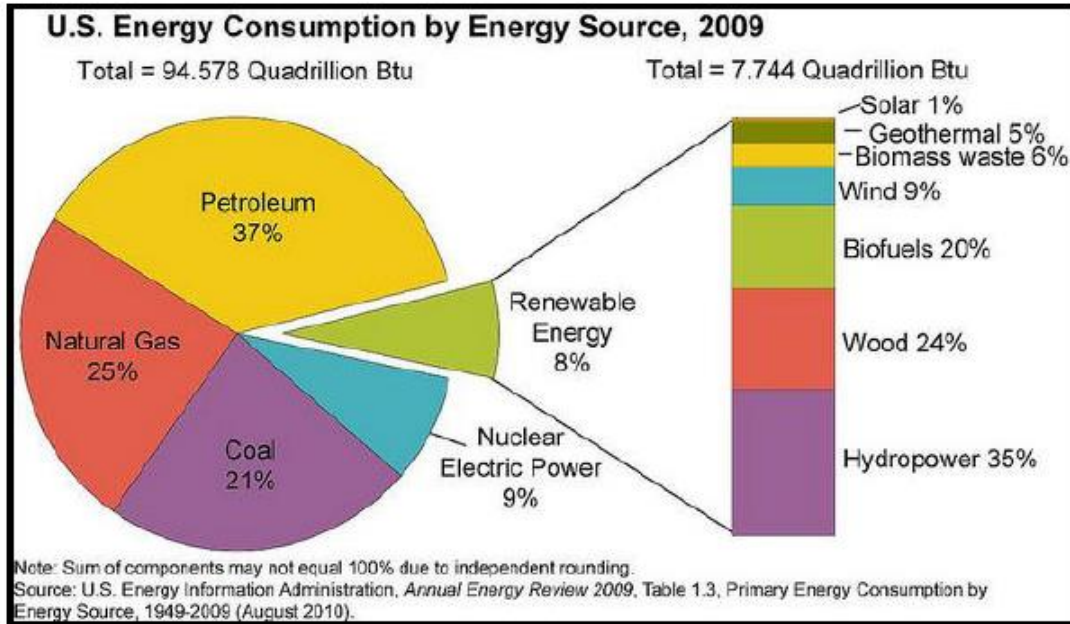


Problem 1

The pie chart below shows the US Energy Consumption by Energy Source for the year 2009¹.



1.1 The source with the highest consumption was

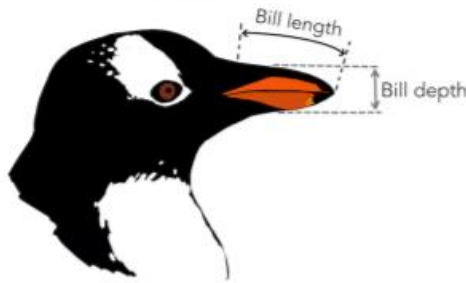
- (a) petroleum.
- (b) natural gas.
- (c) coal.
- (d) renewable energy.

1.2 The combined percent of petroleum and natural gas was

- (a) less than 25% of the total energy consumption.
- (b) between 25% and 50% of the total energy consumption.
- (c) between 50% and 75% of the total energy consumption.
- (d) more than 75% of the total energy consumption.

Problem 2

Data were collected on 344 penguins living on three islands (Torgersen, Biscoe, and Dream) in the Palmer Archipelago, Antarctica. In addition to which island each penguin lives on, the data contains information on the species of the penguin (Adelie, Chinstrap, or Gentoo), its bill length, bill depth, and flipper length (measured in millimeters), its body mass (measured in grams), and the sex of the penguin (female or male). Bill length and depth are measured as shown in the image.¹⁵ (Gorman et al., 2014a) ¹.



2.1 How many cases were included in the data?

2.2 How many numerical variables are included in the data? Indicate what they are, and if they are continuous or discrete.

2.3 How many categorical variables are included in the data, and what are they? List the corresponding levels (categories) for each.

Problem 3

A survey of a random sample of 100 nurses working at a large hospital asked how many years they had been working in the profession. Their answers are summarized in the following (incomplete) table. Fill in the blanks in the table and round your answers to two decimal places:

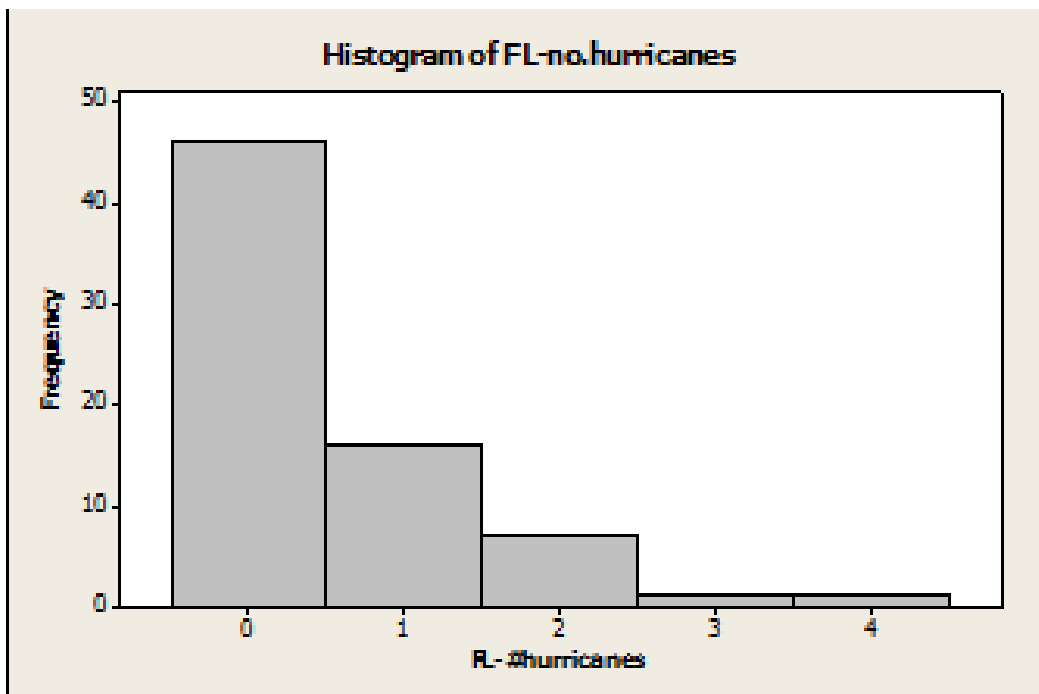
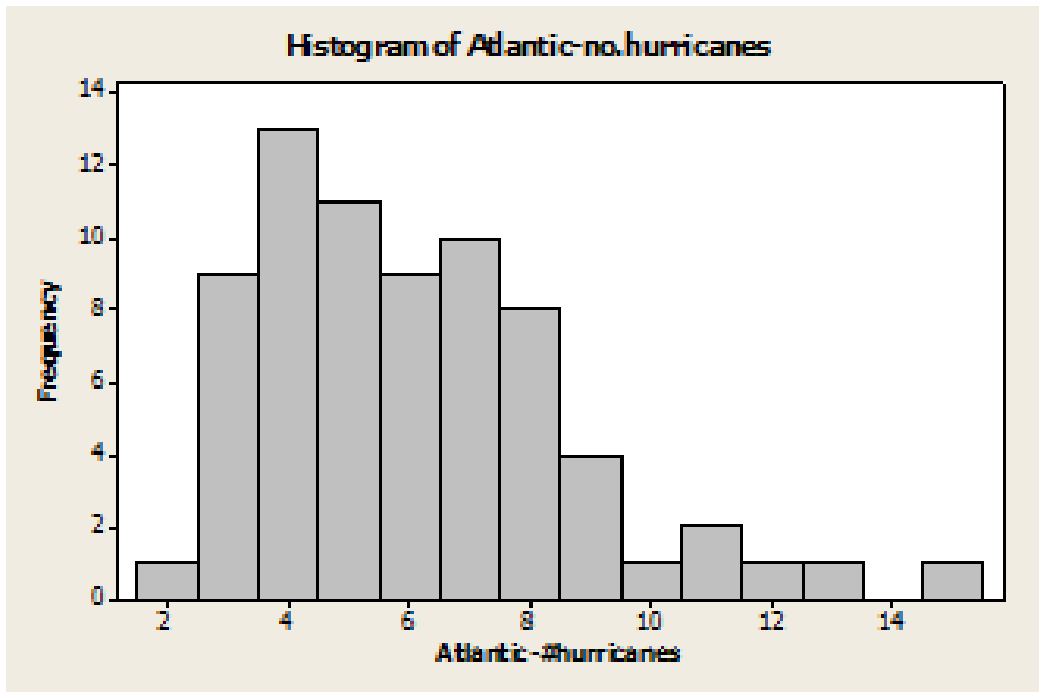
# of years	Frequency	Cumulative frequency	Relative frequency	Cumulative relative frequency
<5			.25	
5-10	30			
>10				

3.1 What proportion of nurses have five or more years of experience?

3.2 What proportion of nurses have ten or fewer years of experience?

Problem 4

The graphs below summarize hurricane activity for the 71 years in the period from 1940 to 2010. The top graph has the data on the number of hurricanes that form each year in the whole Atlantic Basin, and the bottom graph has information on how many of those hurricanes hit the state of Florida (Ripol).



4.1 Which is the variable for the top graph? Which type of variable is it?

4.2 Which is the parameter of interest for the top graph?

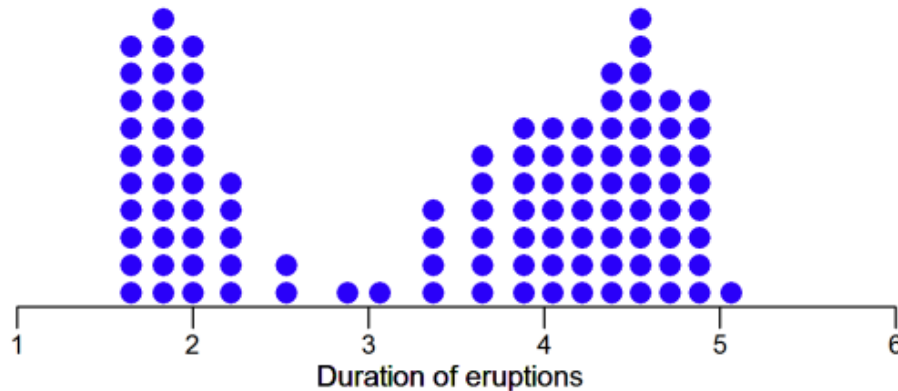
4.3 On an average year we estimated to see about five (5) hurricanes in the Atlantic Basin and none (0) of them hitting Florida. Are the two numbers 5 and 0 hurricanes parameters or statistics?

- a. 5: parameter; 0: statistic
- b. 5: statistic; 0: parameter
- c. Both are statistics
- d. Both are parameters
- e. It is impossible to tell without knowing which is μ and which is \bar{X}

4.4 Which is the value of the median in Florida?

Problem 5

The following plot shows a dot plot of the duration of eruptions in minutes for the Old Faithful Geyser in Yellowstone National Park, Wyoming, USA (Dang).



5.1 What is the minimum value approximately?

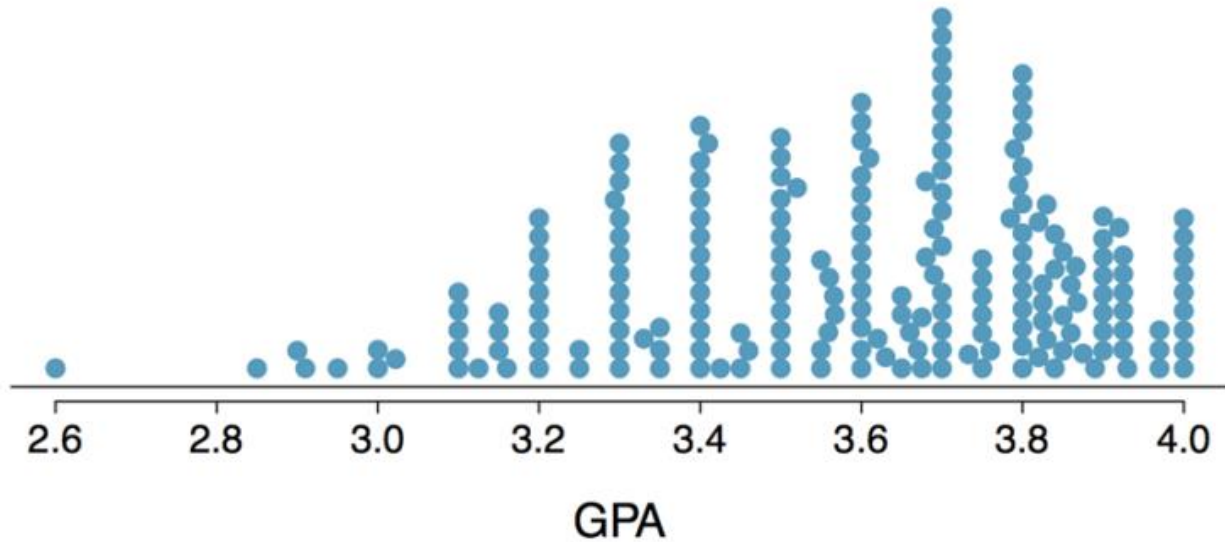
5.2 What is the maximum value approximately?

5.3 What is the range of the data?

5.4 Describe the shape of the distribution.

Problem 6

The following stacked dot plot represents the GPA scores of a group of students in an undergraduate course curriculum at a university.



6.1 Which of the following statements is true for the above stacked dot plot?

- (a) Mode < Median < Mean
- (b) Mode < Mean < Median
- (c) Mean < Mode < Median
- (d) Mean < Median < Mode
- (e) Median < Mean < Mode
- (f) Median < Mode < Mean

6.2 What measures of center and spread are best for the above stacked dot plot?

- (a) Center: Mean; Spread: IQR
- (b) Center: Median; Spread: IQR
- (c) Center: Median; Spread: Range
- (d) Center: Mean; Spread: Standard Deviation
- (e) Center: Median; Spread: Standard Deviation

Problem 7

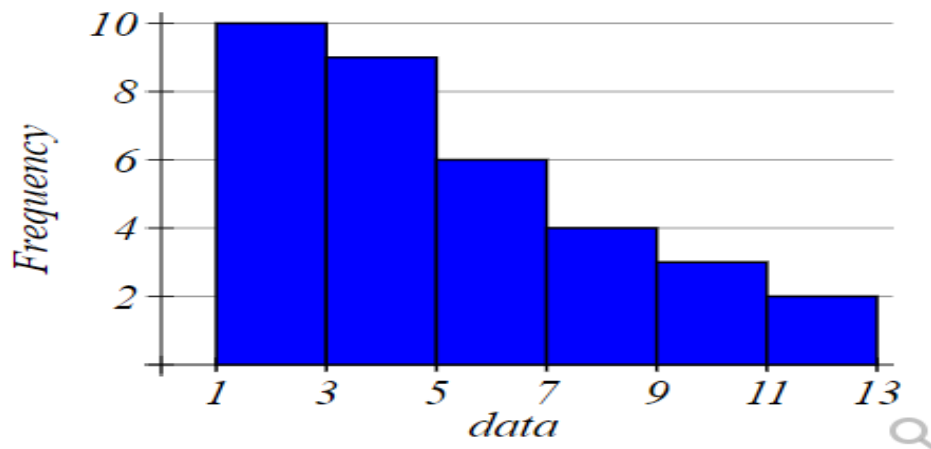
A May 2001 Gallup Poll found that many Americans believe in ghosts and other supernatural phenomena. The poll was based on telephone responses from 1,012 randomly selected adults. The table shows the percentages of people who expressed belief in various phenomena.

Phenomenon	Expressing Belief
Psychic Healing	54%
ESP	50%
Ghosts	38%
Astrology	28%
Channeling	15%

Is it reasonable to conclude that 66% of those polled expressed belief in either ghosts or astrology?

Problem 8

Based on the histogram below:

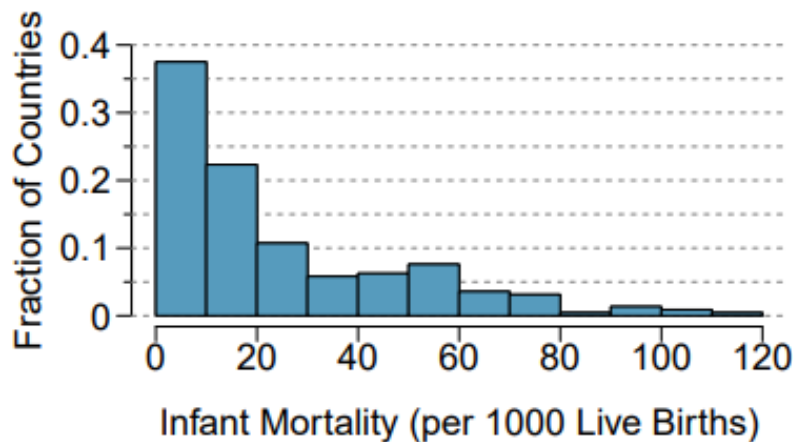


8.1 What is the class or bin width?

8.2 What is the sample size approximately?

Problem 9

The infant mortality rate is defined as the number of infant deaths per 1,000 live births. This rate is often used as an indicator of the level of the health sector in a country. The relative frequency histogram below shows the distribution of estimated infant death rates for 224 countries for which such data were available in 2014¹. (Use this information to answer the following two problems).



9.1 Which is the variable of study? Which type of variable is it?

9.2 What information does the histogram provide?

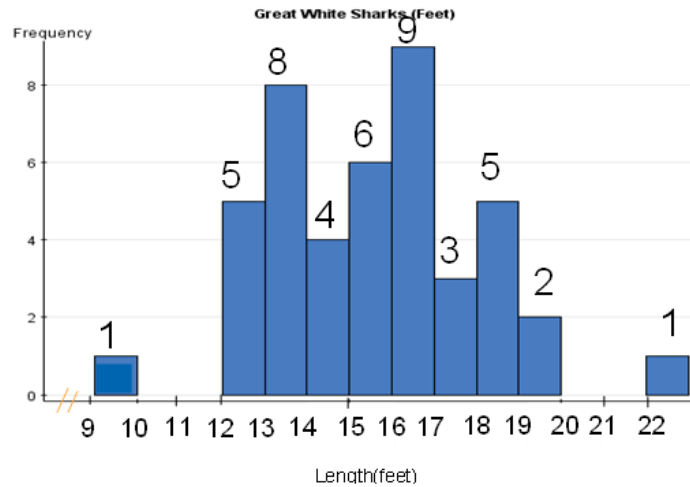
9.3 What proportion of countries has between 0 and 20 infant deaths per 1,000 live births?

9.4 Would you expect the mean of this data set to be smaller or larger than the median? Explain your reasoning

9.5 Which numerical summary is a good measure of the center of the distribution? Explain why.

Problem 10

Below is a histogram of the lengths (measured in feet) of 44 Great White Sharks. (Use this information to answer the following two problems).



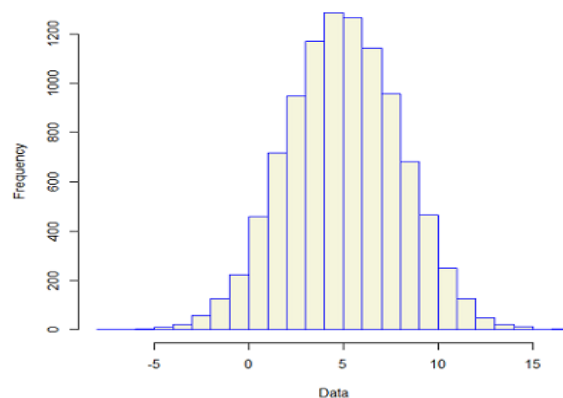
10.1 What proportion of sharks were between 15 and 20 feet? (n=44)

10.2 Which is the exact value of \bar{x} ?

Problem 11

Which numerical summary is a good measure of center for the following distribution?

- a) Median
- b) Mean
- c) Mode
- d) Both median and mean
- e) Median, mean, and mode





Problem 12

- 12.1** What is a reasonable action if an outlier is a legitimate data value and represents natural variability for the group and variable measured?
- 12.2** List statistics that give information only about the location of a dataset.
- 12.3** List statistics that give information only about the spread of a dataset.

Problem 13

In a survey, students are asked how many hours they study in a typical week. The five-number summary measures of the responses are:

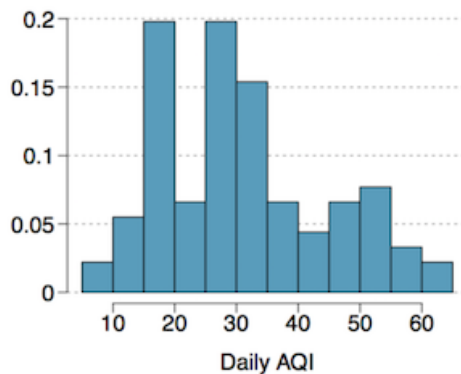
2, 9, 14, 20, 60.

(Use this information to answer the following two problems).

- 13.1** Determine which of the following statements are true for the given dataset. More than one may apply.
- a) The data set has no outliers.
 - b) The data set has exactly one outlier.
 - c) The data set has at least one outlier.
 - d) The data set has exactly two outliers: 2 and 60
 - e) About 50% of the students spend between 9 hours to 14 hours each week studying.
- 13.2** Fill in the blank in the following sentence. About 75% of the students spent at least _____ hours studying in a typical week.
- a) 9
 - b) 14
 - c) 20
 - d) 45
 - e) 60
- 13.3** Which measure of center is not resistant to an outlier in the data?

Problem 14

Daily air quality is measured by the air quality index (AQI) reported by the Environmental Protection Agency. This index reports the pollution level and what associated health effects might be a concern. The index is calculated for five major air pollutants regulated by the Clean Air Act and takes values from 0 to 300, where a higher value indicates lower air quality. AQI was reported for a sample of 91 days in 2011 in Durham, NC. The histogram below shows the distribution of the AQI values on these days.



14.1 Estimate the median AQI value of this sample.

14.2 Would you expect the mean AQI value of this sample to be higher or lower than the median? Explain your reasoning.

Problem 15

Statistics exam scores for 33 students are as follows:

45, 59, 61, 63, 63, 63, 66, 67, 68, 70, 72, 72, 72, 72, 72, 73, 74, 75, 75, 76, 77, 77, 78, 78, 81, 83, 84, 84, 84, 84, 84, 87, 90, 93, 98.

Compute the three quartiles, and the IQR. Check if there is any outlier present in the data.



Problem 16

AIDS data indicating the number of months a patient with AIDS lives after taking a new antibody drug are as follows:

3; 4; 8; 8; 10; 11; 12; 13; 14; 15; 15; 16; 16; 17; 17; 18; 21; 22; 22; 24; 24; 25; 26; 26; 27; 27; 29; 29; 31; 32; 33; 33; 34; 34; 35; 37; 40; 44; 44; 47.

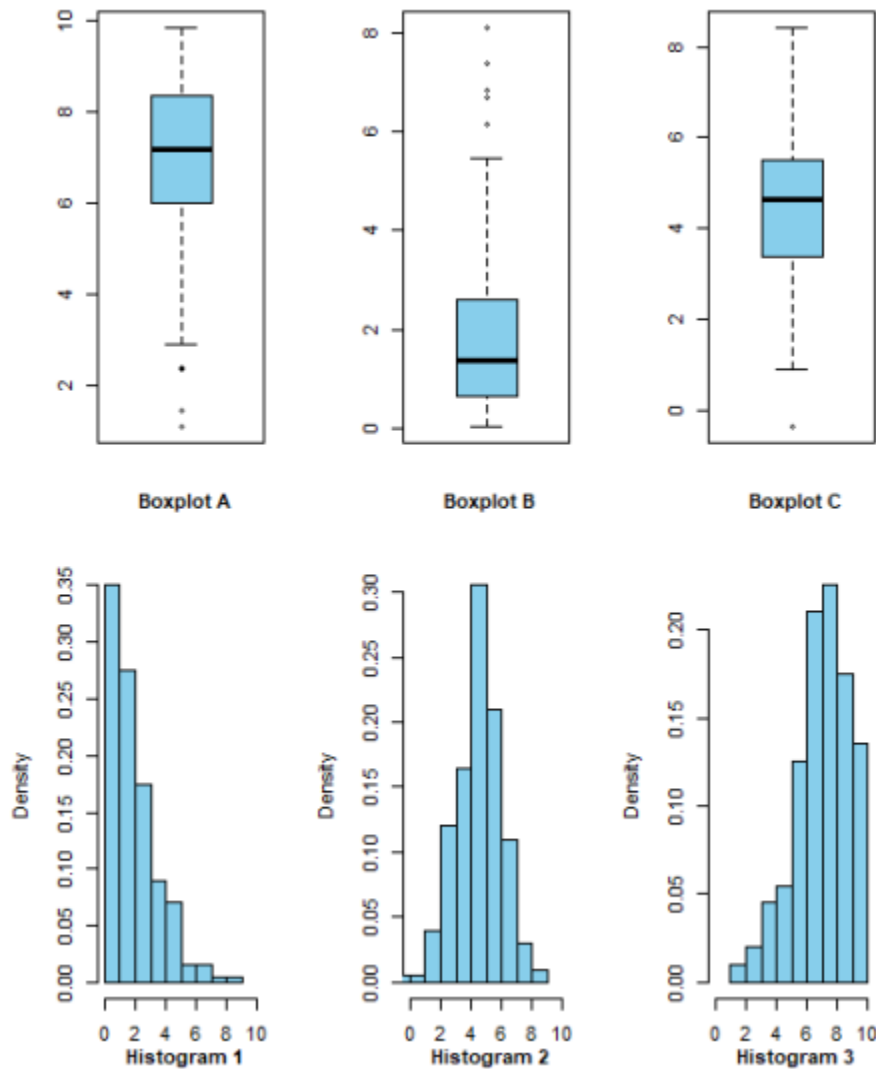
Compute the three quartiles, and the IQR. Check if there is any outlier present in the data.

Problem 17

On a timed math test, the first quartile for time it took to finish the exam was 35 minutes. Interpret the first quartile in this context.

Problem 18

Identify the correct match of the box plots with the histograms given below:



- a) (Histogram 1, Boxplot C); (Histogram 2, Boxplot B); (Histogram 3, Boxplot A).
- b) (Histogram 1, Boxplot B); (Histogram 2, Boxplot A); (Histogram 3, Boxplot C).
- c) (Histogram 1, Boxplot A); (Histogram 2, Boxplot C); (Histogram 3, Boxplot B).
- d) (Histogram 1, Boxplot B); (Histogram 2, Boxplot C); (Histogram 3, Boxplot A).
- e) (Histogram 1, Boxplot C); (Histogram 2, Boxplot A); (Histogram 3, Boxplot B)



Problem 19

Which of the following data sets has the largest standard deviation?

- a) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
- b) 11.75, 12.75, 13.75, 14.75, 15.75, 16.75, 17.75, 18.75, 19.75, 20.75
- c) 0, 0, 0, 0, 0, 10, 10, 10, 10, 10
- d) 51, 52, 53, 54, 55, 56, 57, 58, 59, 60
- e) 88.5, 89.5, 90.5, 91.5, 92.5, 92.5, 93.5, 94.5, 95.5, 96.5, 97.5

Problem 20

Which of the following statements about the sample standard deviation would be incorrect?

- (a) A small value of the sample standard deviation indicates a small amount of variation among the data points.
- (b) If the sample standard deviation is zero, all the observations must be equal to the sample mean, and conversely. (In case of no variability, a measure of spread must be zero.)
- (c) A large value of the sample standard deviation indicates a large amount of variability among the data points, and there are some observations which are away from the sample mean.
- (d) It is resistant to the presence of outliers.
- (e) It works best for symmetric or nearly symmetric distributions.