Section 4: Representing Quantitaive Data

The following maps the videos in this section to the Texas Essential Knowledge and Skills for Mathematics TAC §111.47(c).

4.01 Frequency Tables
- Statistics (1)(E)
- Statistics (4)(B)

4.02 Cumulative Frequency Tables
- Statistics (4)(B)

4.03 Dot Plots and Comparing Distributions
- Statistics (1)(D)
- Statistics (1)(E)
- Statistics (4)(B)

4.04 Stemplots and Comparing Distributions
- Statistics (1)(D)
- Statistics (1)(E)
- Statistics (4)(B)

4.05 Histograms and Comparing Distributions
- Statistics (1)(D)
- Statistics (1)(E)
- Statistics (4)(B)
- Statistics (4)(D)

4.06 Box Plots and Comparing Distributions
- Statistics (1)(D)
- Statistics (1)(E)
- Statistics (4)(B)
- Statistics (4)(D)

Note: Unless stated otherwise, any sample data is fictitious and used solely for the purpose of instruction.

4.01

Frequency Tables

Quantitative data – Data described using numbers that serve as a measurement

- Once data is collected, it needs to be summarized. The _____ of a data set shows the values of the variable and how often they occur.
- Quantitative data can be summarized _____ (using tables or pictures) or _____ (using numbers).
- Graphically summarizing quantitative data allows us to see shape, center, spread, and possible outliers.
- _____ and relative _____ tables display the frequency or relative frequency of each interval.

Consider the following frequency and relative frequency table and make observations about it.

Interval	Frequency	Relative Frequency
[$40, $60)	480	480/1000 = 48%
[$60, $80)	390	390/1000 = 39%
[$80, $120)	130	130/1000 = 13%
Total	1000	1000/1000 = 100%

To construct a frequency table, we divide the observations into _____.

We use intervals with _____ endpoints or choose _____ intervals.

1. Suppose 20 nurses at Medical City Dallas Hospital were randomly sampled and asked how much they spent on food last week while at work. The amounts spent are shown below, sorted in ascending order.

$5	$6	$6	$6	$9	$9	$10	$11	$11	$16
$16	$18	$19	$19	$19	$21	$22	$22	$24	$25

i. Complete the frequency and relative frequency tables below.

Interval	Frequency	Relative Frequency
Total		

ii. How many nurses spent less than $15?

4.02

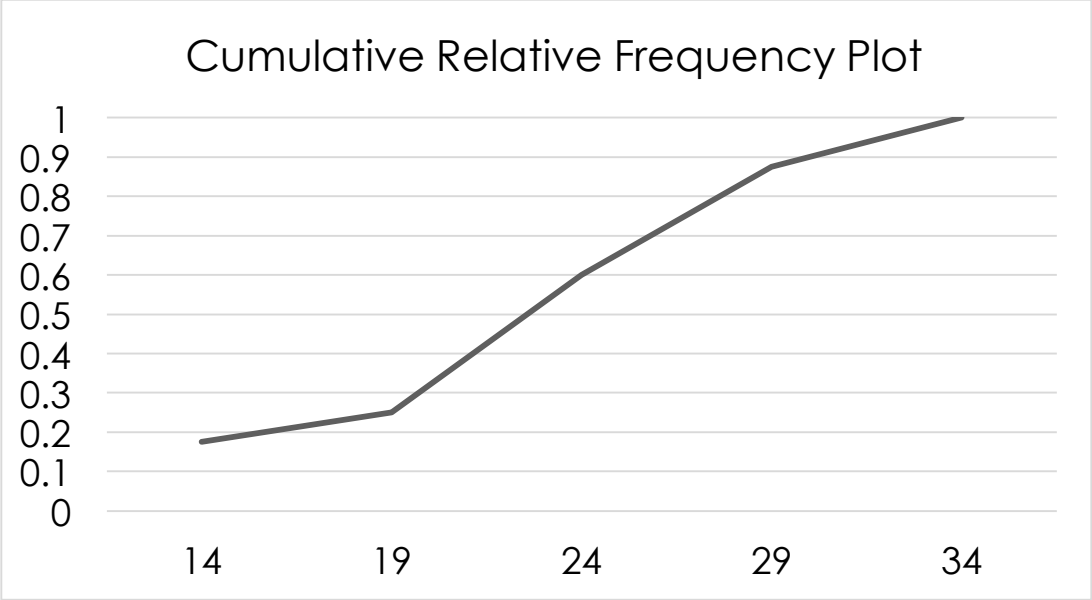
Cumulative Frequency Tables

A ***cumulative frequency distribution*** is used to determine the number of observations below or above a certain value.

Complete the cumulative and relative cumulative frequency tables below.

Interval	Frequency	Relative Frequency	Interval	Cumulative Frequency	Relative Cumulative Frequency
10–14	7	0.175	≤14		
15–19	3	0.075	≤19		
20–24	14	0.350	≤24		
25–29	11	0.275	≤29		
30–34	5	0.125	≤34		
Total	40	1.000			

The cumulative frequency distribution can be displayed using a cumulative relative frequency plot, as in the example below.

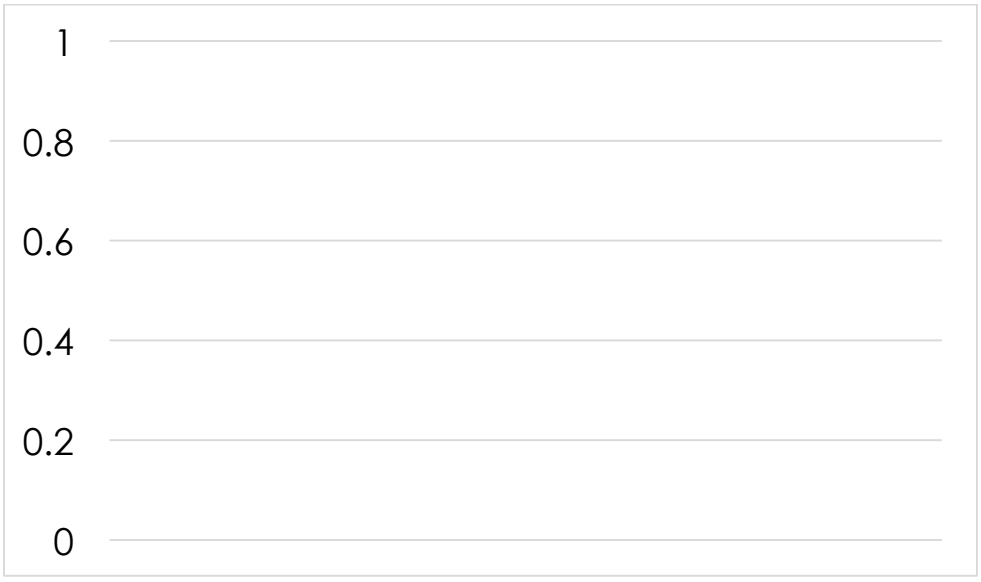


1. Suppose 20 people at the local Kroger supermarket were randomly selected on how much they spent (in dollars). The results are shown below:

79	53	95	48	78	87	102	101	53	58
82	91	40	41	51	61	36	41	85	71

Complete the table below. Then create a cumulative relative frequency plot of the above data.

Interval	Frequency	Relative Frequency	Interval	Cumulative Frequency	Relative Cumulative Frequency
Total					



4.03

Dot Plots and Comparing Distributions

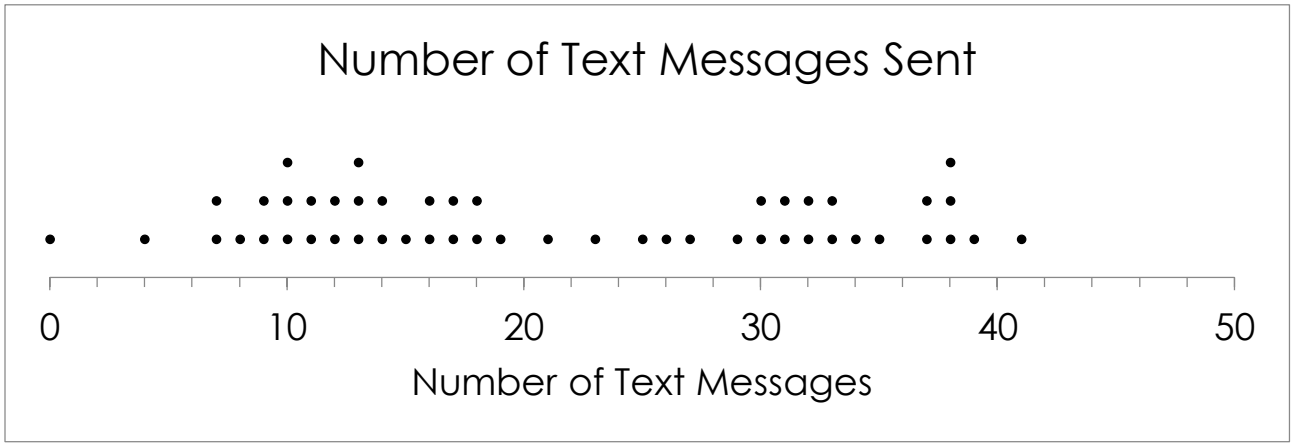
A ***dot plot*** displays each individual observation using a dot.

Why are dot plots not useful for displaying really large data sets?

Imagine that 50 local high school students were randomly sampled and asked how many texts they sent yesterday. The data is shown below.

4	37	9	34	33	11	21	12	10	14
7	38	0	13	27	15	8	13	10	41
14	32	32	25	31	35	31	23	17	18
16	30	10	39	9	11	18	19	12	30
38	16	29	33	13	17	38	26	37	7

A dot plot of the data is shown below.

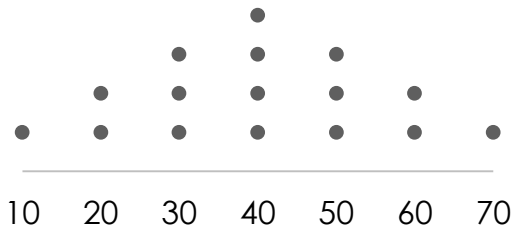


The distribution can be described with respect to shape, center, spread, and potential outliers.

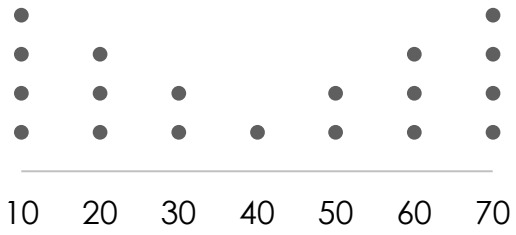
- Shape:
- Center:
- Spread:
- Outliers:

The following are various shapes of distributions. Describe each dot plot in the space provided.

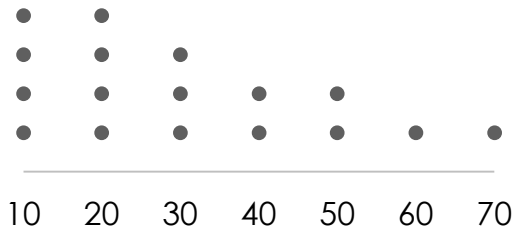
Symmetric (bell-shaped)



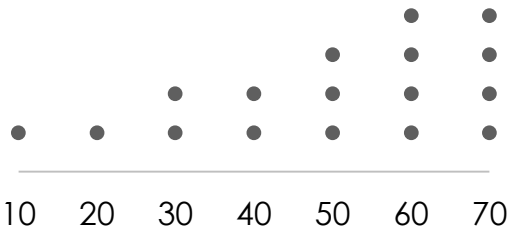
Symmetric (U-shaped)



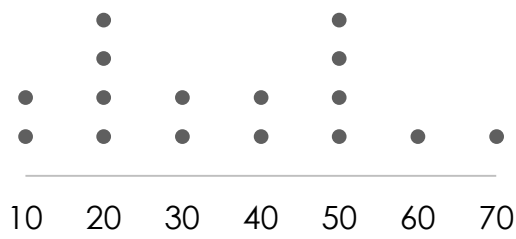
Right-Skewed



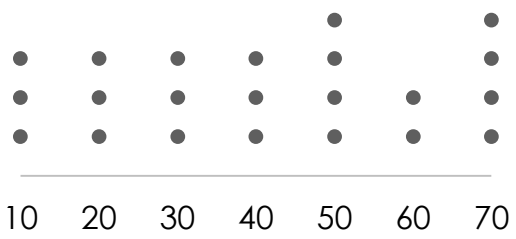
Left-Skewed



Bimodal



Uniform (or rectangular)



1. Suppose two local high schools, School A and School B, compete at Math League every year. School A has won every year. School B believes they continue to lose because they have a team with very diverse ages, instead of having a group with a narrower age cohort.

   The data below give the ages of the 20 students enrolled in Math League at each school.

Ages of the 20 Students in Math League at School A									
19	18	12	17	17	18	14	16	18	17
13	14	18	17	15	16	15	13	19	16

Ages of the 20 Students in Math League at School B									
12	18	13	14	17	12	16	12	13	18
14	13	12	12	15	16	18	15	13	12

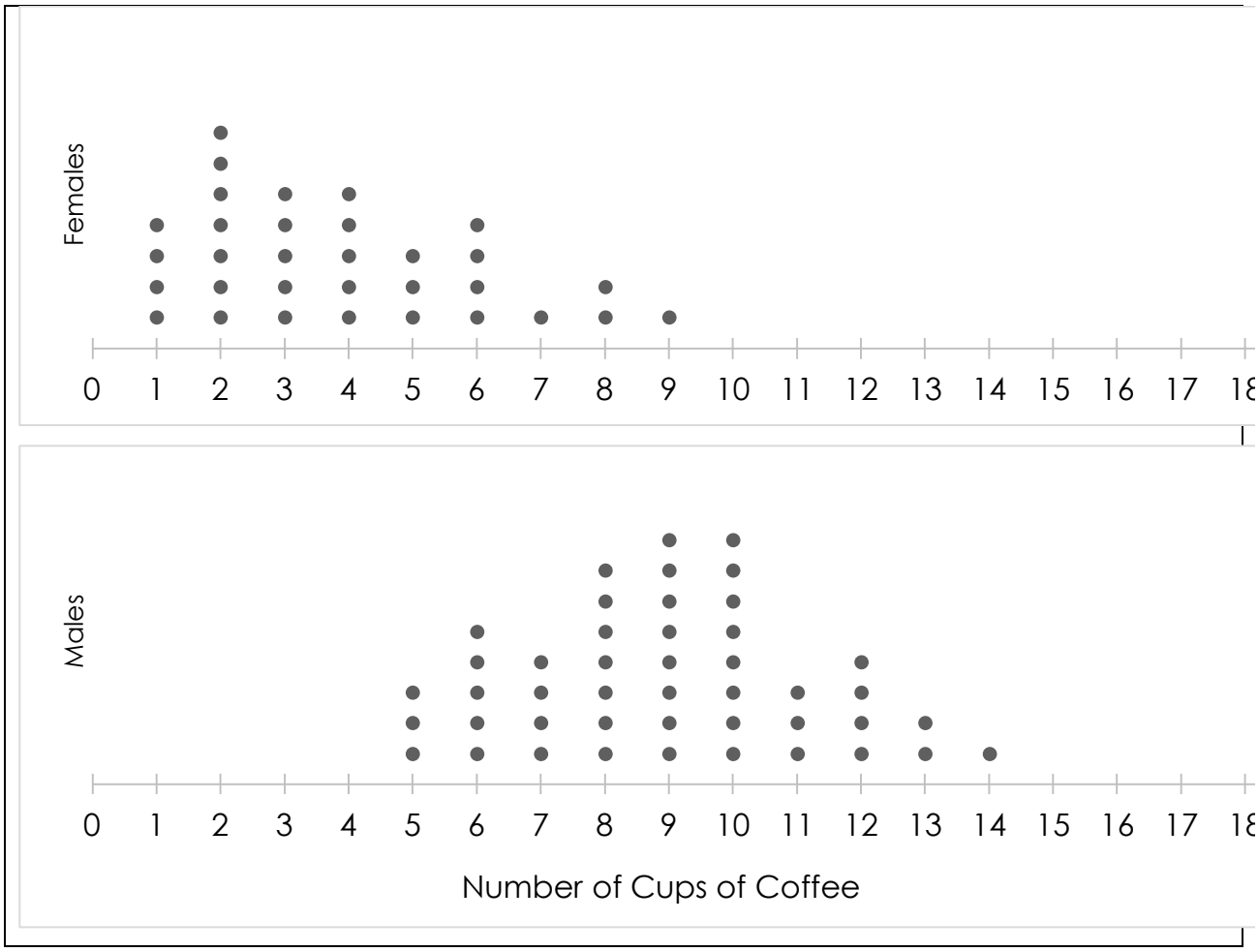
   i. Create a dot plot for each group to compare the distribution of the ages of the 20 students at each school. Comment on the similarities and differences between the two distributions.

School A

School B

   ii. Use the information in the dot plots that you created to discuss the validity of the claims made by School B.

2. Suppose male and female professors from the same college were randomly sampled and asked how many cups of coffee they drink per week. Compare the distributions.



4.04

Stemplots and Comparing Distributions

A ***stemplot*** displays quantitative data, usually from small data sets, using stems and leaves.

- Stemplots always include a key to decode the units of the stem and leaf, to prevent confusion.
- **Example:** 4 | 1 could represent 41, or 4 | 1 could represent 4.1.

Why are stemplots not useful for displaying really large data sets?

Consider the following stemplots and compare their distributions.

Stemplot A	
2	6 7 8 9
3	0 2 3 6 7
4	2 5
5	
6	9

Stemplot B	
1	4
2	5 8 9
3	1 2 3 3 4
4	6 7
5	1

Stemplot C	
1	2
2	7
3	3
4	5 6 7 9 9
5	0 2 3 4

Key: 2 | 6 = 26

1. Suppose 20 people at the local Kroger supermarket were randomly selected on how much they spent (in dollars). The results are shown below:

79	53	95	48	78
82	91	40	41	51
87	102	101	53	58
61	36	41	85	71

Construct and properly label a stemplot of the data set.

2. Suppose 50 college students were randomly sampled and asked how many texts they sent the previous day. The results are shown below.

4	37	9	34	33	11	21	12	10	14
7	38	0	13	27	15	8	13	10	41
14	32	32	25	31	35	31	23	17	18
16	30	10	39	9	11	18	19	12	30
38	16	29	33	13	17	38	26	37	7

Shown below is a stemplot and ***split stemplot*** of the data.

Stemplot of Text Messages

0	0 4 7 7 8 9 9
1	0 0 0 1 1 2 2 3 3 33 4 4 5 6 6 7 7 7 8 8 9
2	1 3 5 6 7 9
3	0 0 1 1 2 2 3 3 4 5 7 7 8 8 8 9
4	1

(4 | 1 represents 41 texts)

Split Stemplot of Text Messages

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

(4 | 1 represents 41 texts)

What is the difference between a stemplot and a split stemplot?

3. Shown below is a back-to-back split stemplot displaying the number of hours of TV watched per week for a hypothetical sample of males and females.

Males		Females
4 4 4 3 3 2 1 1 0 0	0	3
9 7 7 7 6 5 5	0	
4 2 2 1 0 0	1	4 4
6 5	1	5 6 7 7 7 8
1 1	2	1 1 1 1 2 2 2 3
	2	5 5 6 6 7 7 7 8 8 9
0	3	0 0 0 0 0 1 1 1 1 2 4
	3	0 0 1
	4	0

(4 | 0 represents 40)

i. What do the stems and leaves represent in the stemplot?

ii. Compare the shape, center, and spread of the distributions. Are there any outliers?

4.05

Histograms and Comparing Distributions

Histograms display the frequency or relative frequency of data over intervals.

Are histograms useful for displaying small data sets? Explain.

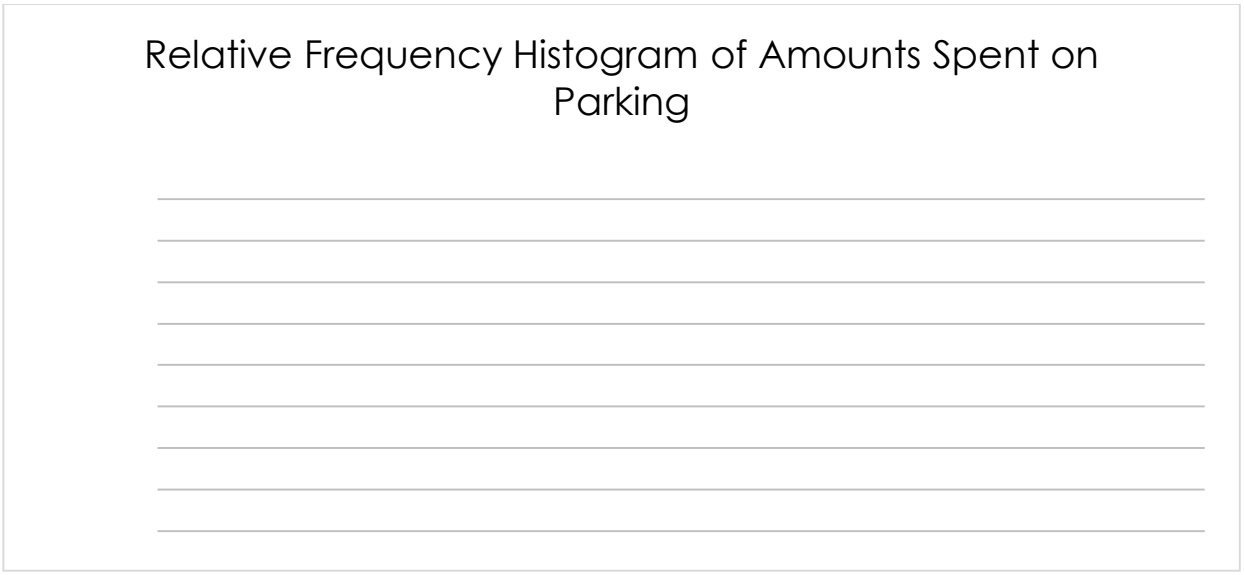
Suppose 20 city residents were randomly sampled and asked how much they spent on city parking last month. The amounts spent are shown below sorted from lowest to highest.

$15	$17	$18	$18	$19	$22	$25	$25	$26	$27
$28	$28	$29	$31	$33	$34	$43	$56	$68	$72

Shown below is a _____ and _____ of the amounts spent.

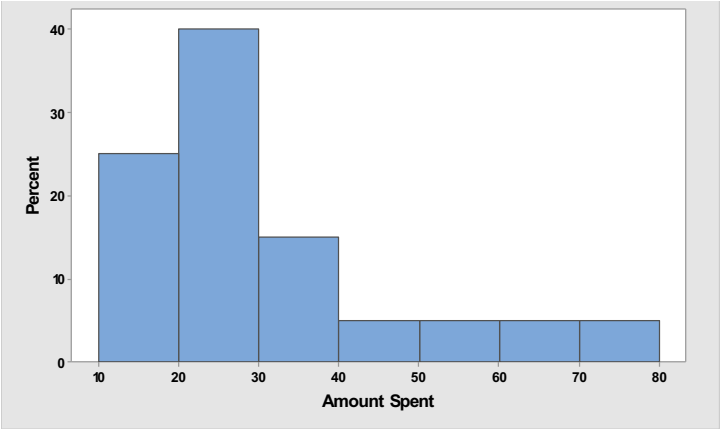
Interval	Frequency	Relative Frequency
[$10, $20)	5	5/20 = 25%
[$20, $30)	8	8/20 = 40%
[$30, $40)	3	3/20 = 15%
[$40, $50)	1	1/20 = 5%
[$50, $60)	1	1/20 = 5%
[$60, $70)	1	1/20 = 5%
[$70, $80)	1	1/20 = 5%
Total	20	20/20 = 100%

1. Construct a frequency histogram and relative frequency histogram from the parking expense data.

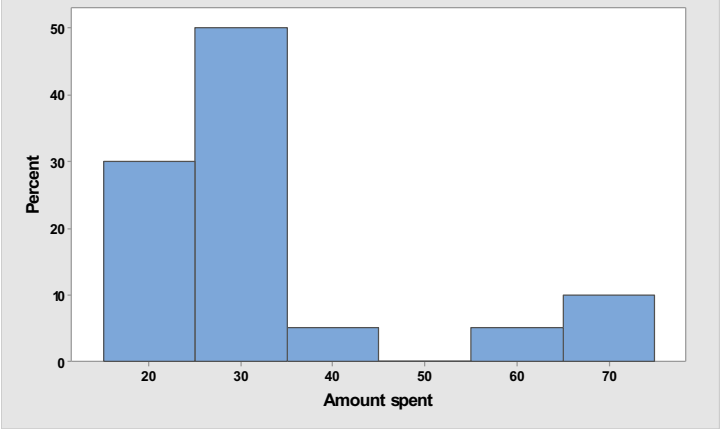


1. The histograms below summarize the same parking expense data from the previous question. Explain the differences between Histogram A and Histogram B.

Histogram A

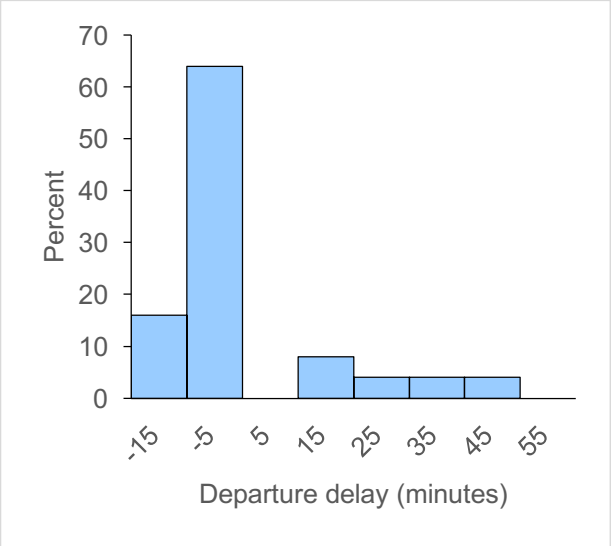


Histogram B

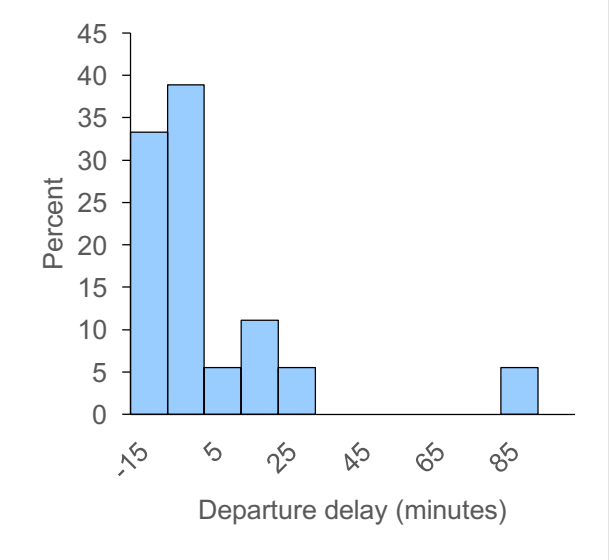


2. The relative frequency histograms below display the delay times of American Airlines flights departing from Houston, Texas, on two different New Year's Days: 25 flights on New Year's Day 2016 and 18 flights on New Year's Day 2017 (Bureau of Transportation Statistics, n.d.).

New Year's Day 2016



New Year's Day 2017



i. Compare the distributions.

ii. What is your advice to a group of friends who plan to fly from Houston on New Year's Day, based on this data?

4.06

Box Plots and Comparing Distributions

A ***box plot*** displays the distribution of a data set based on the five-number summary: minimum, first quartile, median, third quartile, and maximum.

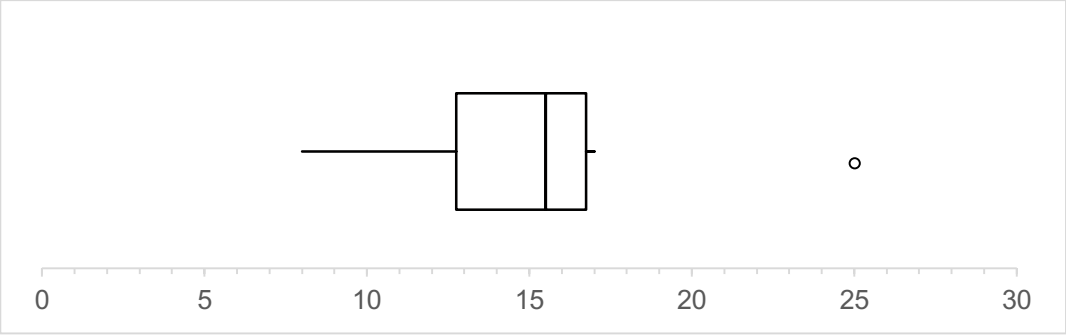
What is the five-number summary?

- _____: the smallest observation in the data set
- _____: the 25th percentile (bottom half median)
- _____: the 50th percentile (middle ordered observation)
- _____: the 75th percentile (top half median)
- _____: the largest observation in the data set

_____ extend to the minimum and maximum values that are not outliers.

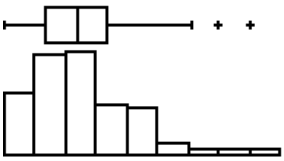
_____ are observations that fall below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$, where IQR is the ***interquartile range***, defined as $IQR = Q3 - Q1$.

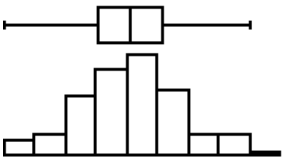
1. Label the components of the box plot below.



2. Consider the following box plots and their data's respective histograms.

   i. Write three observations to compare and contrast the following distributions.

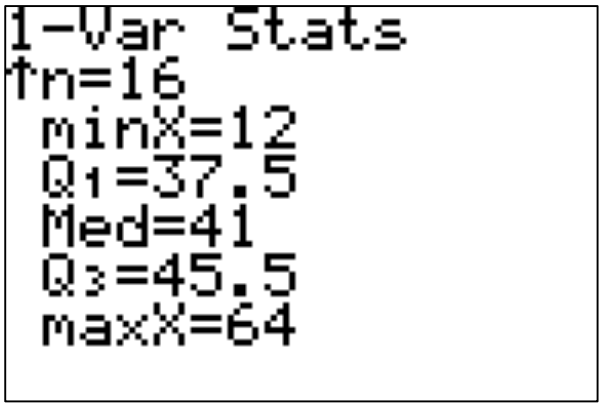




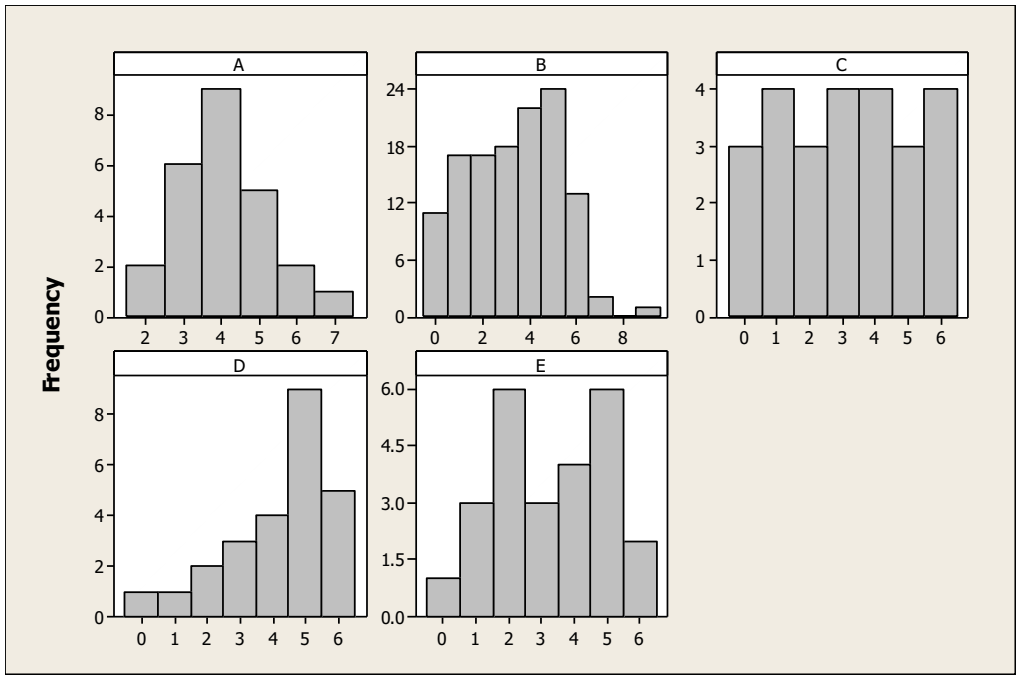
   ii. Provide examples of real-life scenarios that could be represented by each distribution.

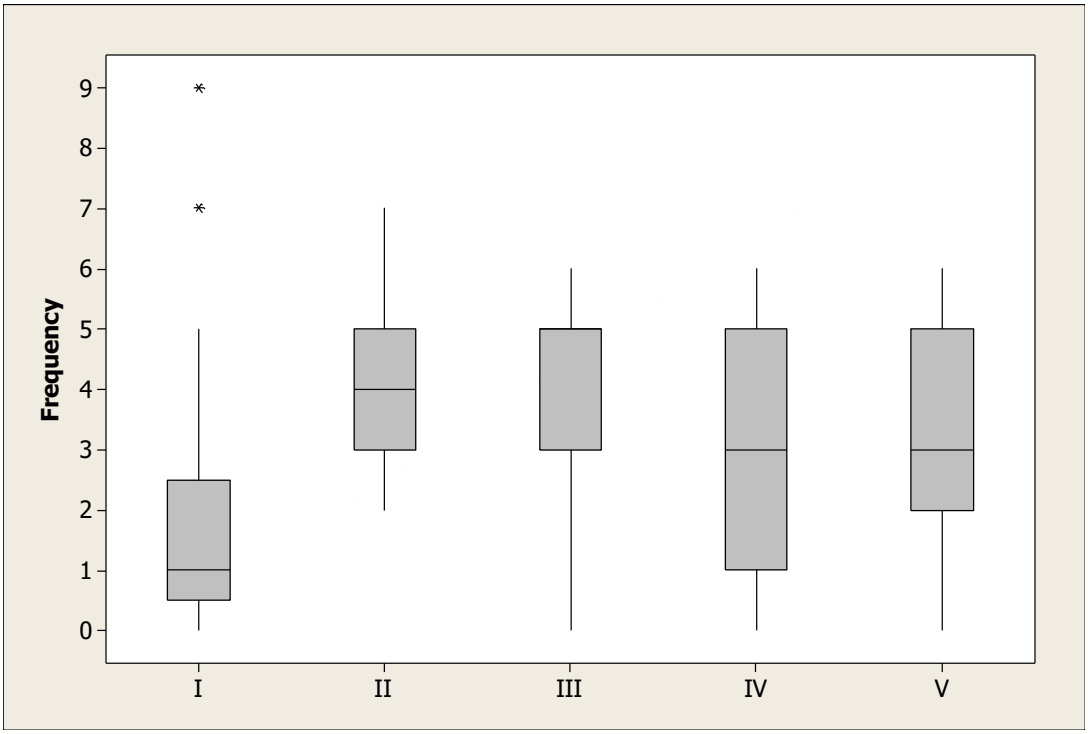
3. Use the TI-84 calculator output on the right to construct a box plot for the sorted data set on the left.

12	25	35	37
38	39	40	41
41	42	44	45
46	46	51	64



4. Match each histogram (A, B, C, D, and E) to its corresponding box plot (I, II, III, IV, and V).





References

Bureau of Transportation Statistics. (n.d.). *Airline On-Time Statistics: Detailed Statistics Departures*. Retrieved from https://www.transtats.bts.gov/ONTIME/Departures.aspx