SECTION 2.2 – Organizing Quantitative Data: The Popular Ways

Classes – Categories for grouping quantitative data.

In summarizing quantitative data, we first determine whether the data are discrete or continuous. If the data are discrete with relatively few different values of the variable, then the categories of data (called classes) will be the observations (as in qualitative data). If the data are discrete, but with many different values of the variable or if the data are continuous, then the categories of data (the classes) must be created using intervals of numbers. We will first present the techniques required to organize discrete quantitative data when there are relatively few different values and then proceed to organizing continuous quantitative data.

(There are several ways to group quantitative data. The text focuses on the following two.)

METHOD 1 – Single-Value Grouping

Grouping quantitative data that uses classes that represent a single possible value. (Discrete data.)

METHOD 2 – Limit Grouping

Grouping quantitative data that uses classes that represent a range of values called class limits.
(Discrete & Continuous data.)

Lower class limit: The smallest value that could go in a class.

Upper class limit: The largest value that could go in a class.

Class Width: The difference between consecutive lower class limits.

SECTION 2.2 – Organizing Quantitative Data: The Popular Ways

☺ **Exercises**:

1) Twenty overall ratings from 1 through 10 of a certain Netflix show are given below:

4, 2, 5, 9, 10, 2, 8, 1, 7, 9, 7, 4, 1, 9, 7, 6, 4, 1, 8, 7

Construct both a frequency distribution and a relative frequency distribution using "Single-Value Grouping" based off this data.

Frequency distribution

Ratings	Tally	Frequency

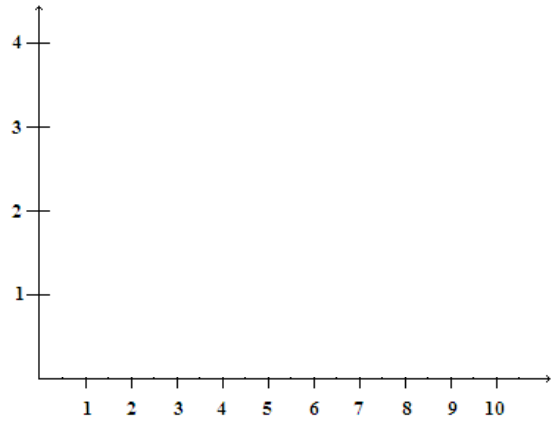
Relative Frequency distribution

Ratings	Tally	Frequency	Relative Frequency

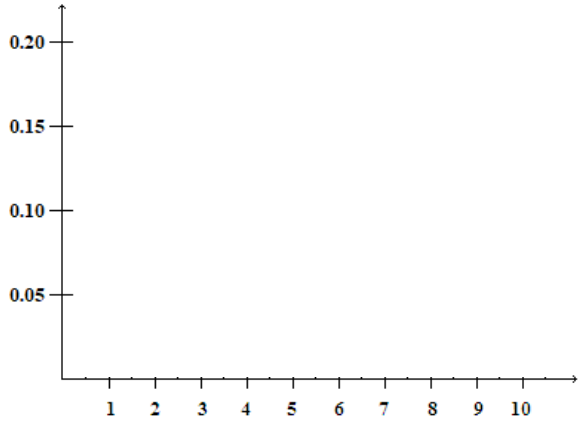
Histogram – A **histogram** is constructed by drawing rectangles for each class of data. The height of each rectangle is the frequency or relative frequency of the class. The width of each rectangle is the same and the rectangles (bars) touch each other.

2) Draw a frequency histogram and a relative frequency histogram for the data above.

Frequency Histogram



Relative Frequency Histogram



SECTION 2.2 – Organizing Quantitative Data: The Popular Ways

☺ **Class Activity & Exercise:**

3) The data will be collected by each member of the class. Each student will write the total number of television sets in their household. Construct both a frequency distribution and a relative frequency distribution using "Single-Value Grouping" based off this quantitative data.

Frequency distribution using "Single Value Grouping"

Number of TV's	Tally	Frequency
0		
1		
2		
3		
4		
5		
6		
7		
8		

Relative Frequency distribution using "Single Value Grouping"

Number of TV's	Tally	Frequency	Relative Frequency
0			
1			
2			
3			
4			
5			
6			
7			
8			

4) Using the same data from above, construct both a frequency distribution and a relative frequency distribution using "Limit Grouping" with a class width of 2 and the first lower class limit of 0.

Frequency distribution using "Limit Grouping" with class width of 2

Number of TV's	Frequency

Relative Frequency distribution using "Limit Grouping" with class width of 2

Number of TV's	Frequency	Relative Frequency

5) Using the same data from above, construct both a frequency distribution and a relative frequency distribution using "Limit Grouping" with a class width of 3 and the first lower class limit of 0.

Frequency distribution using "Limit Grouping" with class width of 3

Number of TV's	Frequency

Relative Frequency distribution using "Limit Grouping" with class width of 3

Number of TV's	Frequency	Relative Frequency

SECTION 2.2 – Organizing Quantitative Data: The Popular Ways

☺ **Exercises**:

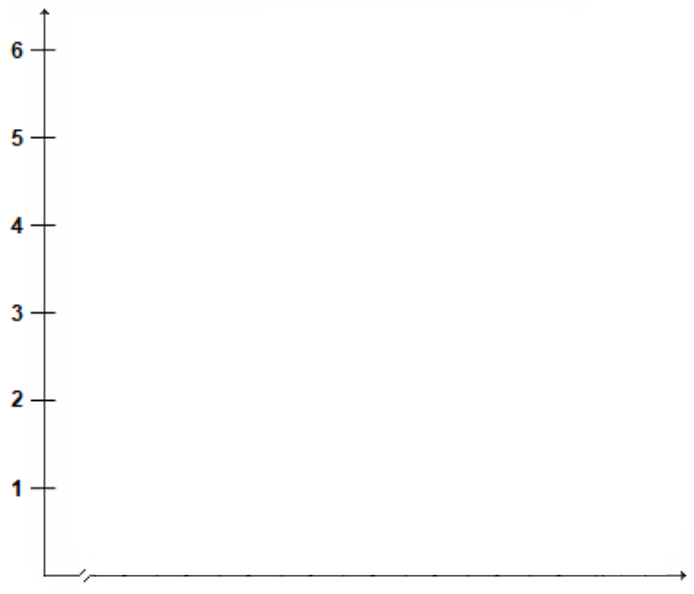
6) Twenty AA batteries were tested to determine how long they would last. The results, to the nearest minute, were recorded as follows:

423, 369, 387, 411, 393, 399, 371, 377, 409, 392,
417, 431, 401, 363, 391, 405, 425, 400, 381, 399

Construct a **frequency distribution** using "Limit Grouping" with the smallest data value representing the lower limit of the first class and using a class width of 9.

Lifetime of battery	Tally	Frequency

7) Draw a frequency histogram for the data above.



SECTION 2.2 – Organizing Quantitative Data: The Popular Ways

☺ **Exercises**:

8) Rock engineers widely use the uniaxial compressive strength (UCS) of rocks in designing surface and underground structures. A precise sample of 20 representative amphibolite samples were drawn from varied rock slopes along Chengdu–Wenchuan highway in Sichuan Province, China. The data in UCS is shown in the table below.

100.6	112	117.5	100.9	94.5	91.5	97.5	100	108.5	87
78	93.9	111.7	114	104.5	118	116.5	115	103.9	110.2

Construct a **frequency distribution** using "Limit Grouping" with the smallest data value representing the lower limit of the first class and using a class width of 8.

Since the data here are not integers, we cannot follow the same class formatting from the previous page. For example, the first class should be 78–85, followed by 86–93, 94–101, 102–109, 110–117, & 118–125. Looking at the observation 93.9 from the data set, this wouldn't fall into any of these classes. The same holds true for the observation 117.5. So, instead we will need to adjust our upper limits of each class to an appropriate number of decimal places.

Frequency distribution using "Limit Grouping" with class width of 8 and lower limit of first class 78

UCS	Tally	Frequency

SECTION 2.2 – Organizing Quantitative Data: The Popular Ways

When collecting your own data and constructing frequency or relative frequency distributions, you will need to create your own class widths along with the lower class limit of the first class. Below are some guidelines in doing so.

Guidelines for Determining the Lower Class Limit of the First Class and Class Width

Choosing the Lower Class Limit of the First Class

Choose the smallest observation in the data set or a convenient number slightly lower than the smallest observation in the data set.

Determining the Class Width

- Decide on the number of classes. Generally, there should be between 5 and 20 classes. The smaller the data set, the fewer classes you should have. When the data set is large, we usually, but not always necessarily want more classes.

- You could also determine the class width by computing

$$\text{Class width} = \frac{\text{largest data value} - \text{smallest data value}}{\text{number of classes}}$$

Round this value **up** to a convenient number. Rounding up may result in fewer classes than were originally intended.

☺ **Exercises**:

9) The following table provides the team earned run average (ERA) for all 15 American League teams from 2017 in Major League Baseball.

AL West			AL Central			AL East	
Team	**ERA**		**Team**	**ERA**		**Team**	**ERA**
Angels	4.20		Indians	3.30		Blue Jays	4.42
Astros	4.12		Royals	4.61		Orioles	4.97
Athletics	4.67		Tigers	5.36		Rays	3.97
Mariners	4.46		Twins	4.59		Red Sox	3.70
Rangers	4.66		White Sox	4.78		Yankees	3.72

Construct a **frequency distribution** using "Limit Grouping" with the smallest data value representing the lower limit of the first class and determine the class width by using the guidelines and the formula on page #2. Use 5 classes.

ERA	Tally	Frequency

SECTION 2.2 – Organizing Quantitative Data: The Popular Ways

Stem-and-leaf plot – A **stem-and-leaf plot** is another way to represent quantitative data graphically. In a stem-and-leaf plot (or *stem plot*), we use the digits to the left of the rightmost digit to form the **stem**. Each rightmost digit forms a **leaf**. For example, a data value of 147 would have 14 as the stem and 7 as the leaf.

Construction of a Stem-and-Leaf Plot

Step 1 The stem of a data value will consist of the digits to the left of the rightmost digit. The leaf of a data value will be the rightmost digit.

Step 2 Write the stems in a vertical column in increasing order. Draw a vertical line to the right of the stems.

Step 3 Write each leaf corresponding to the stems to the right of the vertical line.

Step 4 Within each stem, rearrange the leaves in ascending order, title the plot, and include a legend to indicate what the values represent.

☺ **Example #1**:
The temperature of 30 cities were recorded on a particular day (rounded to the nearest degree):

56, 68, 65, 70, 87, 67, 58, 94, 66, 64, 80, 72, 71, 69, 113
63, 70, 71, 68, 88, 67, 102, 70, 92, 87, 77, 64, 64, 51, 72

To draw a stem-and-leaf diagram, the stems in this particular case are the tens digits' of the temperature scores and the leaves are the ones digits'.

5	681		5	168
6	85764938744		6	34445677889
7	02101072		7	00011227
8	7087	→ → →	8	0778
9	42		9	24
10	2		10	2
11	3		11	3

Legend: 5 | 1 represents 51

Notice that the leaves are recorded in the order in which they appear in the left plot. We then refine the stem and leaf plot by listing the leaves in increasing order, as in the plot to the right.

The 5 | 168 represents the scores of 51, 56, and 58. These are the three scores in the 50's. This trend carries throughout the other remaining lines.

☺ **Exercises:**

Draw a stem-and-leaf diagram for the following set of data.

10) The ages of 15 customers that used coupons at a local supermarket were recorded:
41, 56, 44, 32, 21, 26, 41, 39, 35, 32, 37, 28, 29, 40, 27

SECTION 2.2 – Organizing Quantitative Data: The Popular Ways

☺ **Example #2:**

The GPA's (rounded to the nearest tenth) of a fundamental high school were noted:

2.8, 2.9, 2.0, 2.4, 3.4, 4.0, 3.8, 3.6, 3.8, 2.9, 2.0, 2.8, 2.7, 0.9
4.0, 3.5, 3.5, 2.8, 2.4, 2.3, 3.5, 2.8, 2.9, 3.2, 3.5, 3.7, 4.0, 2.2

Draw a stem-and-leaf plot for the data.

0	9
1	
2	00234478888999
3	2455556788
4	000

Legend: 0 | 9 represents 0.9

☺ **Example #3:**

The data below shows 47 test scores for a class exam:

32, 67, 89, 90, 87, 72, 75, 88, 95, 83, 97, 72, 85, 93, 79, 63, 83, 76, 81, 79, 80, 65, 77, 78
70, 87, 74, 86, 98, 100, 97, 85, 77, 88, 92, 94, 81, 76, 64, 87, 79, 89, 92, 94, 83, 82, 79

Draw a stem-and-leaf plot for the data. (Here, you may use two lines per stem in the event of lengthy leaf's).

3	2
6	3457
7	0224
7	5667789999
8	0112333
8	5567778899
9	022344
9	5778
10	0

Legend: 3 | 2 represents 32

or

3	2
3	
6	34
6	57
7	0224
7	5667789999
8	0112333
8	5567778899
9	022344
9	5778
10	0
10	

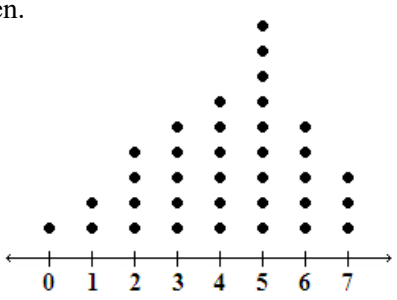
Dotplot

A **dotplot** is a type of display that contains a horizontal number line with a dot used above the number line to represent each value.

☺ **Example #4:**

The number of tornadoes each year in a certain county from 1983 to 2017 is given.

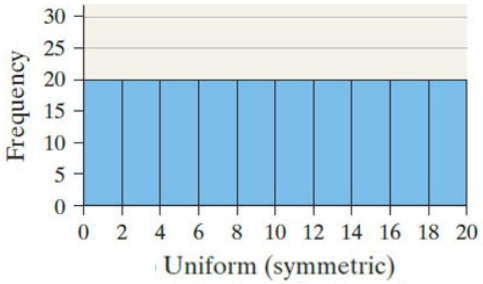
2 , 2 , 4 , 3 , 3 , 0 , 1 , 6 , 5 , 5 , 5 , 3 , 4 , 2 , 6 , 5 , 5 , 7 ,
5 , 3 , 6 , 1 , 2 , 3 , 7 , 4 , 6 , 7 , 4 , 6 , 5 , 5 , 5 , 4 , 4



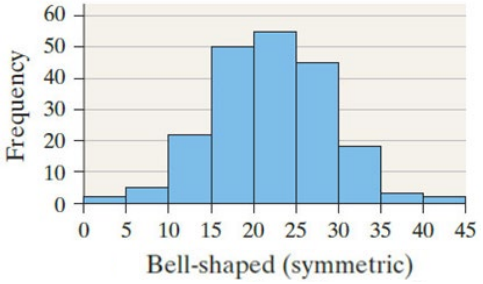
Each dot represents a year in which there were that many tornadoes.

SECTION 2.2 – Organizing Quantitative Data: The Popular Ways

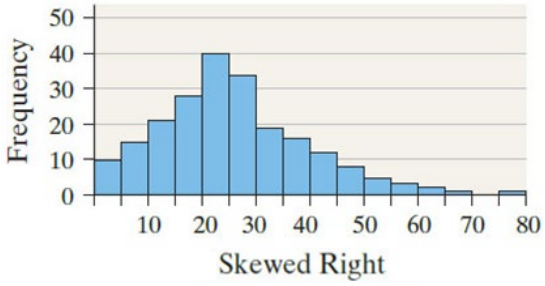
Uniform distribution
The frequency of each value of the variable is evenly spread out across the values of the variable



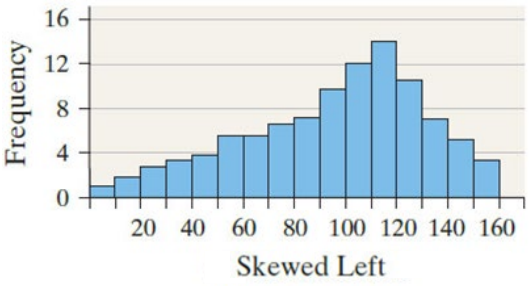
Bell-shaped distribution
The highest frequency occurs in the middle and frequencies tail off to the left and right of the middle



Skewed right
The tail to the right of the peak is longer than the tail to the left of the peak.



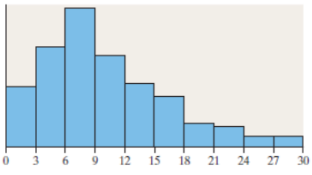
Skewed left
The tail to the left of the peak is longer than the tail to the right of the peak.



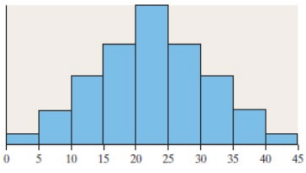
☺ **Exercises:**

For exercises 11 through 14, identify the overall shape of each distribution.

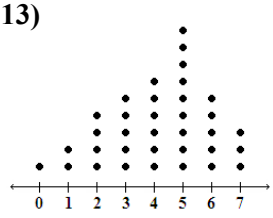
11)



12)



13)



14)

7	01138
8	234679
9	34557
10	13469
11	235888