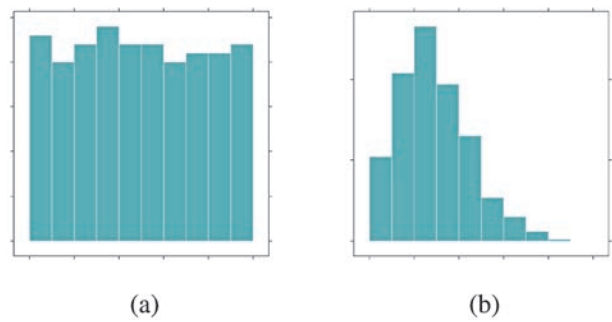


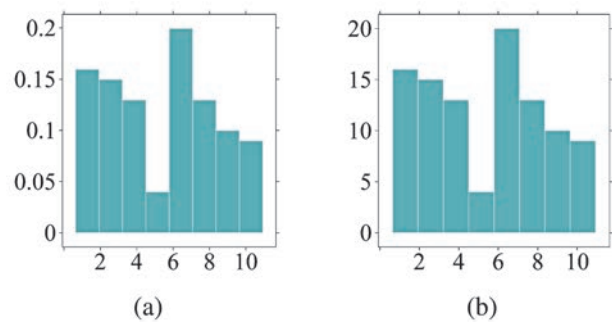
Chapter 2 Test C - Short Answer

Section 2.1 (Visualizing Variation in Numerical Data)

1. [Objective: Know how to make graphs of distributions of numerical variables and interpret the graphs in context.] Below are two histograms. One corresponds to the ages at which a sample of people applied for marriage licenses; the other corresponds to the last digit of a sample of social security numbers. Which graph is which, and why?

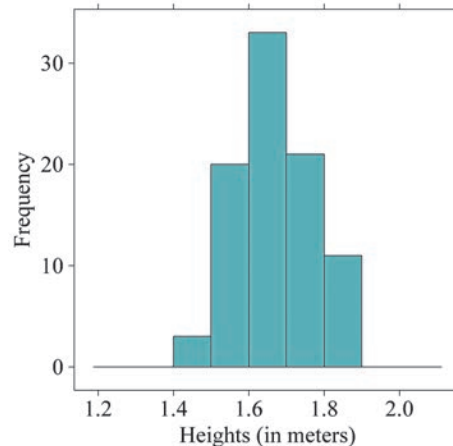


2. [Objective: Understand the difference between frequencies and relative frequencies in a histogram.] The two histograms below display the exact same data. How do the plots differ?



3. [Objective: Understand differences and similarities between dotplots and histograms.] How is a dotplot similar to a histogram? How is it different?

4. [Objective: Understand that a distribution of a sample of data can be displayed multiple ways.] If you were to create a dotplot to display the same data that is represented in the following histogram, how many dots would you draw to represent heights that fall between 1.5 meters and 1.6 meters?



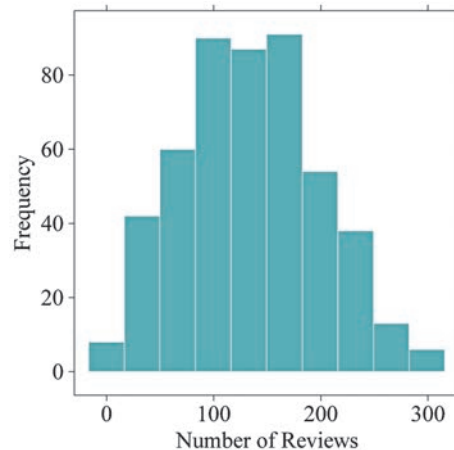
Section 2.2 (Summarizing Important Features of a Numerical Distribution)

5. [Objective: Know what to pay attention to in distributions of numerical data.] When examining distributions of numerical data, what three components should you try to describe?
6. [Objective: Understand modality in distributions.] Describe a scenario in which a distribution could be bimodal. Explain your reasoning.

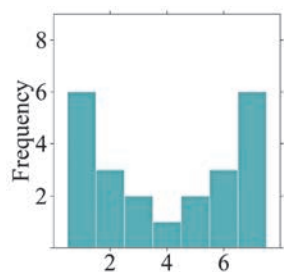
For questions (7) - (9), what would you expect the shape of the distribution described to look like? Explain your reasoning.

7. [Objective: Recognize the shape of a distribution.] The distribution of the household incomes in a large city.
8. [Objective: Recognize the shape of a distribution.] The distribution of scores on an easy test.
9. [Objective: Recognize the shape of a distribution.] The distribution of the time (in minutes) it takes to drive to work using the same route each day.

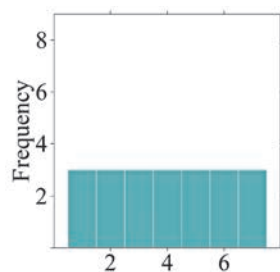
10. [Objective: Understand how to find typical values from a histogram.] The following histogram represents the number of reviews a movie received on a popular website. What is the *typical* number of reviews a movie is expected to receive, according to this distribution? Explain your reasoning.



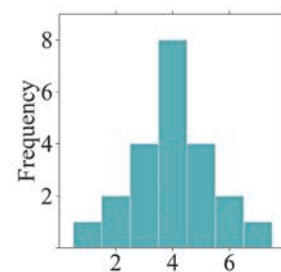
11. [Objective: Determine differences in variability.] Order the following histograms from least to most variability. Explain your reasoning.



(i)

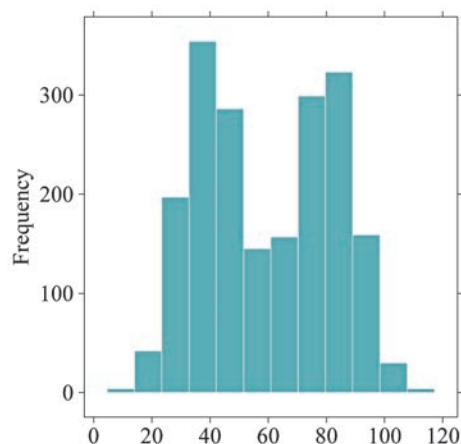


(ii)



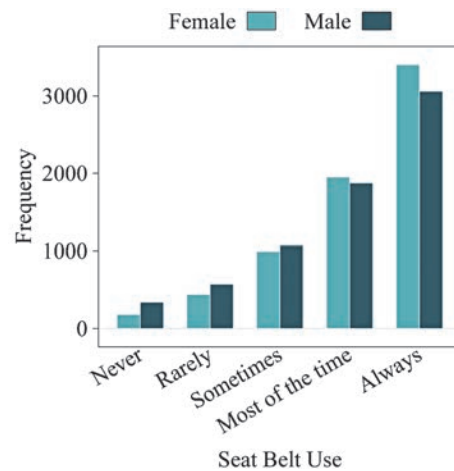
(iii)

12. [Objective: Interpreting typical values of bimodal distributions.] How would you describe the typical value for this histogram? Explain your reasoning.

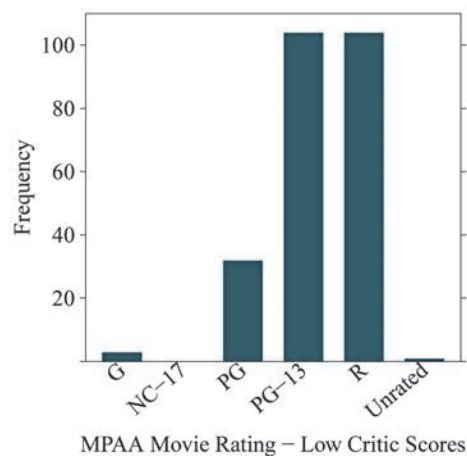
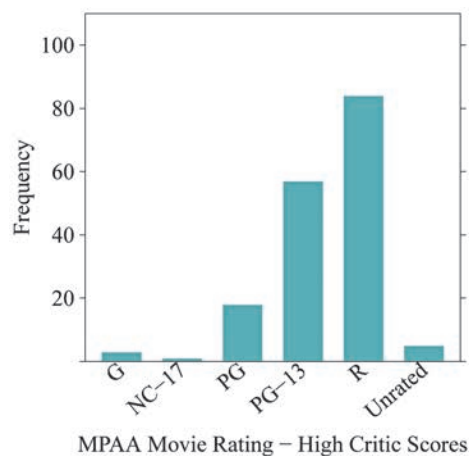


Section 2.3 (Visualizing Variation in Categorical Variables)

13. [Objective: Understand differences between bar charts and histograms.] What is the difference between a bar chart and a histogram?
14. [Objective: Interpreting bar charts.] Using the following bar chart, what can you say about the difference in seat belt use for males versus females?

**Section 2.4 (Summarizing Categorical Distributions)**

15. [Objective: Understand the term *mode* when describing categorical variables.] What does it mean to find the *mode* of a bar chart?
16. [Objective: Determine the variability of categorical data from a bar chart.] The bar charts below depict the MPAA movie ratings of 489 movies, separated by high and low critic scores. Which bar chart shows more variability in MPAA movie ratings? Why?



Section 2.5 (Interpreting Graphs)

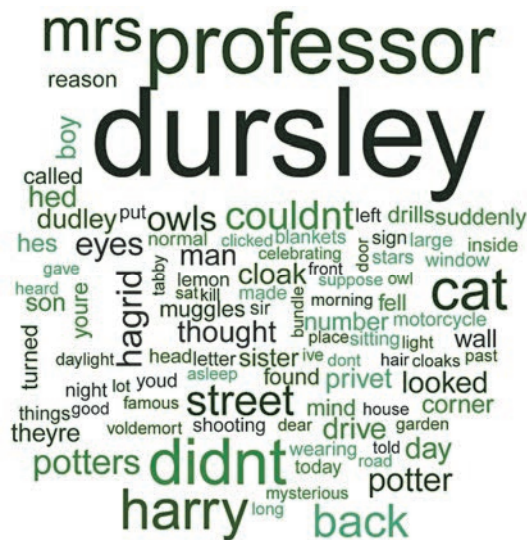
Use the following information to answer questions (17) - (18):

A large state university conducted a survey among their students and received 400 responses. The survey asked the students to provide the following information:

- * Age
- * Year in School (Freshman, Sophomore, Junior, Senior)
- * Gender

17. [Objective: Determine appropriate graph based on variable type.] What type of graph would you use to describe the variable Year in School? Explain your reasoning.
18. [Objective: Determine appropriate graph based on variable type.] What type of graph would you use to describe the variables Gender and Year in School? Explain your reasoning.

A word cloud was created using the first chapter of J.K. Rowling's Harry Potter and the Sorcerer's Stone. (Note that filler words such as "the," "a/an," and "and" were excluded from the plot.) Use the word cloud to answer questions (19) - (20).



19. [Objective: Interpreting word clouds.] According to the word cloud, what is the most common word in the first chapter of Harry Potter and the Sorcerer's Stone? Why?
20. [Objective: Pitfalls of using word clouds.] The words “owls” and “eyes” appear to be similar in size. Does this mean that each of these words is used the same number of times in the first chapter of the book? Why or why not?

Chapter 2 Test C - Answer Key

1. Histogram (a) displays the last digits of social security numbers because all of the values are mostly equally likely. Since the last digit of social security numbers are created randomly, we would expect any digit between 0 and 9 to show up just as often as another digit. Histogram (b) displays the ages at which a sample of people applied for a marriage license. Since most people get married in their early to mid-twenties, but there are also people who wait to get married until a later age, we would expect the distribution to be right-skewed.
2. Histogram (a) uses frequencies to simply count the number of observations at a given value. Histogram (b) uses relative frequencies to show the proportion of observations at a given value.
3. A dotplot and a histogram both show the overall shape of a distribution. They both can help determine a distribution's shape, center, and spread. They differ in terms of appearance in only one way. A dotplot displays a dot to represent each observation in the data, while a histogram uses bars to display intervals of observations.
4. About 20 dots should be drawn because there are about 20 people whose heights fall between 1.5 meters and 1.6 meters, as shown by the frequency value on the y -axis.
5. Shape, center, and spread of the data.
6. Answers may vary. Some examples include: (1) The price of college tuition, including both public and private schools (the different types of colleges would create two modes - private colleges would most likely have higher tuition costs compared to public schools). (2) The heights of all students at a high school (the different genders would create two modes - males are typically taller than females).
7. The distribution of incomes would most likely be right-skewed because most people earn middle-class salaries, but the very wealthy people are likely to earn incomes much higher than average.
8. The distribution of scores on an easy test would most likely be left-skewed because most test-takers will do well on the test, and a few will still do poorly.
9. The distribution of the time it takes to drive to work using the same route each day should be roughly symmetric because the time you leave your house is probably the same each day. The commute times will be very similar on a day-to-day basis.
10. The typical number of reviews a movie will receive is about 130. We know this because the distribution is centered around the value 130 on the x -axis.
11. Least to most variability: (iii), (ii), (i). Histogram (iii) has the least variability because it has more points that are close to the center of the distribution. Histogram (i) has the most variability because it has more points that are far away from the center of the distribution.
12. Since the data are bimodal, there are two typical values - one is about 40 and the other is about 80.
13. A bar chart represents a categorical variable and a histogram represents a numerical variable.

14. Answers may vary. Some examples include: (1) In general, people always wear seat belts. (2) Females wear seat belts more than males. (3) About the same number of males and females report wearing seat belts “sometimes.”
15. The mode can be found by finding the bar, or category, with the most observations. It will be the highest bar in the plot.
16. The “high critic scores” bar chart shows more variability because there are more observations in the different categories than there are for the “low critic scores.”
17. A bar chart because Year in School is a categorical variable.
18. A side-by-side bar chart should be used since these are two categorical variables.
19. The most common word is “dursley” because it is the largest in size.
20. No. A word cloud can only tell us what words are the most common, but it cannot tell us exactly how many times a specific word appeared in the text.