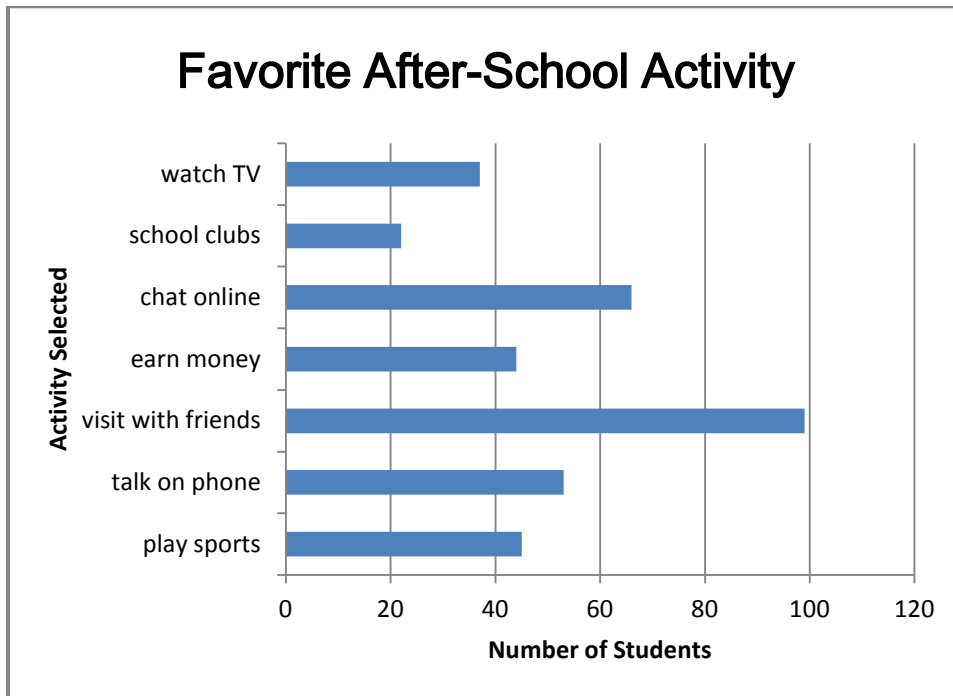


## CHAPTER 5 SOLUTIONS

### Section 5.1

1) a)

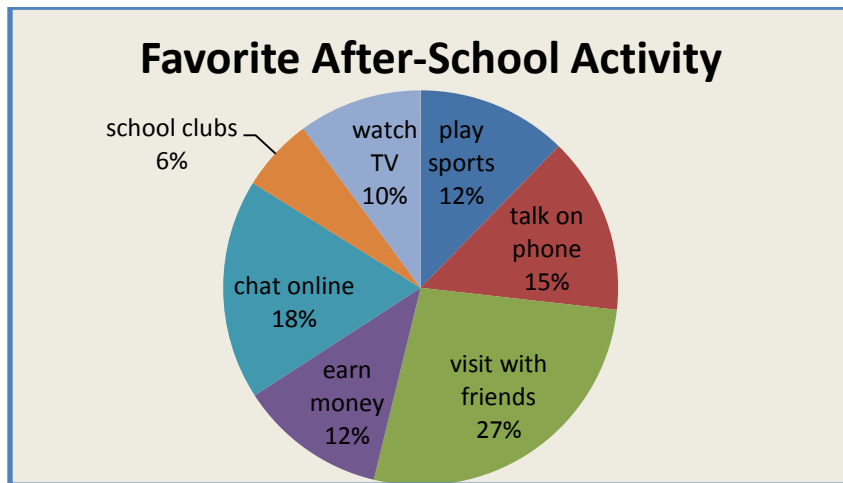


b) A pie chart would be appropriate because the categories add up to the "total" and each student chose exactly one activity. Each sector of the pie chart would represent one specific activity.

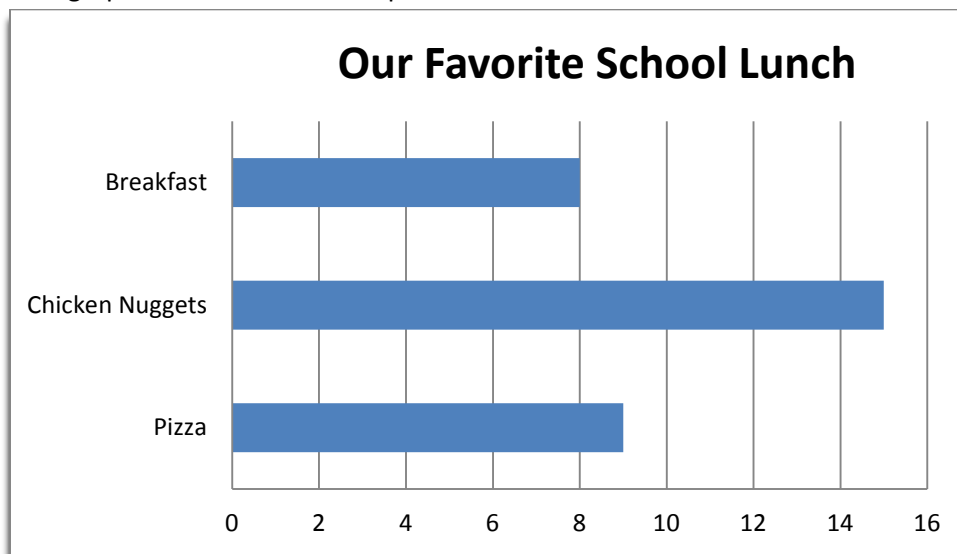
c)

Activity	Number of Students	Percent of total	Central Angle
play sports	45	12.30%	44.3°
talk on phone	53	14.48%	52.1°
visit with friends	99	27.05%	97.4°
earn money	44	12.02%	43.3°
chat online	66	18.03%	64.9°
school clubs	22	6.01%	21.6°
watch TV	37	10.11%	36.4°
TOTALS	366	100.00%	360°

d)



- 2) Answers will vary. During 2007, the majority of bears in North America were black bears. According to this graph, there were 800,000 black bears in North America. The brown bear population was only around 60,000 and there were fewer than 20,000 polar bears.
- 3) a) – c) Answers will be different for every class based upon the Type of Pet survey results.
- 4) a) – c) Answers will be different for every class based upon the Favorite Season survey results.
- 5) a) This graph is misleading because the pictures are different sizes and because they are not lined up. The chicken nuggets are on two lines, so it isn't very clear at first glance that they were a much more popular choice.  
b) A bar graph would be a better representation.



- 6) a) – c) Answers will be different for every class based on the Favorite Food survey results.

### Review Exercises

- 7) a) The individuals are the players.  
b) # = jersey number; POS = position; Player = Player, GP = games played; G = goals; A = assists; P = points; +/- = plus/minus; PIM = penalty minutes; PP = power play goals; SH = shorthanded goals; GW = game winning goals; S = shots; S% = shooting percentage  
c) Position is a categorical variable, player # is neither, and the rest are numerical.

8)  $2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 = 2^5 \cdot 4^5 = 32,768$  possible ways to answer all the questions.

9)  $P(\text{all correct}) = \frac{1}{32,768} = 0.0000305 = 0.00305\%$

**Section 5.2**

1)

	Mean	Median	Mode	range
a)	34.33	29	None	50
b)	44.57	48	22	54
c)	62.8	62	None	100

2) The total weight of the five men is  $167.2 \cdot 5 = 836$  pounds.  $836 - 158.4 - 162.8 - 165 - 178.2 = 171.6$  pounds.

3)

- a)  $5.1 \cdot 12 = 61.2$  feet
- b)  $4.8 \cdot 8 = 38.4$  feet
- c)  $(61.2 + 38.4) / 20 = 4.98$  feet

4) The mean = 31, the median = 32, and there was no mode.

Either the mean or median would be an appropriate measure of central tendency. For example, the median number of advertisements received by each family during the month was 32.

5)

- a) Mean = 63.38, med = 70.5, mode = none, range = 72. If D's start at 63%, Mica will receive a D based on his mean grade.
- b) Mean = 70.88, median = 70.5, mode = 70, range = 24. If C-'s start at 70%, Mica will receive a C- based on his mean grade. Note that the mean, mode, and range all changed but the median did not.
- c) His mean will be 70.3, so he will receive a C-.
- d) His mean will be 74.2, so he will have a C assuming that the C's start at 73%.

6) a) and b) from Table

Car Model	Retail Price	Dealer's Cost	Amt. of Mark-Up	% of Mark-Up
Nissan Sentra	\$24,500	\$18,750	\$5,750	30.7%
Ford Fusion	\$26,450	\$21,300	\$5,150	24.2%
Hyundai Elantra	\$22,660	\$19,900	\$2,760	13.9%
Chevrolet Malibu	\$25,200	\$22,100	\$3,100	14.0%
Pontiac Sunfire	\$16,725	\$14,225	\$2,500	17.6%
Mazda 5	\$27,600	\$22,150	\$5,450	24.6%
Toyota Corolla	\$14,280	\$13,000	\$1,280	9.8%
Honda Accord	\$28,500	\$25,370	\$3,130	12.3%
Volkswagen Jetta	\$29,700	\$27,350	\$2,350	8.6%
Subaru Outback	\$32,450	\$28,775	\$3,675	12.8%

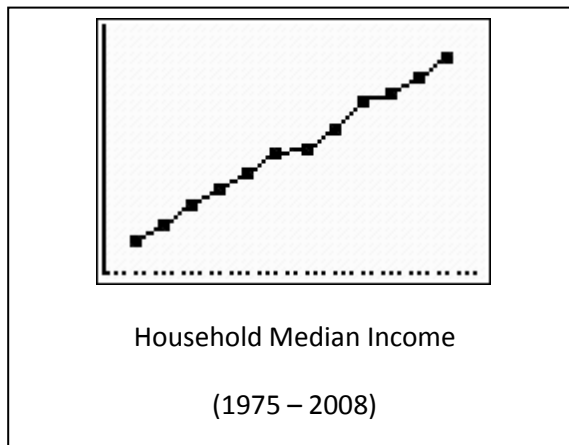
- c) Mean = 16.9% ; median = 14.0%; mode = none; range = 22.1%
- d) No. The order of the first three results is the same for both columns, but the rest were in a different order. For example, The Subaru Outback was in fourth place based on amount of mark-up but in seventh place based on percent of mark-up.

7) Answers will vary. A reasonable answer might be as follows:

The price for platinum has generally been increasing from 1960 to 2005. The price for an ounce of platinum in 2005 is approximately nine times as much as the price was in 1960. It remained fairly constant during the 60's and the 90's. There were several spikes in price, most notably around 1980. It then dropped significantly from 1980 until 1982.

8) Screenshot from a TI-84+

a)



b) Answers will vary. A reasonable answer might be as follows:

The graph shows that the median household income in the U.S. has been increasing at a fairly constant rate from 1975 to 2008. The median income in 2008 is more than four times that of the median household income in 1975. There are no significant spikes or drops in median household income during these 34 years.

## Review Exercises

- 9) This is a permutation because of the “specific horse” assignments. There are  ${}_{10}P_5 = 30,240$  ways to assign the first shift.
- 10) This is a combination because they all going to the same activity. There are  ${}_{10}C_4 = 210$  ways to choose a group of 4 campers to go on to the archery class.
- 11) This is a combination because the order you say the toppings doesn't matter – you get the same 3 topping pizza. There are  ${}_{12}C_3 = 220$  different pizzas possible.
- 12) This is the Fundamental Counting Principle because there are three different categories. There are  $3 \cdot 7 \cdot 8 = 168$  outfits possible.
- 13) This is a permutation because the skiers will all be getting different prizes. There are  ${}_{11}P_3 = 990$  ways to award the prizes.

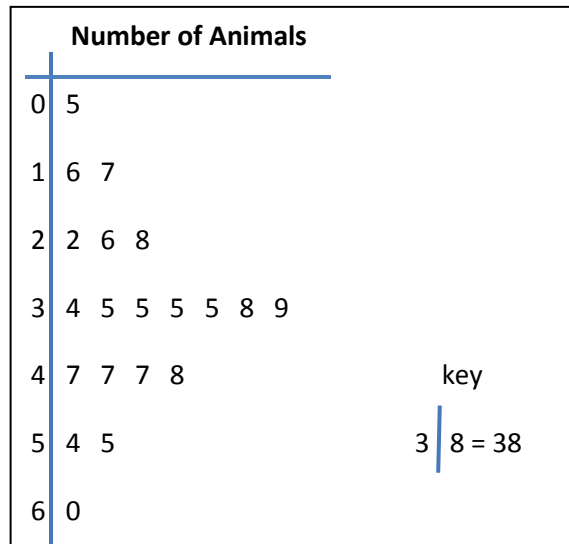
**Section 5.3**

1)

- a)  $\bar{x} = 65$ ; Med = 70; mode = 70; range = 64
- b) The percentage of the paper packaging used in a country that is recycled ranges from 34% in Estonia to 98% in Japan. The distribution is roughly symmetrical, but is very spread out. There appear to be a few low outliers and Japan seems to be a high outlier at 98%. The majority of countries included in this data recycle between 56% and 84% of the paper packaging. Seventy percent is both the mode and median for the percent of paper packaging that is recycled.

2)

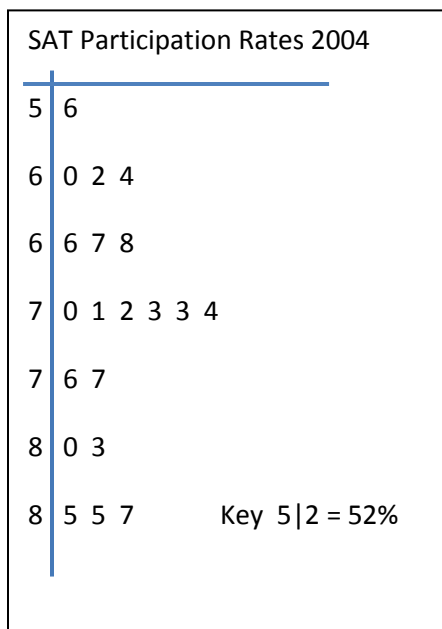
a)



- b) The number of animals treated by the local veterinarian school, over the given 20 day period, ranged from 5 to 60 animals per day. The distribution is quite symmetrical and does not include any outliers. The median number of animals treated per day was 35.

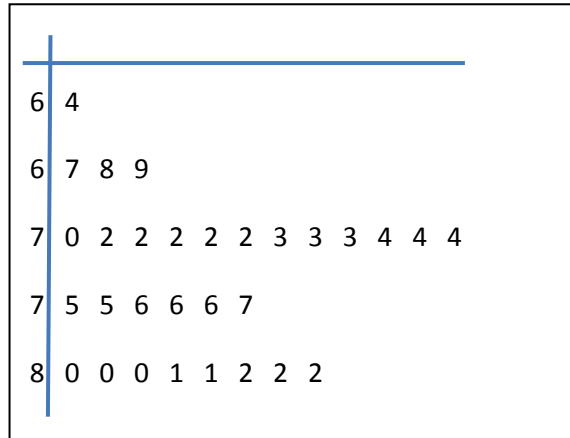
3)

a)



- b) The median is 72.5% of students took the SAT in 2004 for these states.
- c) These are the 20 states (19 states and D.C.) with the highest rates of students participating in the SAT. Therefore, if we included the SAT participation rates from the other 31 states, our mean and median would both be lower.
- d) Among the 20 states with the highest rates of SAT participation, the percent of students who took the SAT during 2004 ranged from 56% in Oregon to 87% in New York. The median percentage of students taking the SAT among these 20 states was 72.5%. The distribution is roughly symmetrical with no outliers. Note that the nine states with the highest SAT participation are all in the northeast portion of the United States.

4) a)



- b) The data appears to be skewed to the left. There is a high concentration of values between 70 and 74.
- c) Answers will vary. A possible answer is that this might represent the ages of the members of the Lion's club.

5)

- a) Mean = 75.62; Med = 77; Mode = 92
- b) The mode would not be appropriate by itself because it is not typical of the majority of ratings. Either the mean or the median would be a better choice. However, because the data set is slightly skewed to the left, the median would probably be a better choice. It makes it clear that half of the game critics gave a score above 77 and half gave a score below 77.

6)

- a)
  - i) This graph is skewed left with a few possible low outliers.
  - ii) This graph is roughly symmetrical with one clear low outlier.
  - iii) This graph symmetrical and bimodal with no clear outliers.
  - iv) This graph is skewed to the right with no clear outliers.
- b) Mean or median
  - i) The median will be greater because the mean is pulled toward the skew and low outliers.
  - ii) The median will be greater because the mean is pulled toward the low outlier.
  - iii) The mean and median would be similar. They would both be where very few actual data values are.
  - iv) The mean will be greater because the mean is pulled toward the direction of the skew.
- c) Answers will vary. Could be ages, costs, test scores, # miles on a car, # songs on cell phones, etc.

7)

a)

Number of Games Played	
1	6
2	6 6 9
3	
4	
5	1 4
6	0 0 1 8
7	0 1 3 4 5 6 7
8	1 2 2 2 2 2

b) Mean = 63.39; med = 71; mode = 82; range = 66

c) The number of games played by these Minnesota Wild players during the 2015-16 regular season ranges from 16 to 82 games. There are four possible low outliers – those below 30 games being played during the season. The median number of games played was 71 and the majority of the players played at least 51 games. The distribution is skewed to the left.

8)

a) +/- statistic. From the website: [http://proicetohockey.about.com/od/scoresandstat1/f/plus\\_minus.htm](http://proicetohockey.about.com/od/scoresandstat1/f/plus_minus.htm)

**Question:** What is the "plus-minus" statistic and how is it calculated?

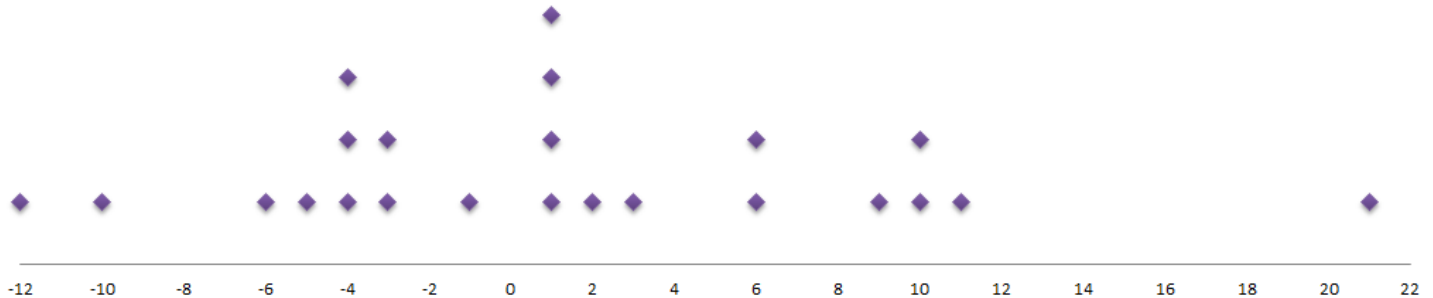
**Answer:** When an even-strength or shorthanded goal is scored, every player on the ice for the team scoring the goal is credited with a "plus." Every player on the ice for the team scored against gets a "minus."

A player's overall total is calculated by subtracting the minuses from the pluses. A high plus total is taken to mean that a guy is a good defensive player. But it's a very broad measurement and there has always been disagreement over how useful a statistic it is.

- Power play goals are not used in calculating plus-minus.
- Shorthanded goals are used in calculating plus-minus.
- Penalty shot goals are not used in calculating plus-minus.

b)

Plus/Minus Statistic  
2015-16 Minnesota Wild (23 players)



- c) The distribution of the plus-minus statistic for these 2015-16 Minnesota Wild players varies greatly and is roughly symmetrical. The median and mode are both +1 point. The plus-minus statistic ranges from -12 for Mikael Granlund to +21 for Erik Haula. Haula's statistic appears to be a possible high outlier among this data.

### Review Exercises

9)

- Population:** All Springfield residents
- Parameter:** The true percent of all Springfield residents who enjoy watching *The Simpsons*.
- Sample:** 1245 Springfield residents surveyed
- Statistic:**  $\hat{p} = \frac{1002}{1245} = 0.805 = 80.5\%$
- Margin of Error:**  $m. e. = \pm \frac{1}{\sqrt{1245}} = \pm 0.028 = \pm 2.8\%$
- Estimated 95% CI:**  $0.805 - 0.028 = 0.777$  and  $0.805 + 0.028 = 0.833$   
The estimated 95% confidence interval is from 0.777 to 0.833 or 77.7% to 83.3%.
- Confidence Statement:** We are 95% confident that the true percent of Springfield residents who enjoy watching *The Simpsons* is between 77.7% and 83.3%.

### Section 5.4

1)

- Approximately 450 employees make \$77,000 or more.
- The bin width is \$11,000.
- Answers will vary but a common answer might be 'A typical employee of this district makes approximately \$50,000.'



2)

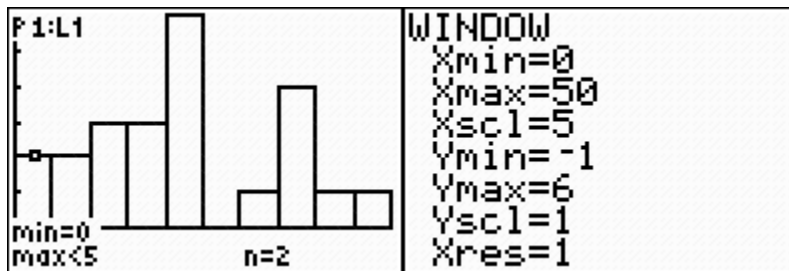
- a) Jessica's weight has ranged from 131 pounds to 149 pounds during this time period. Her median weight was approximately 140 pounds. The graph shows a symmetrical distribution with no outliers.
- b) The range = 2 representing weights from  $137 \leq w < 139$  pounds.
- c) She has been within 5 pound of 140 pounds for 23 of her 30 weight checks or 76.7% of the time.  $23/30 = 0.76667$

3)

- a) The 'over 75' bin is not the same width as the other bins. The first bin should run from 15 to 24 for consistency. All other bins cover a range of ten years.
- b) According to the graph, the percentage of men who spend at least one hour per week exercising decreases with age. The histogram is clearly skewed to the right and does not appear to have any outliers. Over 60% of the men between 16 and 24 years of age exercise at least one hour per week, but fewer than 5% of the men over 75 do. The biggest drop in time spent exercising happened between the 16 – 24 age group and the 25 – 34 age group.

4)

- a) Note that histograms may vary as different students may have different bin widths. The graph below was produced on a TI-84 graphing calculator.



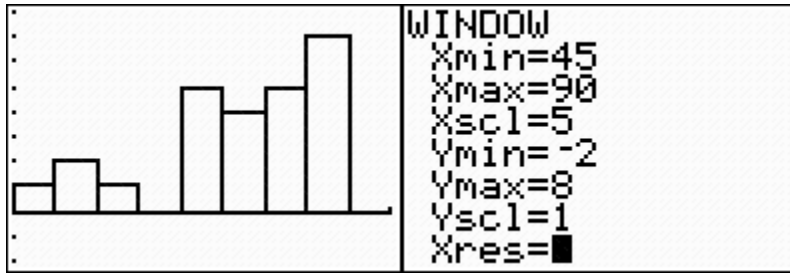
- b) The number of penalty minutes served by these Minnesota Wild players is bimodal, with the majority of these players having less than 25 penalty minutes. The PIM statistic ranged from 0 minutes to 48 minutes for these Wild players, with no outliers. The median and mode are both 20 minutes and the mean is 21.65 minutes.

5)

- a) Answers will vary.
- b) Answers will vary
- c) Answers will vary
- d) Answers will vary
- e) Answers will vary

6)

a) This histogram is a screenshot from a TI-84 graphing calculator.



b) The median will be higher than the mean because the graph is skewed to the left.

c) Range = 35 years; Mean = 71.52 years; Med = 74 years; Mode = 82 years

d) In this case, because the distribution is skewed left, the median is the most appropriate measure of central tendency. The mean is pulled toward the low end by the four countries with very low life expectancies. In addition, the mode is a higher number than most countries have, so it is misleading as well. The median tells us that half of these countries have life expectancies above 74 years and half are below 74 years.

7)

a)

Prize Amount	\$1000	\$250	\$50	\$25	\$0
Probability	1/500	2/500	5/500	10/500	482/500

b) The expected value of a ticket is  $(1/500)1000 + (2/500)250 + (5/500)50 + (10/500)25 + (482/500)0 = \$4$ . Remember, each ticket cost \$10 to purchase so the average ticket is worth  $\$4 - \$10 = -\$6$ . The average ticket will lose \$6.

c) This is not a fair game because the cost to play the game does not match the expected winnings for a ticket. This is quite common in raffles as they are often used as fundraisers.

8)

a)  $P(2 \text{ turquoise}) = \frac{7}{16} \cdot \frac{7}{16} = \frac{49}{256} \approx 0.1914$

b) Simon could get one gold fish and then one of another color or a fish of a different color followed by a gold fish.  $P(\text{exactly one gold}) = \frac{4}{16} \cdot \frac{12}{16} + \frac{12}{16} \cdot \frac{4}{16} = \frac{96}{256} = 0.375$

c)  $P(\text{pink, then gold}) = \frac{5}{16} \cdot \frac{4}{16} = \frac{20}{256} \approx 0.078$

9)

a)  $P(2 \text{ pink}) = \frac{5}{16} \cdot \frac{4}{15} = \frac{20}{240} \approx 0.083$

b) Simon could get one turquoise fish and then one of another color or a fish of a different color followed by a turquoise fish.  $P(\text{exactly one turquoise}) = \frac{7}{16} \cdot \frac{9}{15} + \frac{9}{16} \cdot \frac{7}{15} = \frac{126}{240} = 0.525$  A second way to view this problem would be to use counting methods. There are  ${}_{16}C_2 = 120$  ways to select two fish. There are  ${}_9C_1 = 9$  ways to select a turquoise fish and  ${}_7C_1 = 7$  ways to select a non-turquoise fish. This gives us

$$\frac{9 \cdot 7}{120} = \frac{63}{120} = \frac{21}{40} = 0.525$$

c)  $P(\text{no gold}) = \frac{12}{16} \cdot \frac{11}{15} = \frac{132}{240} = 0.55$

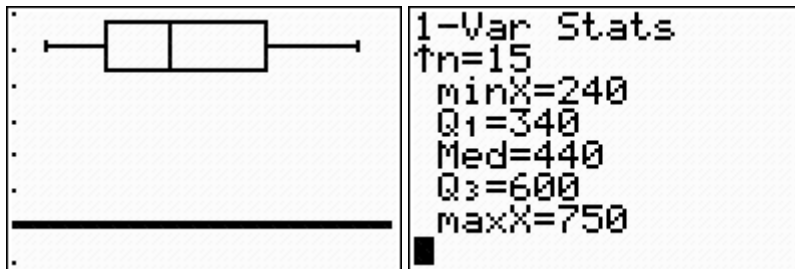
### Section 5.5

1)

- a) Calories from fat: Median = 210 cal; IQR = 130 cal  
b) Cholesterol: Median = 85 mg; IQR = 55 mg

2)

a) 5# sum = { 240, 340, 440, 600, 750} The box plot was generated on a TI-84 graphing calculator.



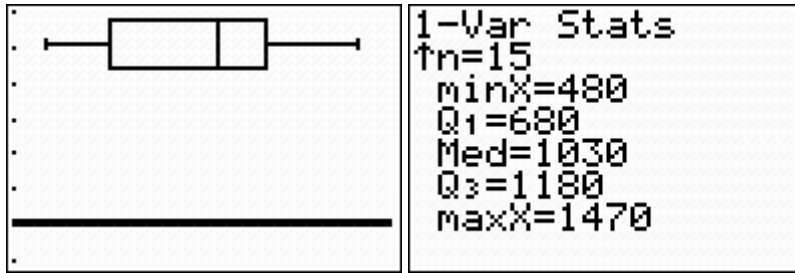
b)

Begin by finding the IQR which is  $600 - 340 = 260$ . For low outliers we have  $340 - 1.5 \cdot (260) = -50$ . There are no low outliers. For high outliers we have  $600 + 1.5 \cdot (260) = 990$  so there are no high outliers either.

c) The calories in these McDonald's sandwiches range from 240 to 750 calories. The median number of calories is 440 and the mean is 474.67 calories for these sandwiches. There are no outliers and the box plot suggests a slight skew to the right.

3)

a) Box plot



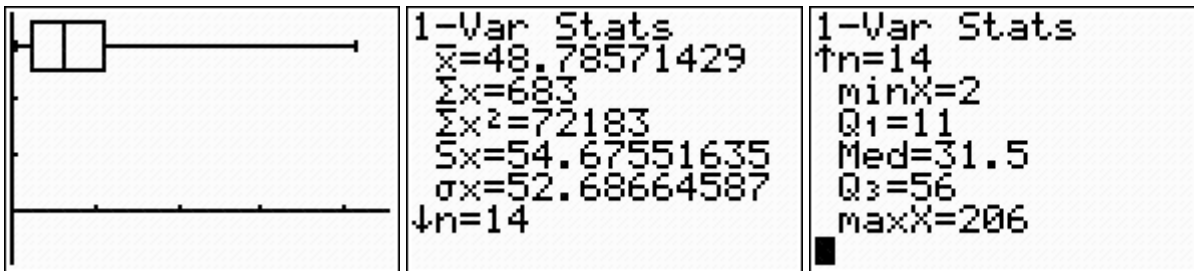
b) Median = 1030 mg; IQR = 500 mg

c) Mean = 992 mg ; and standard deviation = 291.43 mg

d) We saw that the cut off for a high outlier from Example 4 was 1,930 so the Angus Bacon Cheeseburger would be an outlier for sodium content.

4)

a) Five-number summary = {2, 11, 31.5, 56, 206}. The box plot is from a TI-84 graphing calculator.



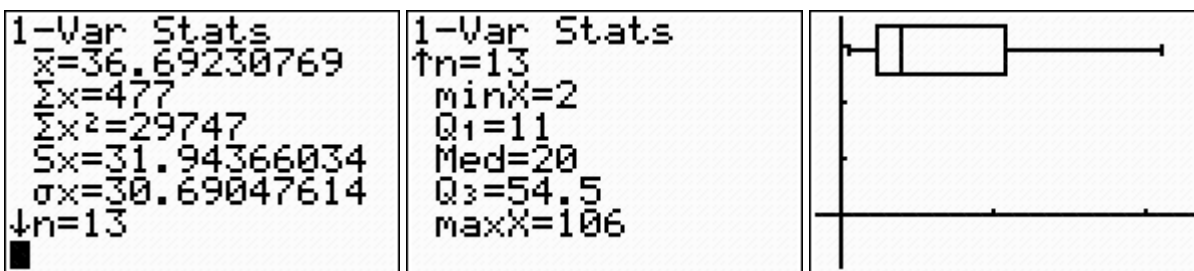
b) IQR = 56 – 11 = 45

For low outliers we have  $11 - 1.5 \cdot (45) = -56.5$  so there are no low outliers.

For high outliers we have  $56 + 1.5 \cdot (45) = 123.5$  so there is one high outlier of 206, the aluminum cans.

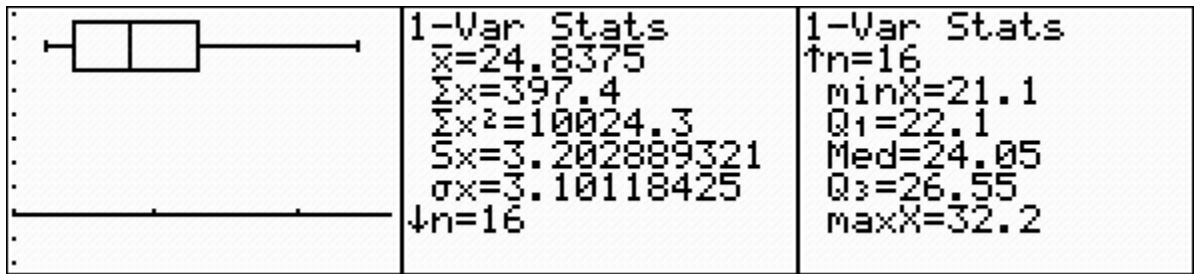
c) The mean = 48.7857 and the standard deviation = 54.6755. The mean is larger than the median because of the outlier.

d) The only outlier was 206 from aluminum cans so we will remove this from the data set. The new five-number summary is now {2, 11, 20, 54.5, 106}. The median, Q3, and max all changed. The mean and standard deviation have also changed. The mean is now 36.69 and the standard deviation is 31.94. A box plot from a TI-84 graphing calculator is shown below.



5)

a) The 5-number summary for the data is = {21.1, 22.1, 24.05, 26.55, 32.2}. The box plot is from a TI-84.

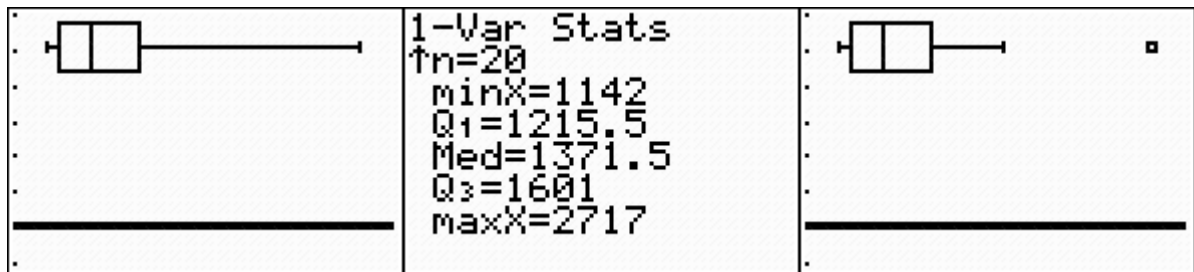


b) The 'box' shows that the middle 50% of the mean travel times to work are between 22.1 minutes and 26.55 minutes for people in these cities.

c) The mean number of minutes it takes for workers in these cities to get to work ranges from 21.1 minutes in Roseville to 32.2 minutes in Albertville. The distribution for travel times is skewed to the right, but does not include any times that are outliers. The median travel time for these workers is 24.05 minutes.

6)

a) The five-number summary is {1142, 1215.5, 1371.5, 1601, 2717}. The box plot was generated on TI-84. Note that the second box plot treated the outlier differently than the rest of the data.



b) The IQR = 385.5. For low outliers we have  $1215.5 - 1.5(385.5) = 647.25$  so there are no low outliers. For high outliers we have  $1601 + 1.5(385.5) = 2179.5$  so there is one high outlier, the Burj Khalifa in Dubai at 2717 feet.

c) The heights of these tall buildings range from 1142 feet to 2717 feet tall. The distribution is skewed to the right. The Burj Khalifa is the tallest building in the world and is over 700 feet taller than the next tallest building. Among these buildings it is a high outlier. Without this value, the rest of the heights form a much more symmetrical distribution. The median height of the buildings is 1371.5 feet tall, while the mean is 1470.25 feet tall.

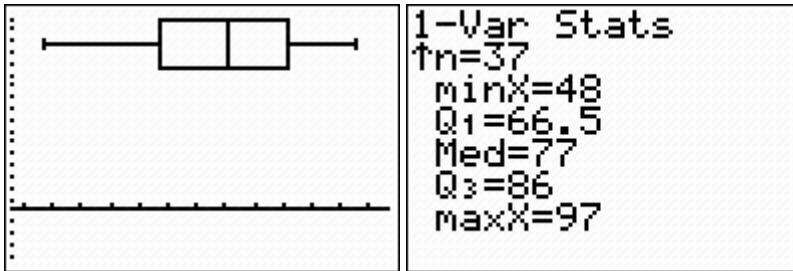
d) The middle 50% of these buildings are between 1215.5 feet tall and 1601 feet tall.

e) The five number summary is {54, 76, 88, 103, 163} so the range = 109 floors and the IQR = 27 floors.

f) For low outliers we have  $76 - 1.5(27) = 35.5$  so there are no low outliers. For high outliers we have  $103 + 1.5(27) = 143.5$  so there is one high outlier and it is once again Burj Khalifa in Dubai.

7)

- a) The five-number summary is {48, 66.5, 77, 86, 97}.
- b) The box plot from a TI-84 is shown below.

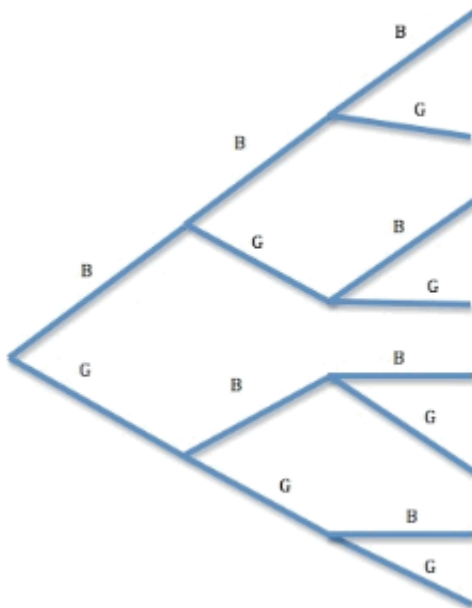


- c) The median critics' rating for the Wow So Fit game was 77 on a scale of 1 to 100, with 100 being the highest rating. The ratings for this game range from 48 to 97. The distribution is slightly skewed left with no outliers.
- d) The "box" part tells us that the middle 50% of critics gave the Wow So Fit Game a rating between 66.5 and 86 on a scale of 1 to 100.

8)

- a) Bias – The person is claiming that the ratings are too high due to the critics receiving a free game.
- b) Validity – The person is claiming that the rating system does not make sense as a way to measure the games because of the lack of guidelines.
- c) Reliability – The person is claiming that the rating system cannot be relied upon to give consistent results.

9)  $S = \{BBB, BBG, BGB, BGG, GBB, GBG, GGB, GGG\}$



10)

- a)  $P(B, G, B) = 0.5 \cdot 0.5 \cdot 0.5 = 0.125$
- b)  $P(2 \text{ Girls}) = P(GGB) + P(GBG) + P(BGG) = 3 \cdot 0.5 \cdot 0.5 \cdot 0.5 = 0.375$
- c)  $P(\text{At Least 1 Boy}) = 1 - P(\text{No Boys}) = 1 - 0.5 \cdot 0.5 \cdot 0.5 = 1 - 0.125 = 0.875.$

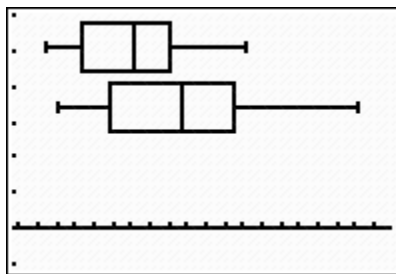
**Section 5.6**

1)

- a) The 5-number summary for Daily Total Fat % is {12, 22, 36, 45, 66} while the 5-number summary for Daily Saturated Fat % is {15, 29, 49, 63, 96}.

<pre>1-Var Stats ↑n=15 minX=12 Q1=22 Med=36 Q3=45 maxX=66</pre>	<pre>1-Var Stats ↑n=15 minX=15 Q1=29 Med=49 Q3=63 maxX=96</pre>
---	---

- b) The box plots below are from a TI-84. The top plot is for Daily Total Fat %.



- c) The median percent of daily value for saturated fat for these McDonald's sandwiches is 13% higher than that of total fat. The median for the saturated fat data is higher than the third quartile for total fat. The saturated fat percentages are more spread out (15% to 96% of daily value) than the total fat percentages (12% to 66% of daily value). Neither distribution has any outliers.

2)

- a) For boys, the standard deviation = 3.796 inches, the range = 12 inches, and the IQR = 7 inches. For the girls, the standard deviation = 3.215 inches, the range = 11 in, and the IQR = 5 in.
- b) The girls' heights are less spread out than the boys' heights are. All three of the measures of spread from part a) are smaller for girls than they are for boys.
- c) For the boys, the mean = 69.133 inches, the median = 69 inches, and the modes = 67 and 73 inches. For the girls, the mean = 63.215 inches, the median = 63.5 inches, and the modes = 61, 64, and 66 inches.
- d) The median height for girls is 5.5 inches less than it is for boys and the mean height for girls is about 5.9 inches less than it is for boys.
- e) The shape of the distributions of heights is roughly symmetrical for both boys and girls. Neither group has any outliers.

3)

a)

Class #3		Class #4
	4	1
	9 6 6	1
	4 4 3 1	2
	8 6 5	2
	3 2 0	3
9 8 8 7 7 7 5 5 5	3	3
	2 2 1	4
	6 5	4
	1	5

**key 4|3 = 43**

- b) For Class 3, the five-number summary is {14, 24, 35, 38.5, 51}.  
For Class 4, the five-number summary is {20, 27, 31, 38, 46}.
- c) For Class 3, the mean = 32.03, the standard deviation = 9.77, the modes are 35 and 37, the range = 37, and the IQR = 14.5.  
For Class 4, the mean = 32.91, the standard deviation = 6.51, there are many modes (25, 27, 29, 30, 31, 35, 37), the range = 26, and the IQR is 11.
- d) Class 4 had more consistent scores on this common assessment because they only had a range of 26, compared with a range of 37 for Class 3. Class 3 had a higher median and maximum score, but it also had a lower minimum and quartile one score. Overall, Class 3 had students who scored very high as well as students who scored very low. Neither class had any scores that are considered to be outliers.

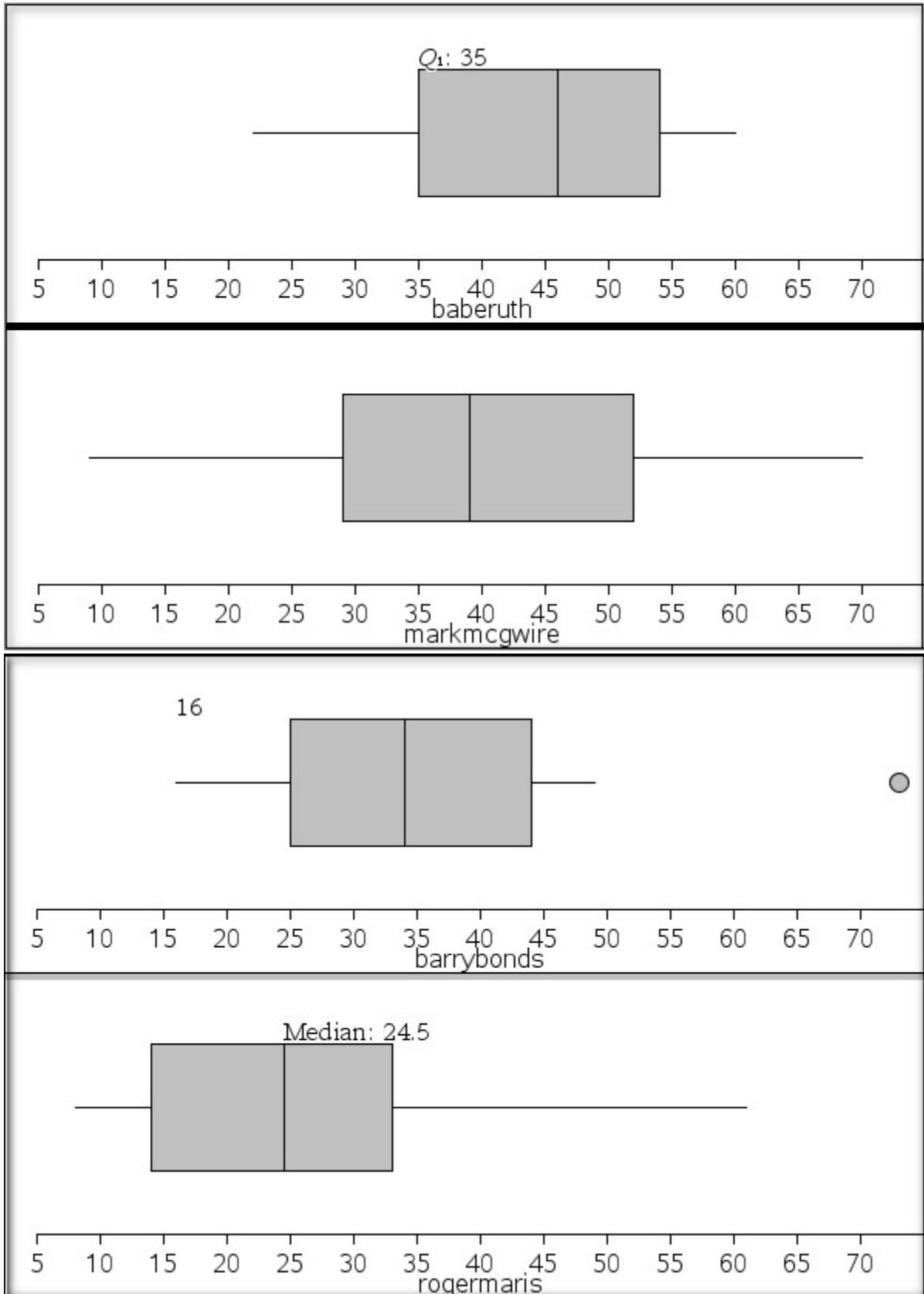
4)

a)

Player	Mean	St. Dev.	Range	IQR	Min	Q1	Med	Q3	Max
Babe Ruth	43.93	11.25	38	19	22	35	46	54	60
Mark McGwire	38.67	18.12	61	23	9	29	39	52	70
Barry Bonds	36.06	13.45	57	19	16	25	34	44	73
Roger Maris	26.1	15.61	53	19	8	14	24.5	33	61



b)



c) Outlier tests:	$Q1 - 1.5(IQR)$	and	$Q3 + 1.5(IQR)$
Babe Ruth:	$35 - 1.5(19) = 6.5$	and	$54 + 1.5(19) = 82.5$ ; no low or high outliers
Mark McGwire:	$29 - 1.5(23) = -5.5$	and	$52 + 1.5(23) = 86.5$ ; no low or high outliers
Barry Bonds:	$25 - 1.5(19) = -3.5$	and	$44 + 1.5(19) = 72.5$ ; no low, but one high outlier of 73
Roger Maris:	$14 - 1.5(19) = -14.5$	and	$33 + 1.5(19) = 61.5$ ; no low or high outliers

d) Babe Ruth was much more consistent as he had the smallest range. His median is higher than all three of the others. Also, his first quartile is higher than the median for Bonds and the third quartile for Maris. Even though Barry Bonds has the highest number, he is not necessarily the best home-run hitter among these players because his high number of home runs was an outlier for him. Also, his median is lower than both McGwire and Ruth. Roger Maris and Barry Bonds both have distributions that are skewed to the right. Mark McGwire has a fairly symmetrical graph and Babe Ruth's graph is the only one that is skewed to the left.

5) The small cars had the biggest range for gas mileage. Vans had the smallest range.  
 The vans had one low outlier and the small cars had two high outliers for gas mileage.  
 The Q1 for small cars is higher than the Q3 for all other types of cars. This means that 75% of small cars get better gas mileage than at least 75% of each of the other types of cars.  
 For the vans, the Q3 and maximum value are the same.  
 The compact and midsize graphs are fairly symmetrical.  
 The small and sporty graphs are both skewed to the right.  
 The large vehicles are skewed to the left.

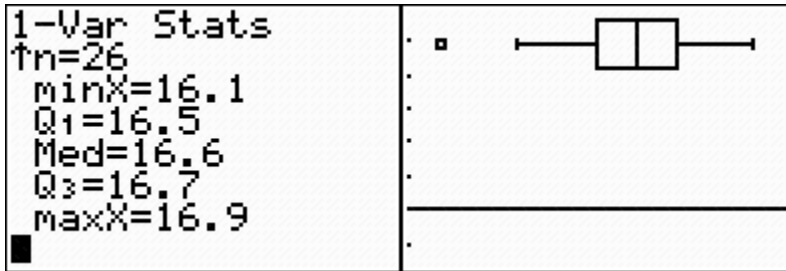
6)

- Graph I: symmetrical and bell-shaped  
 Graph II: symmetrical and bell-shaped  
 Graph III: symmetrical and bimodal  
 Graph IV: symmetrical and uniform
- The mean and median of all of these graphs is around 52.
- The mean, median, and range would be the same for all of these graphs.
- Graph I would have the smallest standard deviation because it is the least spread out and has the most observations near the mean. Graph III would have the largest standard deviation because it is the most spread out and has the most observations far from the mean.
- The mean and standard deviation would be most appropriate Graphs I and II because they are symmetrical with bell-shaped distributions. The five-number summary would be more appropriate for Graphs III and IV because they are not bell-shaped and have many observations far from the mean.

**Section 5.7 Review**

- C. The mean is pulled toward the direction of the skew but the median is not pulled.
- The mean = 16.592 ounces, the standard deviation = 0.165 ounces, the mode = 16.6 ounces, and the range = 0.8 ounces.

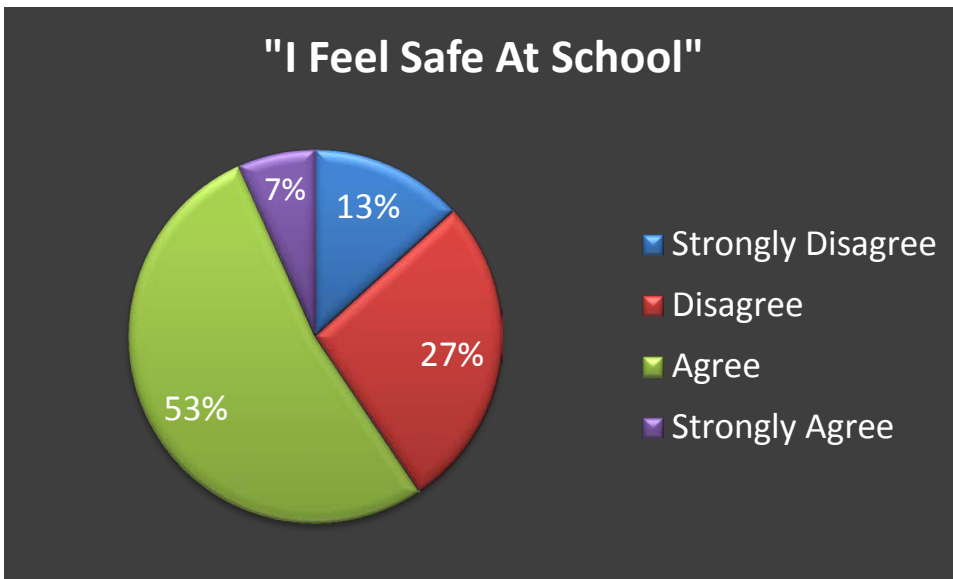
- b) The five-number summary = {16.1, 16.5, 16.6, 16.7, 16.9}. The box plot from a TI-84 graphing calculator is given below.



- c) The dot plot is more informative because you can see every actual value. The mode and the shape are both very clear in the dot plot.
- d) 88.5% of the packages weighed less than 16.75 ounces so he will lose money 11.5% of the time.

3)

- a) The pie chart below was created on Excel.

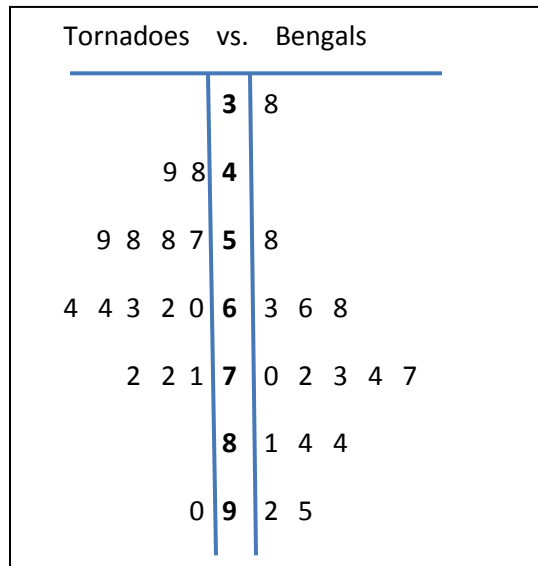


- b) Only 7% of the students surveyed said that they “*strongly agree*” that they feel safe at school. The most common reaction to the question among these students was “agree”, with 53% of the students responding this way. However, 2 out of every 5 of the students said that they “disagree” or “strongly disagree” with this statement.
- c) The committee should probably be concerned because 40% of the students do not feel safe at school. This is nearly half of the student body who are expressing concerns for their safety.

- 4) B)  
 5) D)  
 6) A)  
 7) C)  
 8) A

- 9) B
- 10) A
- 11) D
- 12) E
- 13) B
- 14) C
- 15) A
- 16) E
- 17) B
- 18)

a) Back-to-back stem plot



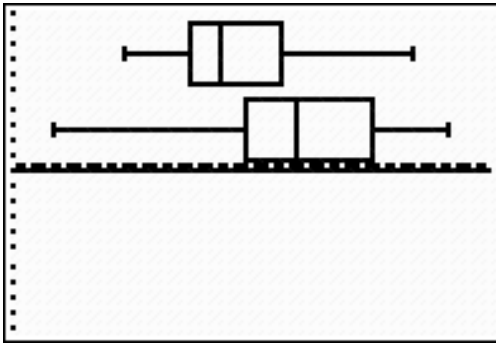
b) The five-number summary for the Tornadoes = {48, 58, 62, 71, 90}, the mean = 63.133 points and the standard deviation = 10.315 points.

<b>1-Var Stats</b> $\bar{x}=63.13333333$ $\Sigma x=947$ $\Sigma x^2=61277$ $Sx=10.3154992$ $\sigma x=9.965719018$ $\downarrow n=15$	<b>1-Var Stats</b> $n=15$ $\min X=48$ $Q_1=58$ $\text{Med}=62$ $Q_3=71$ $\max X=90$
---	---

The five-number summary for the Bengals is {38, 66, 73, 84, 95}, the mean = 73 points and the standard deviation = 14.147 points.

<pre> 1-Var Stats x̄=73 Σx=1095 Σx²=82737 Sx=14.14718548 σx=13.66747965 ↓n=15 </pre>	<pre> 1-Var Stats ↑n=15 minX=38 Q1=66 Med=73 Q3=84 maxX=95 </pre>
--	---

- c) The top box plot is for the Tornadoes and the bottom is for the Bengals.



- d) The Bengals generally scored more points than the Tornadoes did in these 15 basketball games. The median number of points for the Bengals is higher than the third quartile for the Tornadoes. Also, the mean number of points is 10 points higher for the Bengals. The Bengals have one low outlier while the Tornadoes have one high outlier. Even though the spreads are similar, the Bengals only had one game with less than 58 points, but the Tornadoes only had one game with more than 72 points. Based on only these points, the Bengals appear to be the better basketball team over this time period.
- e) These are only the points for one team. Did the teams win or lose these games? How many points did the opposing teams score? Were any players out due to illness or injury?

19)

- a) These graphs are time plots.
- b) The cellphone graph shows a steady rise in the number of cell phones per capita starting after 1985. It is increasing at a fast pace, but appears to be leveling off somewhat.

The landlines graph shows a rise in the number of landlines from 1985 to around 2000. After 2000 the number of landlines per capita has been dropping off at a fairly steep rate.

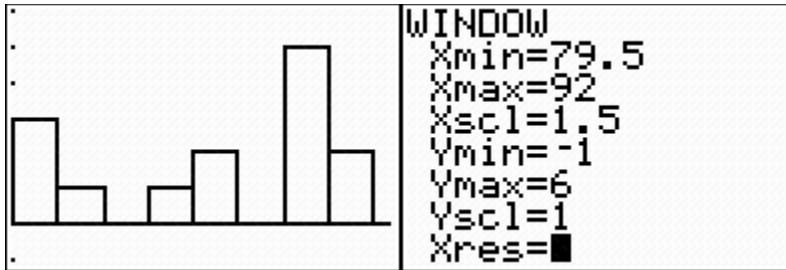
- c) Both were rising from 1985 to 2000. After 2000, the cellphones increased at a faster rate while the number of landlines declined. This shows that as cellphones have become more prevalent and many people no longer have landlines. The highest that landlines reached was 0.7 per capita, but the number of cell phones has surpassed 0.85 per capita and the trend shows that it is still increasing.

- d) The approximate number of cell phones per capita in 1997 was 0.2 while it was 0.75 in 2005. The trend suggests there will be about 1.15 cell phones per capita in 2018.
- e) The peak number of landlines per capita was about 0.7 in 2000. The model predicted about 0.25 landlines per capita in 2015.

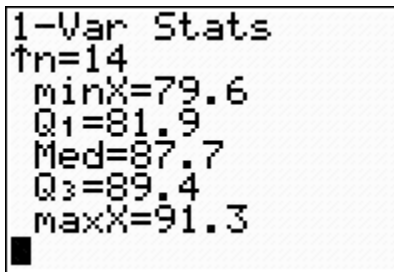
20) The histogram below was produced on a TI-84.

a) Histogram

Window used



b) The five-number summary is {79.6%, 81.9%, 87.7%, 89.4%, 91.3%}.



c) The IQR is  $89.4 - 81.9 = 7.5$  so we will use  $Q1 - 1.5(7.5)$  and  $Q3 + 1.5(7.5)$  for our outlier calculations.

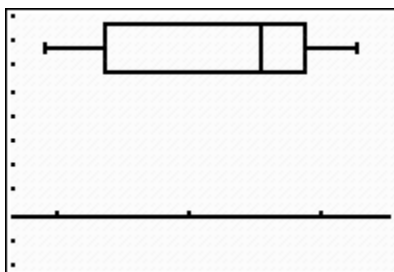
$$81.9 - 1.5(7.5) = 70.65$$

There are no low outliers.

$$89.4 + 1.5(7.5) = 100.65$$

There are no high outliers.

d) The box plot below was generated on a TI-84.



e) The range = 11.7%, the IQR = 7.5%, and the modes are 73% and 85%.

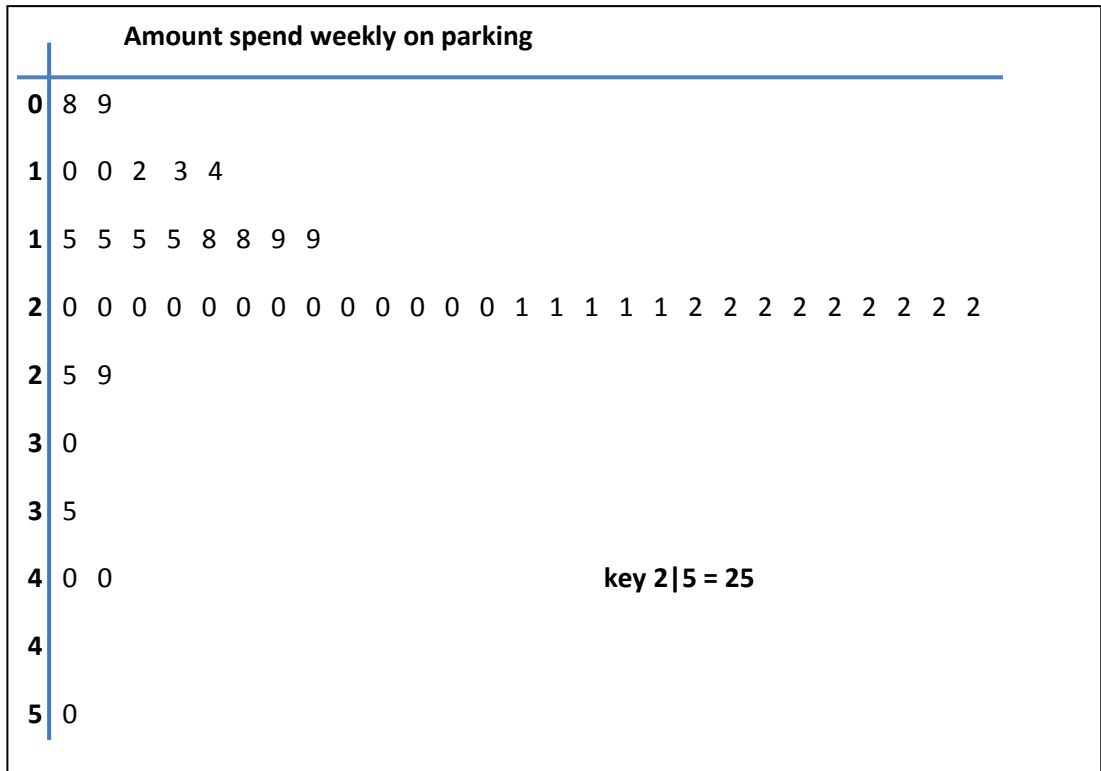
f) The mean = 86.27% and the standard deviation = 4.23%.

g) The mean is 1.4% less than the median.

- h) The five number summary would be more appropriate, because the graph is not symmetrical and bell-shaped. There are many values far from the center of the distribution which will have a strong effect on the standard deviation.
- i) The percent of people 25 years old and older who are high school graduates for these states ranges from 79.6% in Mississippi to 91.3% in Minnesota. The distribution for this 2010 data is not symmetrical, but doesn't contain any outliers. The histogram shows that these states form three different groupings. The median rate of high school graduates for these states is 86.27% for 2010 according to the U.S. Census website.
- j) According to this Census data, Minnesota has 91.3% of its residents over 25 years of age who are high school graduates. Minnesota is the highest among these states.

21)

a)



b) The 5-number summary is {8, 18, 20, 22, 50}.

c) Our IQR is  $22 - 18 = 4$ .

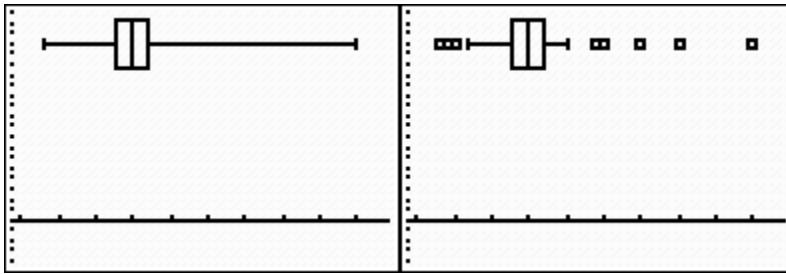
$$18 - 1.5(4) = 12$$

There are four low outliers of \$8, \$9, \$10, and \$10 dollars per week.

$$22 + 1.5(4) = 28$$

There are 6 high outliers of \$29, \$30, \$35, \$40, \$40, and \$50 dollars per week.

d) The box plot and box plot with outliers was created on a TI-84.



e) The range = \$42, the IQR = \$4, and the mode = \$20.

f) The mean = \$20.90 and the standard deviation = \$7.65.

g) The mean is just less than one dollar larger than the median. This may be due to the excess of outliers on the right side of the graph.

h) In this case, the five-number summary would be more appropriate. This is because the distribution has a large number of outliers and is spread out significantly from the mean.

i) The amount of money spent weekly on parking by this sample of 50 employees has a large spread. It ranges from \$8 to \$50 per week. Interestingly, the IQR is quite small. The middle 50% of these people spend between \$18 and \$22 dollars weekly on parking. There are ten outliers with four on the low end (\$8, \$9, \$10 & \$10) and six on the high end (\$29, \$30, \$35, \$40, \$40, & \$50). If the outliers are ignored, the graph is fairly symmetrical. The median and mode for the amount spent on parking is \$20 per week. The majority of employees spend less than \$22 per week.