – 1.5 $\times$ IQR = 0.6449 – 11.2659 <0, so no values are more than 1.5 IQR's below Q1. Also, Q3 + 1.5 $\times$ IQR = 8.1555 + 11.2659 = 19.4214, so there are no high outliers. This rule is rather conservative – most people would easily call the United States' value (18.9144) a far outlier, and perhaps Canada would be considered an outlier, too.

2.52: There are no salaries greater than $9,906,875. This is the salary that is 1.5 IQRs greater than Q3.

2.53: Any of the 11 incomes more than $873.75 would be considered an outlier by the 1.5 IQR rule.

2.54 and 2.55 are Web-based exercises.