2.2 Graphical summaries of data

Luis Talavera

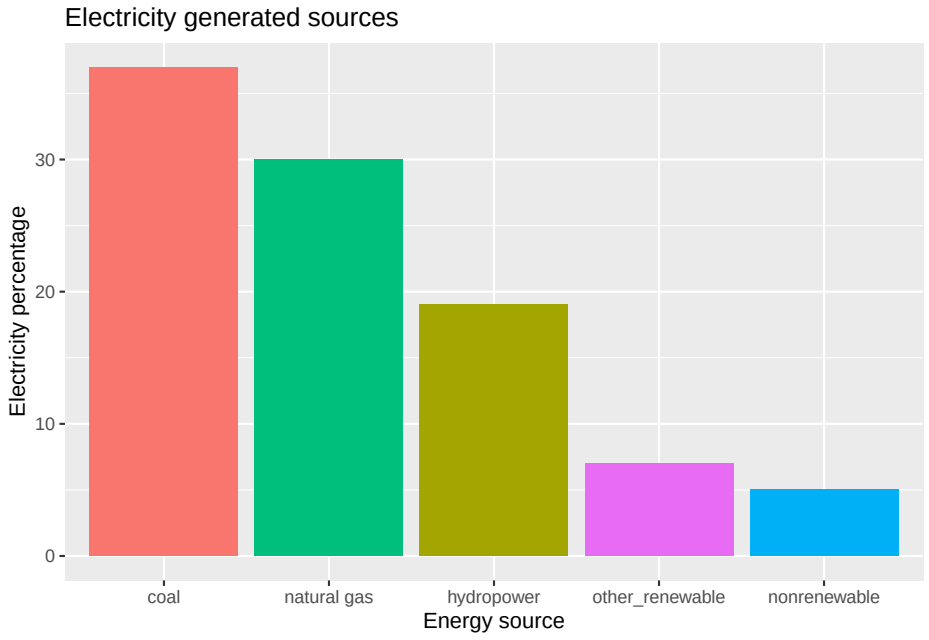
2022-12-18

Exercises

2.10 Generating Electricity In 2012 in the United States, most electricity was generated from coal (37%), natural gas (30%), or nuclear power plants (19%). Hydropower accounted for 7% of the total electricity produced; other renewable sources such as wind or solar power accounted for 5%. Other nonrenewable sources (such as petroleum) made up the remaining 2%. (Source: http://www.eia.gov/ electricity/annual/html/epa_01_01.html)
a. Display this information in a bar graph.
b. Which is easier to sketch relatively accurately, a pie chart or a bar chart?
c. What is the advantage of using a graph to summarize the results instead of merely stating the percentages for each source?
d. What is the modal category?

Ans.

a.



b. A bar chart
c. The modal category can be easily identified as well as the distribution of the data.

d. Coal

2.11 What do alligators eat? The bar chart is from a study investigating the factors that influence alligators' choice of food. For 219 alligators captured in four Florida lakes, researchers classified the primary food choice (in volume) found in the alligator's stomach in one of the categories—fish, invertebrate (snails, insects, crayfish), reptile (turtles, baby alligators), bird, or other (amphibian, mammal, plants). (Data available on the book's website.)
a. Is primary food choice categorical or quantitative?
b. Which is the modal category for primary food choice?
c. About what percentage of alligators had fish as the primary food choice?
d. This type of bar chart, with categories listed in order of frequency, has a special name. What is it?

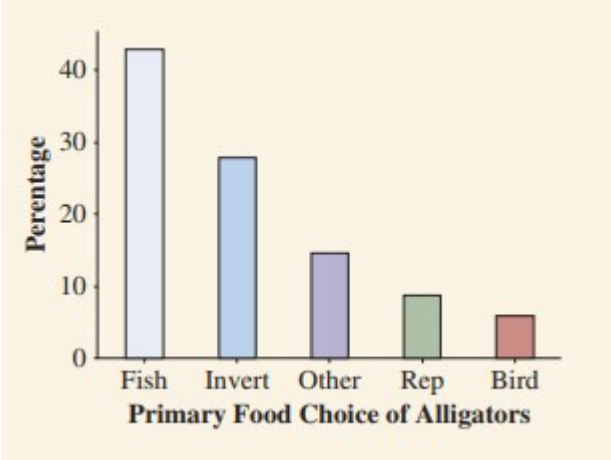


Figure 1: Primary Food Choice of Alligators

Ans.

a. Primary food is a categorical food
b. Fish
c. ~42%
d. Pareto chart

2.12 Weather stations The pie chart (constructed using EXCEL) shown portrays the regional distribution of weather stations in the United States.
a. Do the slices of the pie portray (i) variables or (ii) categories of a variable?
b. Identify what the two numbers mean that are shown for each slice of the pie.
c. Without inspecting the numbers, would it be easier to identify the modal category by using this graph or the corresponding bar graph? Why?

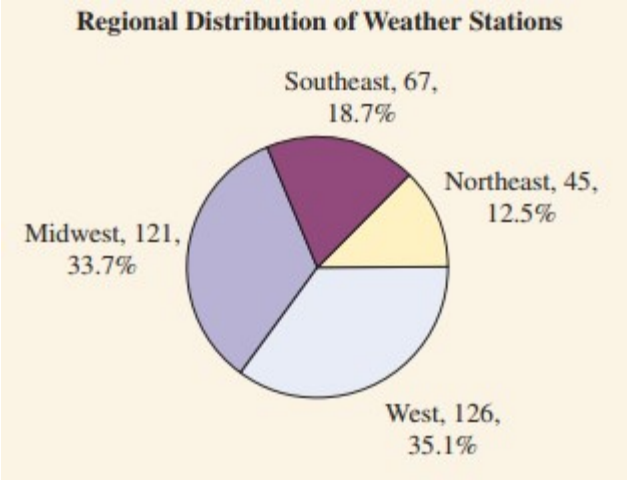


Figure 2: Regional Distribution of Weather Stations

Ans.

a. Categories of a variable.
b. The absolute and relative frequency.
c. It would be easy using a bar graph, without the numbers both Midwest and West can be considered as the modal category.

2.13 France is most popular holiday spot Which countries are most frequently visited by tourists from other countries? The table shows results according to Travel and Leisure magazine (2005).
a. Is country visited a categorical or a quantitative variable?
b. In creating a bar graph of these data, would it be most sensible to list the countries alphabetically or in the form of a Pareto chart? Explain.
c. Does either a dot plot or stem-and-leaf plot make sense for these data? Explain.

Most Visited Countries, 2005

Country	Number of Visits (millions)
France	77.0
China	53.4
Spain	51.8
United States	41.9
Italy	39.8
United Kingdom	24.2
Canada	20.1
Mexico	19.7

Source: Data from *Travel and Leisure* magazine, 2005.

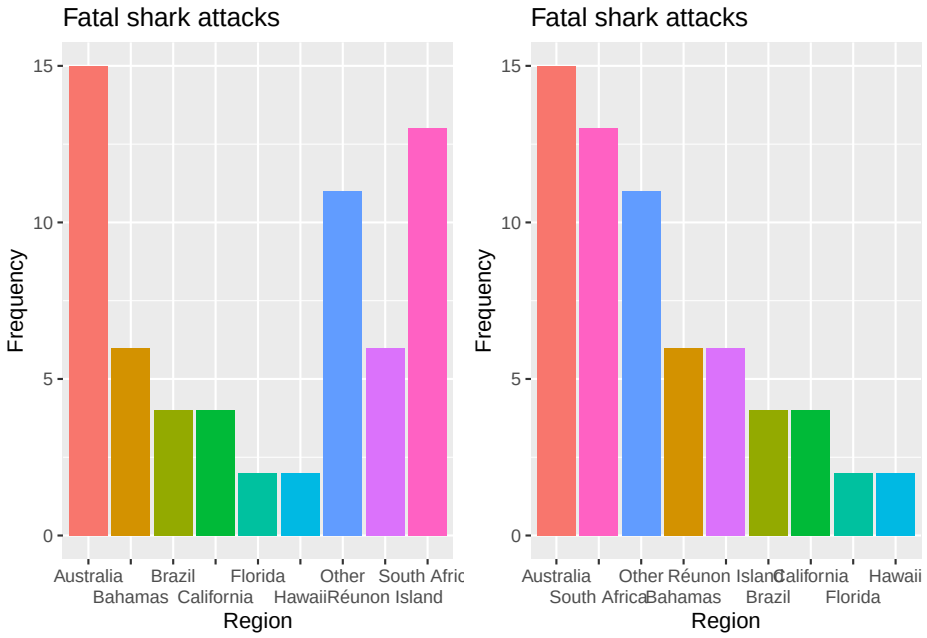
Figure 3: Regional Distribution of Weather Stations

Ans.

a. A categorical variable.

b. Pareto chart form is better, it displays the most visited countries at the beginning of the table
c. They does not make sense. These graphs are not for categorical data.

2.14 Pareto chart for fatal shark attacks The data shown in Exercise 2.9 give frequencies of fatal shark attacks for different regions. Using software or sketching, construct a bar graph, ordering the regions (i) alphabetically and (ii) as in a Pareto chart. Which do you prefer? Why?



Ans. I prefer Pareto form chart due to displays most frequent categories at the left and least important categories at the right.

2.15 Sugar dot plot For the breakfast cereal data given in Table 2.3, a dot plot for the sugar values (in grams) is shown:
a. Identify the minimum and maximum sugar values.
b. Which sugar outcomes occur most frequently? What are these values called?

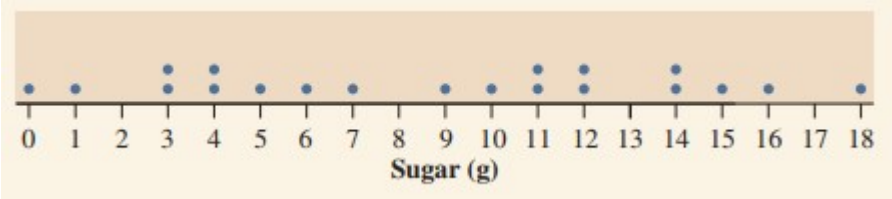


Figure 4: Breakfast cereal sugar data

Ans.

a. 0g and 18g
b. 3, 4, 11, 12, 14. These values are called modes.

2.16 Spring break hotel prices For a trip to Miami, Florida, over spring break in 2014, the data below (obtained from travelocity.com) show the price per night (in U.S. dollars) for various hotel rooms.

239	237	245	310	218	175	330	196	178
245	255	190	330	124	162	190	386	145

a. Construct a stem-and-leaf plot. Truncate the data to the first two digits for purposes of constructing the plot. For example, 239 becomes 23.

b. Reconstruct the stem-and-leaf plot in part a by using split stems; that is, two stems of 1, two stems of 2, and so on. Compare the two stem-and-leaf plots. Explain how one may be more informative than the other.

c. Sketch a histogram by hand (or use software), using 6 intervals of length 50, starting at 100 and ending at 400. What does the plot tell about the distribution of hotel prices? (Mention where most prices tend to fall and comment about the shape of the distribution.)

Ans.

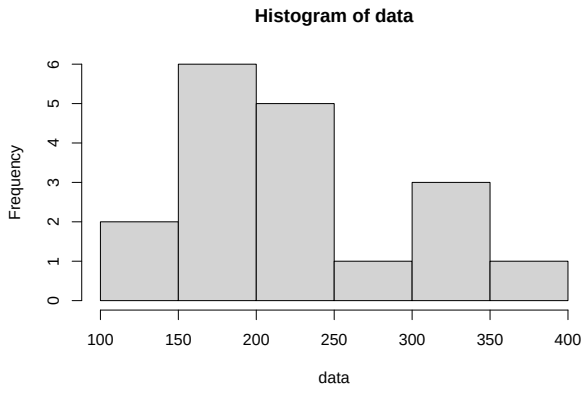
a.

```
The decimal point is 2 digit(s) to the right of the |

1 | 2568899
2 | 0244556
3 | 1339
```

b.

```
The decimal point is 2 digit(s) to the right of the |

1 | 2
1 | 568899
2 | 0244
2 | 556
3 | 133
3 | 9
```

c.



Exercises 2.17, 2.18 and 2.19 were similar, since a leaf-stem chart is rarely used I decided to skip those exercises

2.20 Histogram for sugar For the breakfast cereal data, the figure below shows a histogram for the sugar values in grams.
a. Identify the intervals of sugar values used for the plot.
b. Describe the shape of the distribution. What do you think might account for this unusual shape? (Hint: How else are the cereals classified in Table 2.3?)
c. What information can you get from the dot plot or stem-and-leaf plot of these data shown in Exercises 2.15 and 2.19 that you cannot get from this plot?
d. This histogram shows frequencies. If you were to construct a histogram by using the percentages for each interval, how (if at all) would the shape of this histogram change?

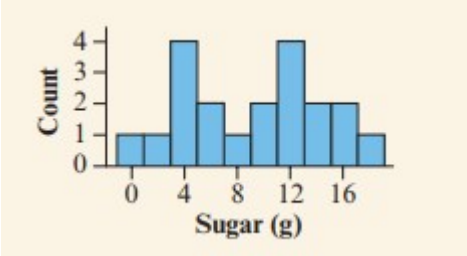


Figure 5: Breakfast cereal data

Ans.

a. 0-2, 2-4, 4-6, 6-8, 8-10, 10-12, 12-14, 14-16, 16-18
b. Is a bimodal shape
c. The exact numbers
d. It would only change the y axis but the shape will keep the same

2.21 Shape of the histogram For each of the following variables, indicate whether you would expect its histogram to be symmetric, skewed to the right, or skewed to the left. Explain why.
a. The price of a certain model of smartwatch in different stores in your district
b. The amount of time students use to take an exam in your school
c. The grade point average (GPA) in your academic program this year
d. The salary of all the employees in a company.

Ans.

a. symmetric, the majority of stores will sell the smartwatch at relatively close prices.
b. Skewed to the left, most of students spend the same amount of time to study, just a few who didn't compromised study for a short time.
c. symmetric, most of students will share GPA.
d. Skewed to the right, the differences betwwen salaries could be so large.

2.22 More shapes of histograms Repeat the preceding exercise for
a. The winner's score in a basketball game during an NBA season.
b. The distance from home to school for students in a specific school
c. The number of attempts a young adult needs to pass a driving license test
d. The number of times an individual requests a password reset for the forgotten password of his or her email account.

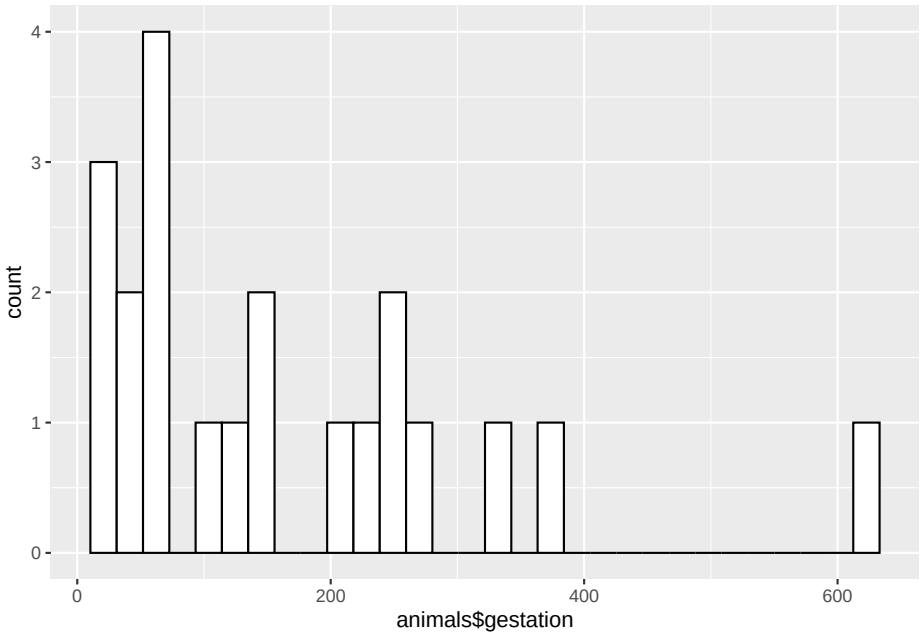
Ans.

a. symmetric
b. symmetric
c. Skewed to the right.
d. Skewed to the right.

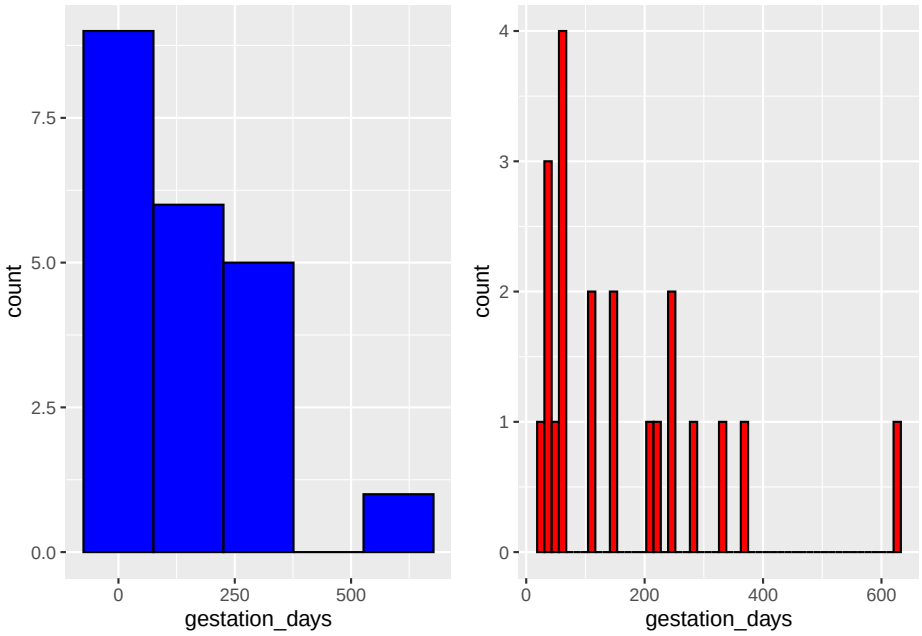
2.23 Gestational Period The Animals data set at the book's website contains data on the length of the gestational period (in days) of 21 animals. (Source: Wildlife Conservation Society)
a. Using software, plot a histogram of the gestational period.
b. Do you see any observation that is unusual? Which animal is it?
c. Is the distribution right-skewed or left-skewed?
d. Try to override the default setting and plot a histogram with only very few (e.g., 3 or 4) intervals and one with many (e.g., 30) intervals. Would you prefer either one to the histogram created in part a? Why or why not?

Ans.

a.



b. Ass, is humans
c. Right-skewed
d.



I think is case-dependent, more bins more smooth and more informative but too many bins make the plot difficult to understand.

2.24 How often do students read the newspaper? Question 14 on the class survey (Activity 3 in Chapter 1) asked, "Estimate the number of times a week, on average, that you read a daily newspaper."
a. Is this variable continuous or discrete? Explain.
b. The histogram shown gives results of this variable when this survey was administered to a class of 36 University of Georgia students. Report the (i) minimum response, (ii) maximum response, (iii) number of students who did not read the newspaper at all, and (iv) mode.
c. Describe the shape of the distribution.

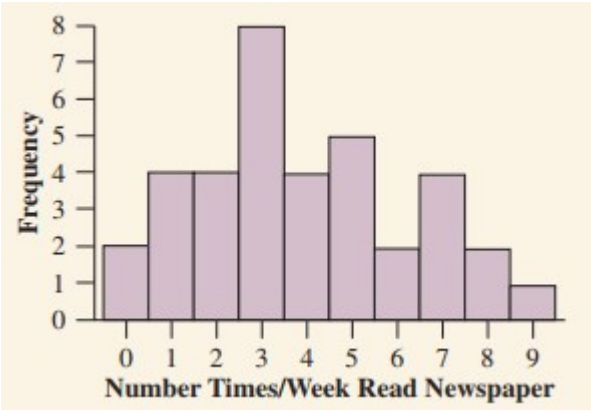


Figure 6: Number Times/Week Read Newspaper 36 Gerogia students

Ans.

a. Discrete, is a countable quantity (1, 2, 3, 4, etc).

b. `i. 0.`
`ii. 9.`

iii. 2.
iv. 3

c.Skewed to the right

2.25 Blossom widths A data set (available at the book's website) analyzed by the famous statistician R. A. Fisher consisted of measurements of different varieties of iris blossoms. Histograms representing the widths of the petals of the two species, Iris setosa and Iris versicolor, are on the next page.

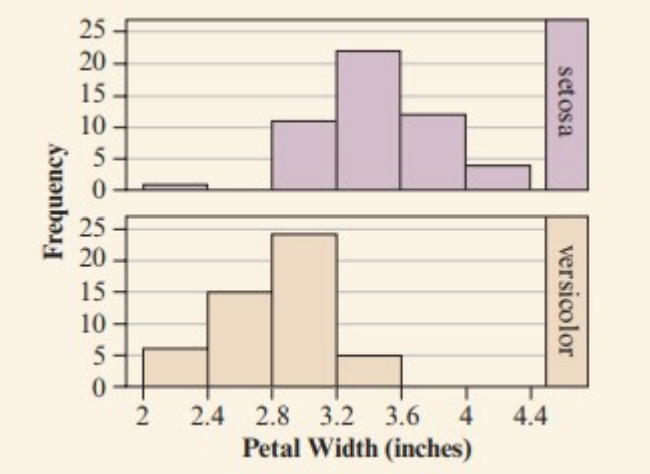


Figure 7: Iris setosa and Iris veriscolor widths

a. Describe the shape of the distribution of setosa petal widths.
b. Describe the shape of the distribution of versicolor petal widths.
c. Of the 50 versicolor blossoms in the data set, approximately what percentage has a petal width less than 3.2 inches?
d. Is it possible to determine accurately the percentage of versicolor blossoms with a width of more than 3 inches? Why or why not?

Ans.

a. Setosa: Skewed to the left.
b. Versicolor: Skewed to the right.
c. ~ 92%
d. Using the graph, yes it shows the frequency of blossoms with a width of more than 3 inches is almost 5.

2.26 Central Park temperatures The first figure shows a histogram of the Central Park, New York, annual average temperatures from 1869 to 2012.

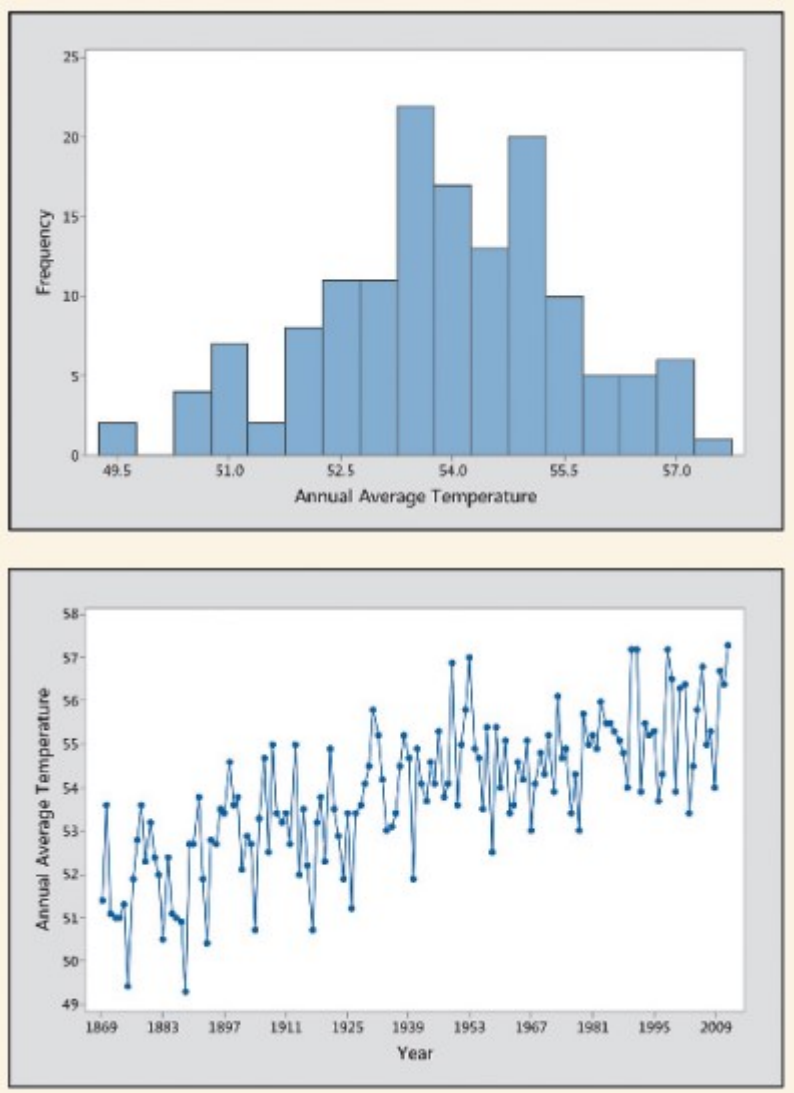


Figure 8: Central Park, New York, annual average temperatures from 1869 to 2012.

a. Describe the shape of the distribution.
b. What information can the time plot above show that a histogram cannot provide?
c. What information does the histogram show that a time plot does not provide?

Ans.

a. Symmetric
b. A decreasing or increasing tendency of averages.
c. Where most of the data falls into.

2.27 Is whooping cough close to being eradicated? In the first half of the 20th century, whooping cough was a frequently occurring bacterial infection that often resulted in death, especially among young children. A vaccination for whooping cough was developed in the 1940s. How effective has the vaccination been in eradicating whooping cough? One measure to consider is the incidence rate (number of infected individuals per 100,000 population) in the United States. The table shows incidence rates from 1925 to 1970.

Incidence Rates for Whooping Cough, 1925–1970	
Year	**Rate per 100,000**
1925	131.2
1930	135.6
1935	141.9
1940	139.6
1945	101.0
1950	80.1
1955	38.2
1960	8.3
1965	3.5
1970	2.1

Source: Data from Historical Statistics of the United States, Colonial Times to 1970, U.S. Department of Commerce, p. 77.

Figure 9: Incidence Rates for Whooping Cough, 1925–1970

a. Sketch a time plot. What type of trend do you observe? Based on the trend from 1945 to 1970, was the whooping cough vaccination proving effective in reducing the incidence of whooping cough?
b. The figure shown is a time plot of the incidence rate from 1980 to 2012, based on numbers reported by the Center for Disease Control and Prevention.3 Describe the time series. Is the United States close to eradicating whooping cough? What do you think contributes to the recent trend? (Data available on the book's website in the pertussis.csv file.)

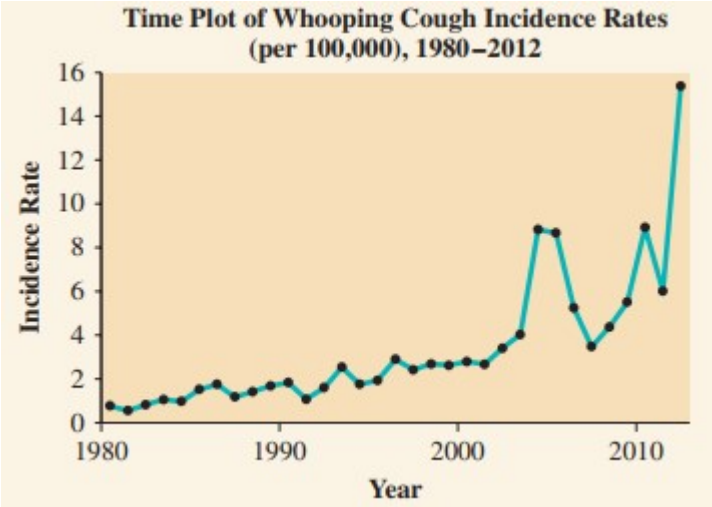


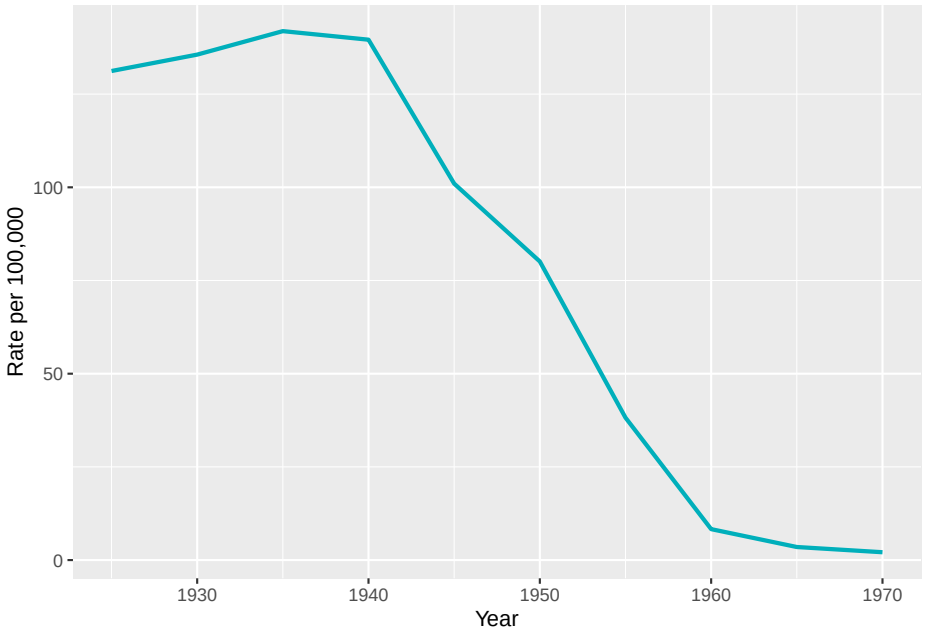
Figure 10: Incidence Rates for Whooping Cough, 1980-2012

c. Would a histogram of the incidence rates since 1935 address the question about the success of the vaccination for whooping cough? Why or why not?

Ans.

a.

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
```

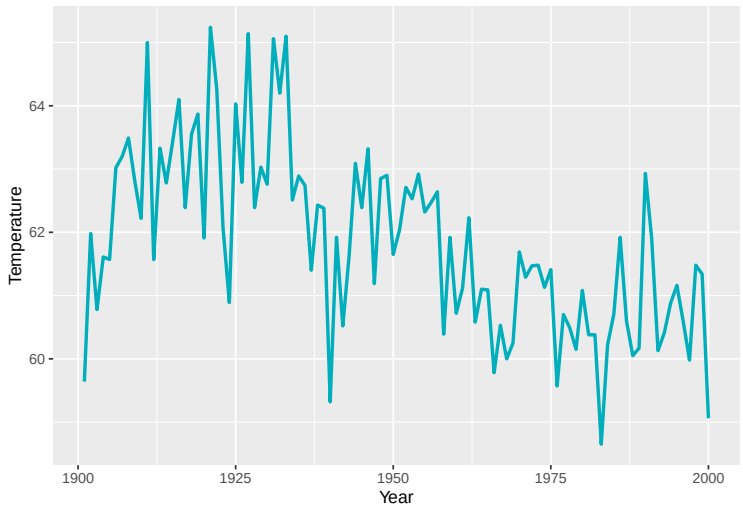


Yes, the rates has been decreasing over the years

b. No, US is not close to eradicate the disease due to people stop from getting the vaccine.
c. By the plots, we could se the campaign was effective and after the campaign stops the rates increased.

2.28 Warming in Newnan, Georgia? Access the Newnan, GA Temps file on the book's website, which reports the average annual temperatures during the 20th century for Newnan, Georgia. Construct a time plot to investigate a possible trend over time. Is there evidence of climate change?

Ans.



Well, the temperature in Newman Georgia has been decreasing over the years, but a temperature records from an specific place can't be used to determine a global climate change.