# Unit 1 Answer Key

1.1 Homework Answer Key

1. The population is all 2022 Tesla Model S cars, the sample is the 5 specific model S cars that are being tested on.

2.
   a. The population of interest is 1584 SI students.
   b. The sample is the 150 students who were polled.

3.
   a. The researchers were trying to learn if the new surgical process would improve the outcomes for Alzheimer's patients. Likely questions motivating the research would be about the effectiveness of current treatments and the risks involved in letting Alzheimer's go untreated.
   b. It seems to have been conducted reasonably. Just because the study was too small to show statistically significant results does not mean it was too small – it could be that it was difficult to obtain experimental subjects in an ethical manner, or it could be that any effect actually was insignificant (so if there appeared to be that treated patients had an effect while non-treated ones steadily declined, that effect was not due to the surgery, rather it was potentially random or due to some other factor). The additional information I might want is how many patients were in the control group, and the actual results of the experimental group. (For example, what does "essentially held their own" mean? Does it mean that most of the eleven experience very slight to no decline, and a couple had a decline similar to, or worse than, the control group? Or did it mean something else?)

4.
   a. The researchers were trying to learn how effective a garlic supplement was at reducing the risk of getting a cold. Some questions that may have motivated their research could be "How effective are current strategies preventing colds?", "what is the cost of using a supplement to prevent a cold compared to the effectiveness of that supplement?", and "How effective is the use of a garlic supplement to prevent colds?"
   b. It seems like it may have been conducted reasonably, but you would need a lot more information to be sure. How many people were in the study and in each group? How were those people assigned to the groups? What were the actual proportions of people who did or did not get colds in each group, and was the difference between the results large or small? How did they control for things like the health and immune system of the patients and exposure to the colds?

5.
   a. Numerical, discrete.
   b. Numerical, continuous.
   c. Categorical.
   d. Numerical, continuous.
   e. Categorical.

6.
   a. The variable being tested is how many minutes of homework teachers believe that they are assigning. It is quantitative and continuous.
   b. The population of interest is the SI faculty (all 150 members). The sample is all of the members of the SI faculty.
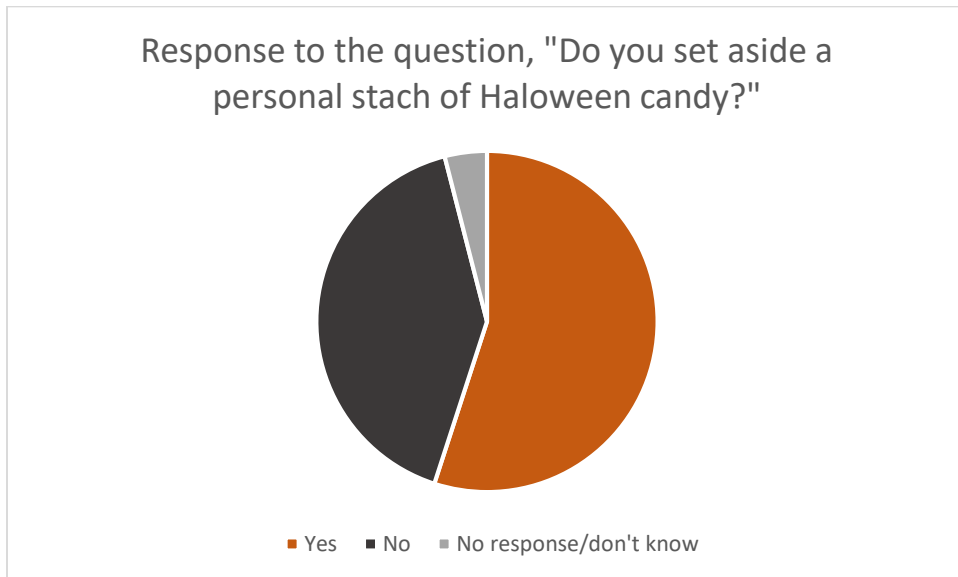   c. This sample is a census, because every member of the population is represented in the sample.
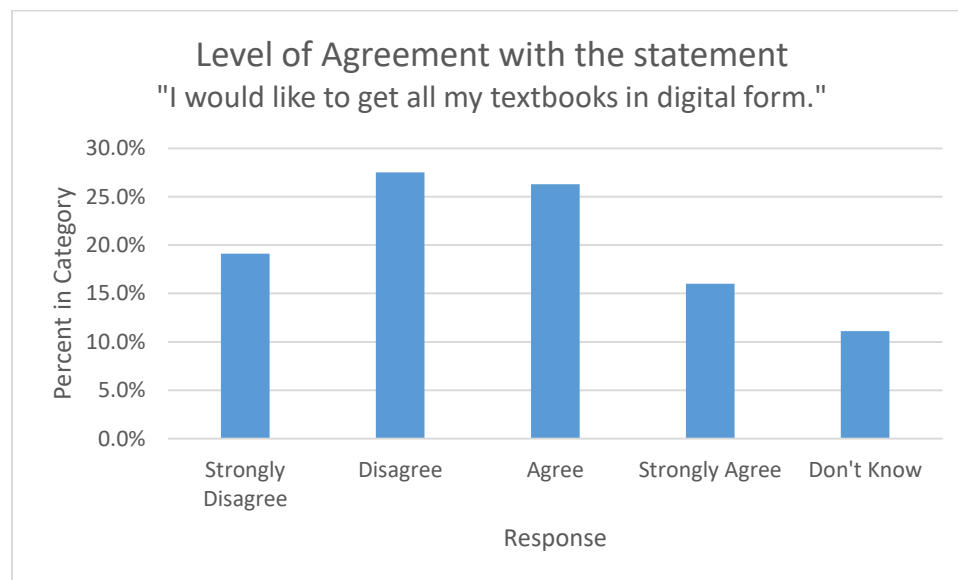
1.2 Homework Answer Key

1.
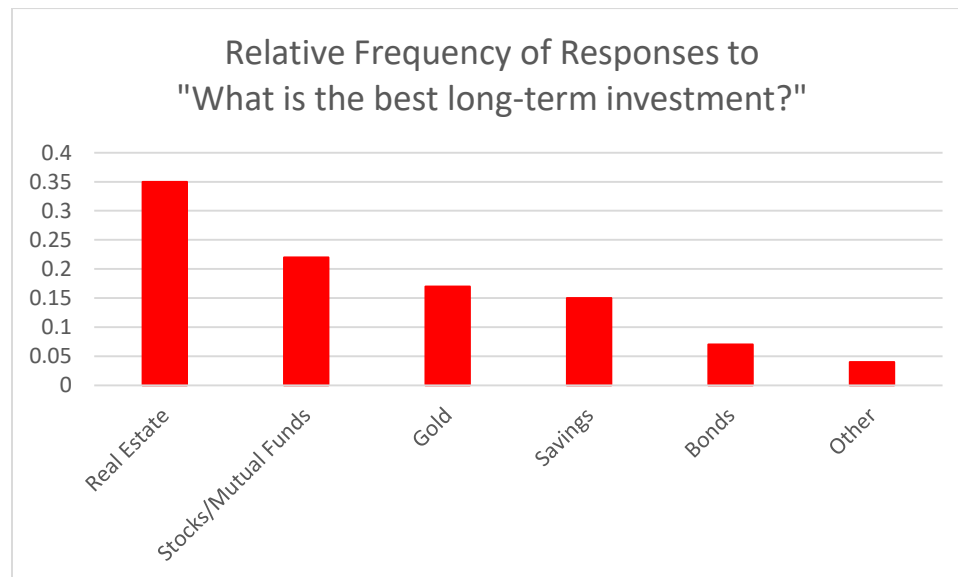
a.

b. More than one-third of all respondents answered that real estate was best. While this was not a majority (more than 50%), it was much more than any other response – more than 1.5 times greater than the next highest category. Gold and savings were close to the same in the response rate.

2.

Response to the question, "Do you set aside a personal stach of Haloween candy?"

Yes   No   No response/don't know

3.

Level of Agreement with the statement
"I would like to get all my textbooks in digital form."

Percent in Category

Strongly Disagree | Disagree | Agree | Strongly Agree | Don't Know
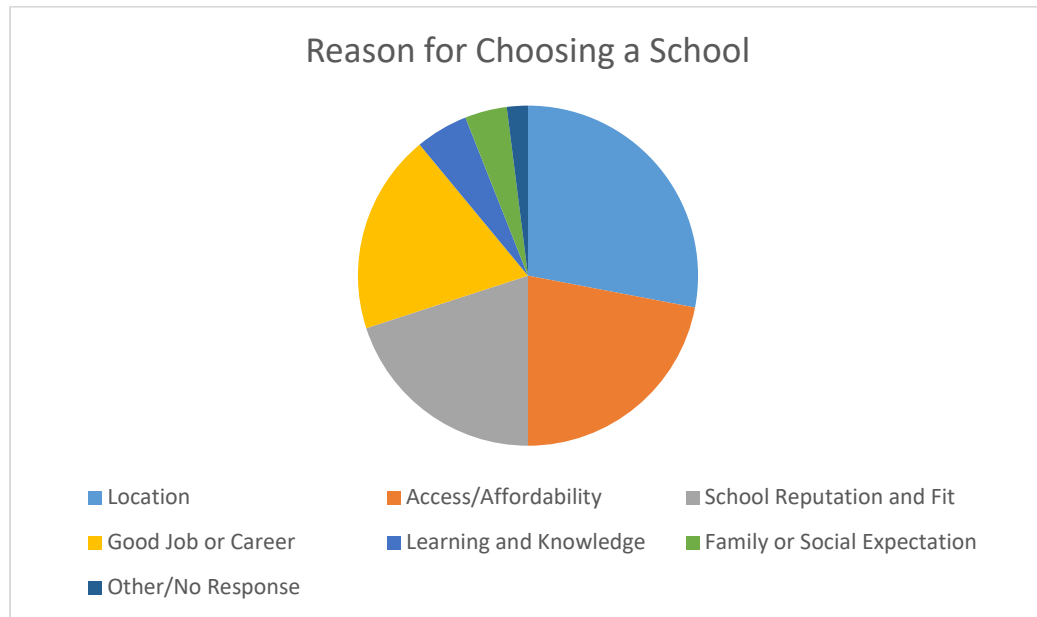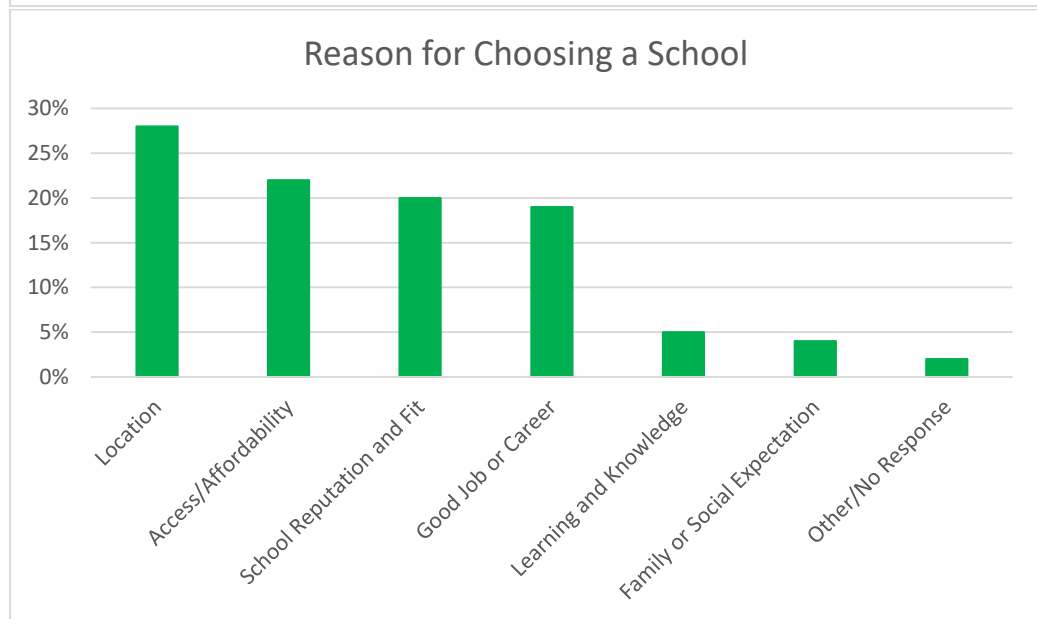
Response

a.
b. There are a lot of possibilities, but they should emphasize that there is only a slight majority who don't want digital textbooks and that the distribution of results is divided. An example is given below:

"Students Divided on Wanting Digital Textbooks"

4.

Reason for Choosing a School

- Location
- Good Job or Career
- Other/No Response
- Access/Affordability
- Learning and Knowledge
- School Reputation and Fit
- Family or Social Expectation

a.

Reason for Choosing a School

b.

c. The bar chart seems to be a better way of displaying the data – since several of the categories are similar in magnitude, it is difficult to detect this difference on the pie graph. In addition, there are a few very small values on the pie chart that are difficult to represent.

5.

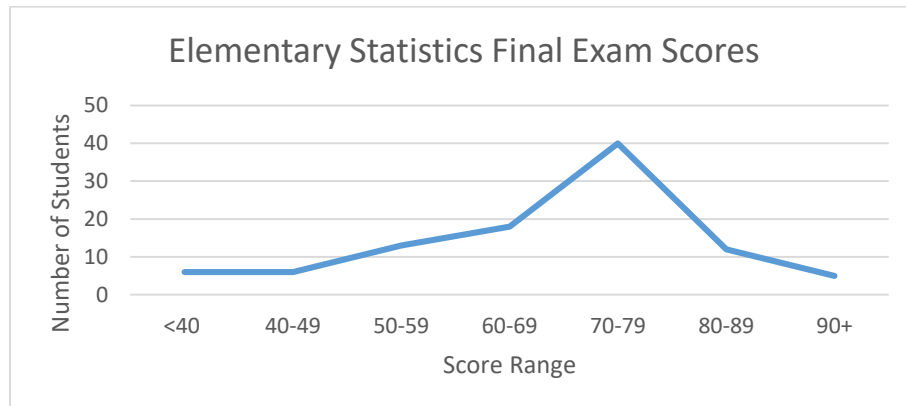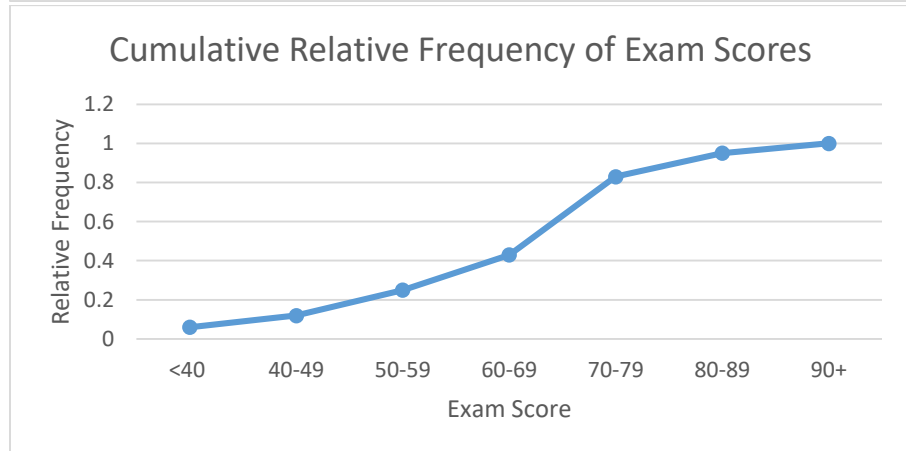Ticket Sales for the Top 5 Movies of 2021
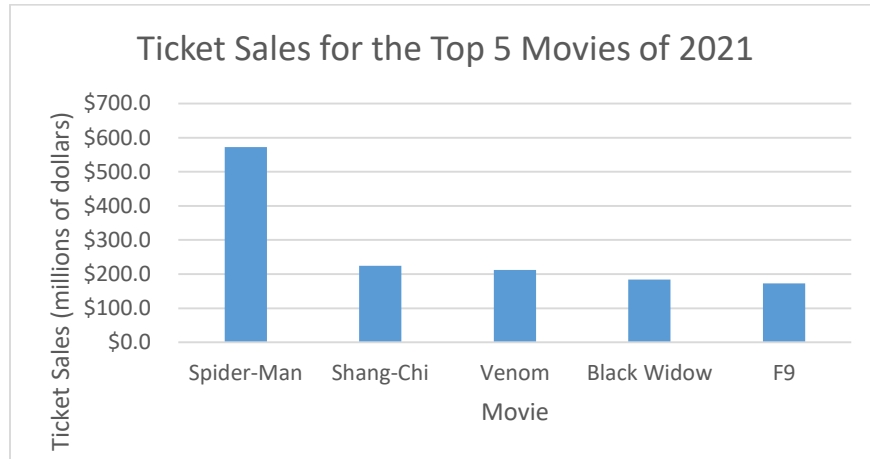
a.

b. The Spider-Man film made a lot more on ticket sales than the other films – which were all relatively similar to each other (in fact, if you have to add the next three films together and you will only exceed the ticket sales by $47.1 million in ticket sales) – so the Spider-Man movie did very well relative to the other films of 2021.

6.

a. 100 students   b. 18 students   c. 43 students   d. 0.43

Elementary Statistics Final Exam Scores

e.

Cumulative Relative Frequency of Exam Scores

f.

1.3 Homework Answer Key

1.
   a. Discrete
   b. Continuous
   c. Discrete
   d. Discrete
   e. Continuous

2.
   a. Brand and color of motorcycle are categorical.
   b. Weight of motorcycle is numerical and continuous.
   c. Price and number of motorcycles previously purchased are numerical and discrete.
   d. A bar chart, because it is used to display categorical data – dot plots are for numerical data.
   e. A histogram, because it is used to display numerical data – bar charts are for categorical data.

3.
   a. The response is qualitative, as it is a verbal (not numerical) response.
   b. This is categorical data, not numerical (qualitative, not quantitative).
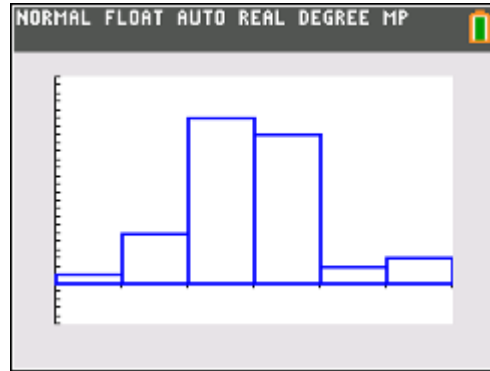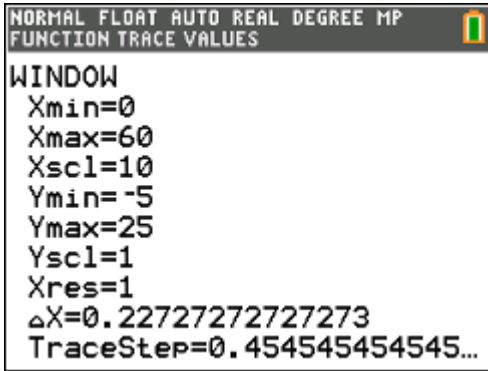   c. There must be a mistake because the categories total to 120%, not 100%.

4.
   a. **Key:** 2|2.8 = 22.8 cents

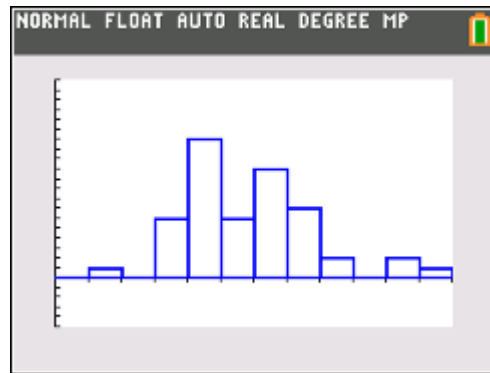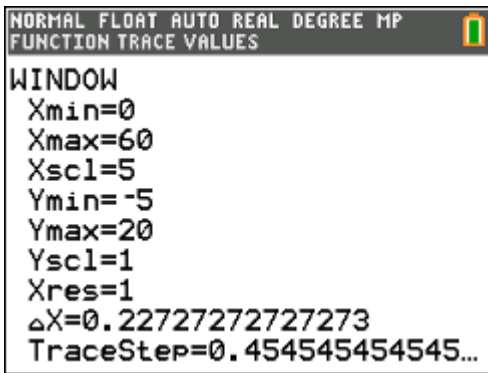| Stem	Leaves
0	9.0
1	6.0, 6.2, 7.4, 8.4, 8.9, 9.0
2	0.0, 0.0, 0.1, 2.8, 3.0, 3.0, 3.3, 3.8, 3.8, 4.0, 4.0, 4.0, 4.8, 5.0, 6.0, 6.3, 7.4, 7.9, 8.6
3	0.0, 0.0, 0.5, 0.6, 0.7, 1.4, 1.8, 2.8, 2.9, 3.0, 4.4, 5.1, 5.7, 6.0, 6.4, 6.9, 8.5, 9.1
4	0.5, 1.4
5	2.4, 3.3, 8.6

   b. **Key:** 2|3 = 23 cents (data is truncated)

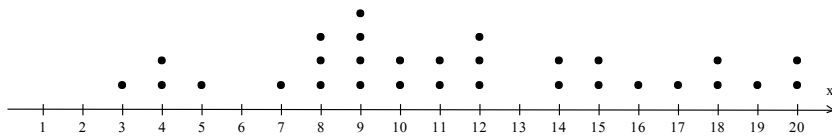| Stem	Leaves
0	9
1	6 6 7 8 8 9
2	0 0 0 2 3 3 3 3 3 4 4 4 4 5 6 6 7 7 8
3	0 0 0 0 0 1 1 2 2 3 4 5 5 6 6 6 8 9
4	0 1
5	2 3 8

c. I have included the window used as well as the histogram below:

NORMAL FLOAT AUTO REAL DEGREE MP
FUNCTION TRACE VALUES

WINDOW
 Xmin=0
 Xmax=60
 Xscl=10
 Ymin=-5
 Ymax=25
 Yscl=1
 Xres=1
 ΔX=0.22727272727273
 TraceStep=0.454545454545…

d. I have included the window used as well as the histogram below:

NORMAL FLOAT AUTO REAL DEGREE MP
FUNCTION TRACE VALUES

WINDOW
 Xmin=0
 Xmax=60
 Xscl=5
 Ymin=-5
 Ymax=20
 Yscl=1
 Xres=1
 ΔX=0.22727272727273
 TraceStep=0.454545454545…

5.

a.

b. 1L|2 means a score of 12 on the quiz

| Stem	Leaves
0L	3 4 4
0H	5 7 8 8 8 9 9 9 9
1L	0 0 1 1 2 2 2 4 4
1H	5 5 6 7 8 8 8 9
2L	0 0

c.

NORMAL FLOAT AUTO REAL DEGREE MP
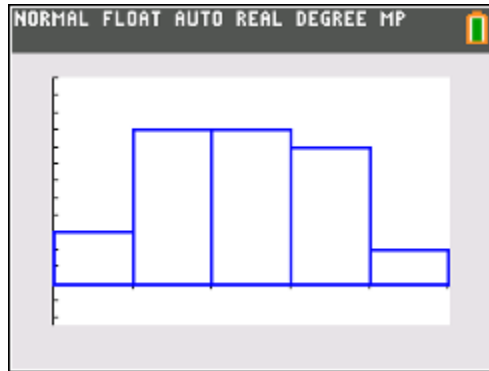
d. The stem-and-leaf plot and the histogram look essentially the same, just rotated 90° from each other. This is essentially because the stems and the "bin" or "class" width for the histograms (1L puts in values of 10 through 14 as leaves for that stem, and the histogram is showing data from 9.97 to 14.96, which includes the exact same numbers from that stem of the stem-and-leaf display).

6.

a.

NORMAL FLOAT AUTO REAL DEGREE MP
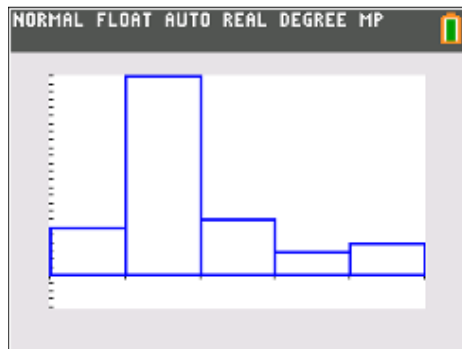
Here is what I used for my window (I did ZoomStat first, then adjusted from there)

Xmin = 20

Xmax = 70

Xscl = 10

Ymax = 25

b. If this is being used to advocate a position that relies on wind speed, using the maximum value could be problematic – what if the maximum wind speed was an outlier in one or more of the years? Knowing the average yearly wind speed and how variable the wind was would likely be better data for power generation – where you need consistent wind at a certain velocity to produce stable amounts of power.
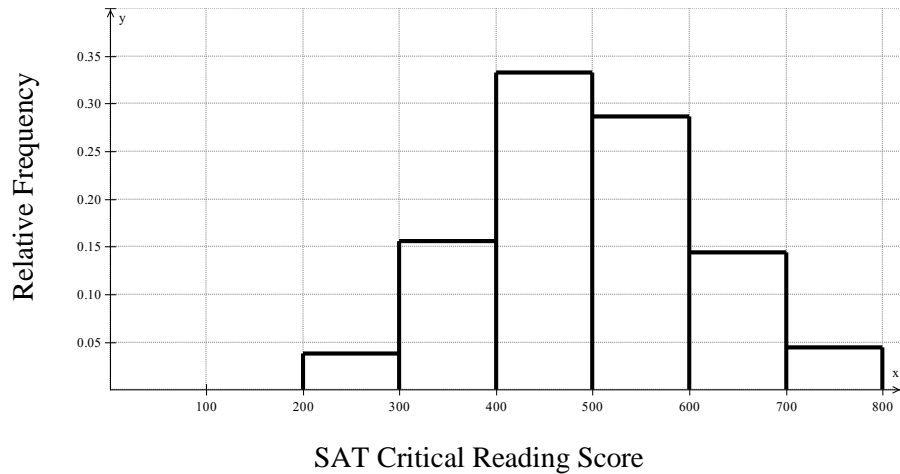
7.

a.

SAT Critical Reading Scores for Male Students

SAT Critical Reading Score
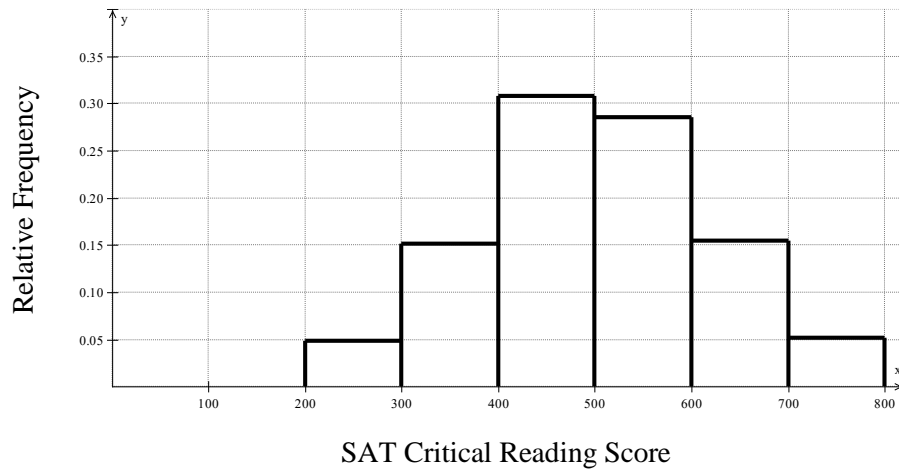
**SAT Critical Reading Scores for Female Students**

SAT Critical Reading Score

b. The distributions are very similar in overall shape, with both being roughly symmetrical. The male has a slightly higher relative frequency for the lowest and highest scores, and the females have a slightly higher relative frequency for the 300 to 500 range of scores.

1.4 Homework Answer Key
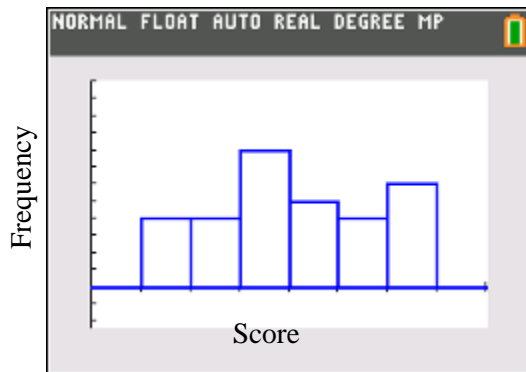
1.

   a.

Number of Quit Attempts
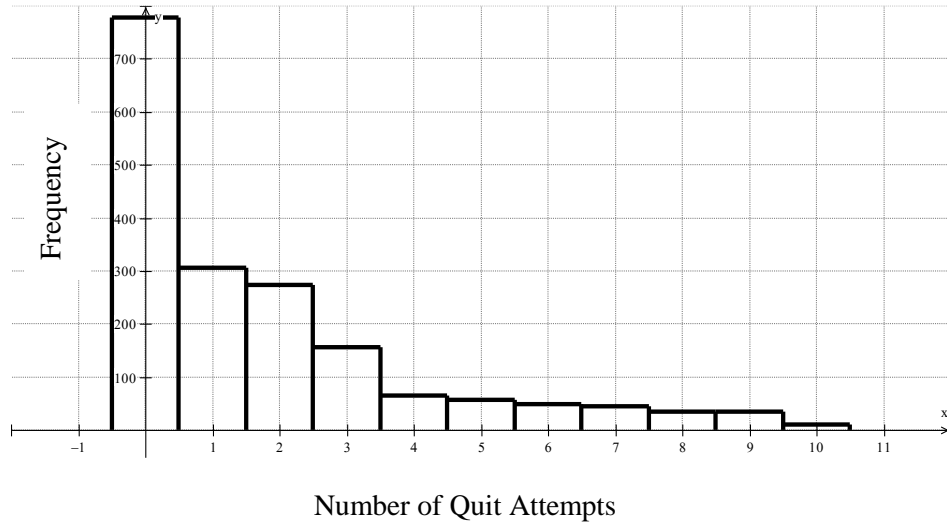
   b. Shape: Positively skewed (or right skewed), no gaps, unimodal
       Outliers: There do not appear to be any outliers
       Center: The mode is 0, median is 1 (there are 1813 values in the list, so the median
       would be the 907th value), mean is 1.76
       Spread: The spread is from 0 to 10, and the range is 10

2.

   a.

   x-axis is scaled at 3 units
   per mark

   y-axis is scaled at 1 unit
   per mark

   b. Shape: Not skewed, roughly symmetrical (though the second mode at the last class of
       data makes it not really great symmetry), bimodal.
       Outliers: There do not appear to be any outliers
       Center: The median is 11, mean is 11.77
   Spread: The spread is from 3 to 20, and the range is 17

3.
   a. The data is left-skewed.
   b. There is a gap from 1 to 4, and 1 appears to be an outlier.
   c. The range is 7, the spread is from 1 to 8.
   d. The median and mode are both 7 (there are 18 values so the median would be the average of the 9th and 10th values). The mean is 6.68.

4. Shape: Positively (right) skewed.
   Outliers: 79 appears to be an outlier, and there is a gap between 49 and 79 minutes.
   Center: The median is 19 minutes, while the mean is 29.67 minutes
   Spread: The spread is from 4 minutes to 79 minutes, while the range is 75 minutes.

5.
   a. Positively Skewed
   b. Roughly Symmetric
   c. Negatively Skewed
   d. Positively Skewed

6.
   a.
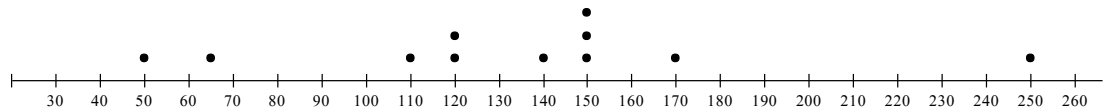
   b. Mean: 134.1, Median: 140, Mode: 150

7.
   a. The fact that the mean is much higher than the median means the data is very positively (or right) skewed. That is, there are likely a few weddings that are very expensive shifting the average much higher.
   b. I would agree that it is grossly misleading – most people think of the average as "what most people spend", so the perception would be that the average couple spends about $27,000 on their wedding, when half of all couples spend about $18,000 or less.
   c. The wedding industry might report the data this way to encourage people to spend more on their weddings – if a customer thinks that most people are spending more than they are, they might want to spend more to "fit in".
   d. This is correct, because "minority" means less than 50%. Since 50% of the data is above the median (which is $18,086), and assuming that some people do spend between the median and the average (which is very likely), then a minority of people would spend the mean or above.

1.5 Homework Answer Key

1.

   a. Since the mean is higher than the median, and the highest value is very far from the median and Q3, this data is positively (or right) skewed.
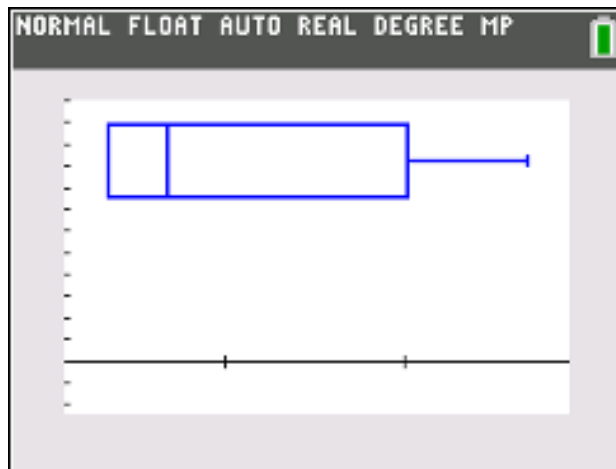
   b.
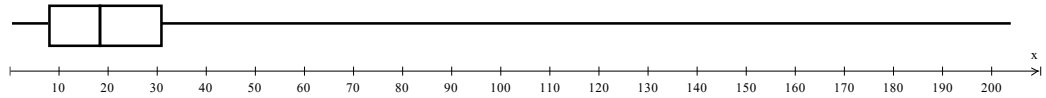
   c. Since the Inter-Quartile Range (IQR) is 24, anything higher than 24 more than Q3 is, by definition, an outlier. Since Q3 = 31, anything higher than 55 minutes is an outlier. Since the maximum is 205 minutes, there must be at least one outlier.
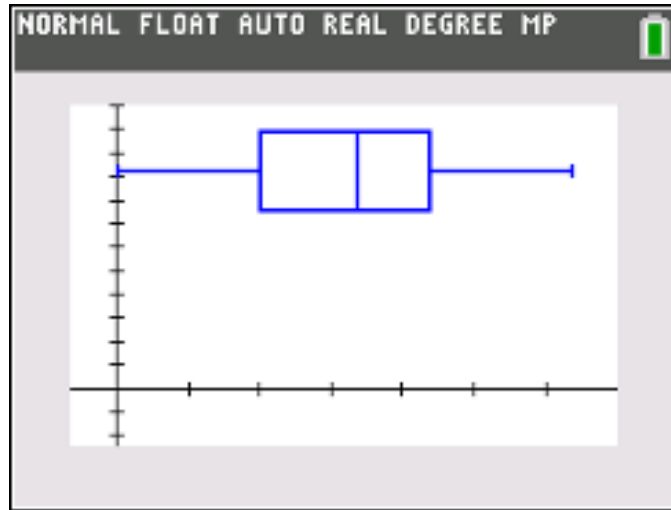
2.

   a. Minimum: 7, Lower Quartile: 7, Median: 8, Upper Quartile: 12, Maximum: 14

   b. The minimum and the lower quartile are the same because there are 18 data values, which means there are 9 values below the median, so the 5th value is the lower quartile. Since the first 5 values are all 7, then the minimum and the lower quartile are the same.

3.
   a. Minimum: 0, Q1: 6, Median: 10, Q3: 13, Maximum: 19
   b.

   c. There are no outliers in the data. The Inter-Quartile Range is 7, so outliers would have to be below –4.5 grams of sugar (which is impossible), or above 23.5 grams of sugar, but the maximum is only 19 grams of sugar, so there are no outliers.

4.
   a. Minimum: 40, Q1: 50, Median: 60, Q3: 75, Maximum: 120
   b. Yes, 120 is an outlier.
   c. The IQR is 25, Q3 + 1.5(IQR) = 112.5, so 120 is an outlier.

5.
   a. Minimum: 3 points, lower quartile: 8 points, median: 11 points, upper quartile: 16 points, maximum: 20 points
   b. The interquartile range is 8 points, which means outliers would have to below –4 points or above 28 points. Since no values fall into that range, there are no outliers.
   c. Sd
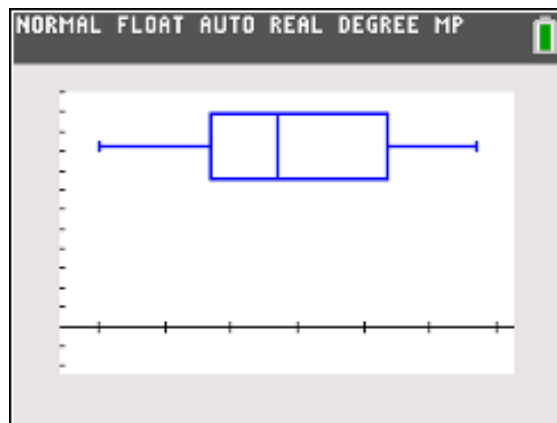
1.6 Homework Answer Key

1.
   a. +/– 1 standard deviation: 42 minutes and 54 minutes
      +/– 2 standard deviations: 36 minutes and 60 minutes
   b. Because the collection is normally distributed, 68% of the values fall between 42 and 54 minutes, and 95% of the values fall between 36 and 60 minutes.
   c. Since 66 minutes is 3 standard deviations above the mean, only the top 0.15% of his collection would be longer, so we would assume that not very many albums in his collection would be this length.

2. It is likely not normally distributed, for a normal distribution 95% of the values fall within 2 standard deviations of the mean. That would mean 95% of the values would be in between –20.5 minutes and 83.5 minutes. Since this would require a large portion of the data to be negative (which is impossible), the distribution cannot be normal.

3.
   a. Since the 98th percentile is 2 standard deviations above the mean, that would mean 74 minutes of time to correct tests. So he should begin correcting 74 minutes before 9 am, which would be 7:46 am
   b. Since this amount of time (50 minutes) is one standard deviation below the mean, this is associated with the 16th percentile, so he would have to have a correction time in the 16th percentile or below.

4. Since the mean is 509 and the 16th percentile is 415, the standard deviation must be approximately 94 (since the 16th percentile is associated with 1 standard deviation below the mean).
   a. A score of 603 is the 84th percentile (84th percentile is associated with 1 standard deviation above the mean).
   b. The approximate standard deviation is 94 (this is not exact, because the 16th percentile is not *exactly* one standard deviation below the mean, it is approximately one standard deviation below the mean).
   c. An exam score of 600 would have a z–score of 0.97 and an exam score of 480 would have a z–score of –0.31.
   d. An exam score of 227 would be a z–score of about –3.0, or 3 standard deviations below the mean. Since about 99.7% of the data should fall between +/– 3 standard deviations of the mean, only about 0.15% of the tests should score this low. Therefore, we would not expect to see many scores below 227.

5. The rat's z–score for Stimulus 1 is –0.83, and its z–score for Stimulus 2 is –1.00, so its response to the second stimulus is relatively more rapid compared to other rats (the higher absolute z–score means the value is further from the mean).

1.7 Homework Answer Key

1.
    a. Your ACT score has a z–score of 1.8, and your SAT had a z–score of 1.57, so your score compared to the rest of the population taking the test is better on the ACT.
    b. Your friend's ACT score has a z–score of 0.8, and your SAT had a z–score of 0.65, so your friend's score compared to the rest of the population taking the test is better on the ACT. You should advise him to send the ACT score.

2.
    a. The mean and median for the black graph is approximately 2.5 (they are the same because the graph is symmetrical). For the red graph, the median is probably around 3.9, while them mean is more likely around 3.6.
    b. The black graph has no skew, it appears very symmetric (and approximately normal), while the red graph appears to be skewed left (or negatively skewed).
    c. Since both the mean and the median appear to be quite a bit higher for the red graph than for the black graph, it appears that the tutoring had a positive effect. In addition, shifting from no skew to left skewed means that there are fewer lower performing students in the population represented by the black graph.

3.
    a. The range of both datasets is 65, but the spreads are different. The spread of the left side is from 23 to 88, while the spread of the right side is 33 to 98.
    b. Since both datasets have 18 values, the median will be the average of the ninth and tenth values. The left set has a median of 56 while the right set has a median of 66.
    c. There do not appear to be any outliers in either data set.
    d. Answers may vary – if you want to wait less, you should go to the clinic whose data is represented on the left side. Depending on the cause of the wait time, you may choose differently – for example, what if the wait was longer at the clinic represented on the right side because they gave more attention to each patient, so that may be a choice that someone would make (depending on data not presented in the problem).

4. Answers are approximates since the boxplots don't necessarily align perfectly with the scale.
    a.

Boxplot **a**	Boxplot **b**
Minimum:30	Minimum: 85
Q1: 41	Q1: 95
Median: 51	Median: 108
Q3: 66	Q3: 113
Maximum: 95	Maximum: 120

Boxplot **a**:
Shape is skewed right, no outliers, median is about 51, spread is from 30 to 95 and the range is 65

Boxplot **b**:
Shape is roughly symmetric (maybe a very slight left skew), no outliers, median is about 108, spread is from 85 to 120 with a range of 35.

Comparison:
Boxplot **b** is less skewed than boxplot **a**, and most of the values in boxplot **a** are higher than the maximum value of boxplot **b** – since the lower quartile of **a** equals the maximum value of **b**, that means 75% of the data in **a** is higher than the data in **b**. The median of **a** is more than twice the median of **b**, and there is less variability in **a**, because it has a smaller spread and range.

b.
    i. True.
   ii. False, data set **b** seems to be skewed right.
  iii. True.
  iv. False, the maximum for **b** is 95, and the minimum for **a** is 85
   v. True.